The
University
Of
Sheffield.

Automatic
Control &
Systems
Engineering.

Control in Portfolio Optimization

Matius Chong

9th May 2025

Supervisor: Dr Paul Trodden

A dissertation submitted in partial fulfilment of the requirements for the degree of BEng Computer Systems Engineering

Abstract

The mathematical formulation of portfolio optimization is built up from basic principles in the Markowitz mean-variance framework. The return and risk of a portfolio of financial assets is modelled as the expectation and variance of a discrete random variable respectively using Modern Portfolio Theory. Then, an investor's goals are formulated as an objective function with constraints as a convex optimization problem. Single period and multi-period models are considered, and control is introduced via Model Predictive Control and Reinforcement Learning. We use historical data from Nov 2023 to Nov 2024 of the "Magnificent 7" stocks in the US market from Yahoo Finance as a numerical example to illustrate these concepts. Accurate estimates of future returns are shown to be the deciding factor in practical portfolio optimization.

Individual Contribution

As portfolio optimization is a widely established field, there were a wide array of state-of-the-art literature using advanced mathematics that I felt I could not contribute much to. As such, this paper focuses on its mathematical derivations and its modelling factors. The novelty of this paper are as follows:

- Illustrating basic concepts in Modern Portfolio Theory and the classic Markowitz framework using real data of the "Mag 7" tech stocks.
- Calafiore (2008) MPO's single decision variable framework is used with point estimates and no-serial correlation assumption to formulate the 2-period, 4-period case, and MPC model in section 5.2, 5.3, chapter 6, and tested with real data.
- As far as I am aware, a novel RL algorithm is developed using basic RL concepts emulating the traditional mean-variance objective with a risk penalty.

The project was done completely over code in Python and Vscode as the development environment software. All code can be found online at (https://github.com/mateusy7/portopt).

Contents

Chapter 1 Introduction ... 5

1.1 Background and Motivation ... 5

1.2 Aims and Objectives ... 5

1.3 Project Management ... 6

1.4 Overview of Report ... 7

Chapter 2 Literature Review ... 8

Chapter 3 Concepts and Terminology ... 10

3.1 Return ... 10

3.2 Risk ... 10

3.3 Efficiency ... 10

Chapter 4 Modelling stocks with Modern Portfolio Theory ... 12

4.1 Two asset portfolio ... 12

4.2 Risk-free asset and the Tangent Portfolio ... 13

4.3 Multi-asset portfolio ... 15

Chapter 5 Portfolio Optimization with Convex Optimization ... 20

5.1 Single Period Optimization (SPO) ... 20

5.2 Multi Period Optimization (MPO) ... 22

5.3 Effect of length of periods in multi-period models ... 26

Chapter 6 Model Predictive Control and practical concerns ... 30

Chapter 7 Reinforcement Learning 33

Chapter 8 Conclusions 37

References 38

Chapter 1 Introduction

1.1 Background and Motivation

Portfolio optimization is the process of allocating capital across a selection of financial assets to achieve specific investment objectives, most commonly maximizing expected return for a given level of risk. It forms the foundation of modern investment management and is used across diverse contexts—from algorithmic trading to the long-term asset allocation strategies of pension funds and endowments. By systematically balancing the trade-off between risk and return, portfolio optimization enables investors to make data-driven, rational decisions aligned with their goals and risk tolerance.

1.2 Aims and Objectives

The aim of this paper is to explain the mathematical formulation of portfolio optimization from first principles and build up to current methods. Secondly, to show how predictions and control variables affect the system. Finally, what the characteristics of different models are and to compare their performance. Treatment of the subject will be limited to discrete time.

Basic Objectives

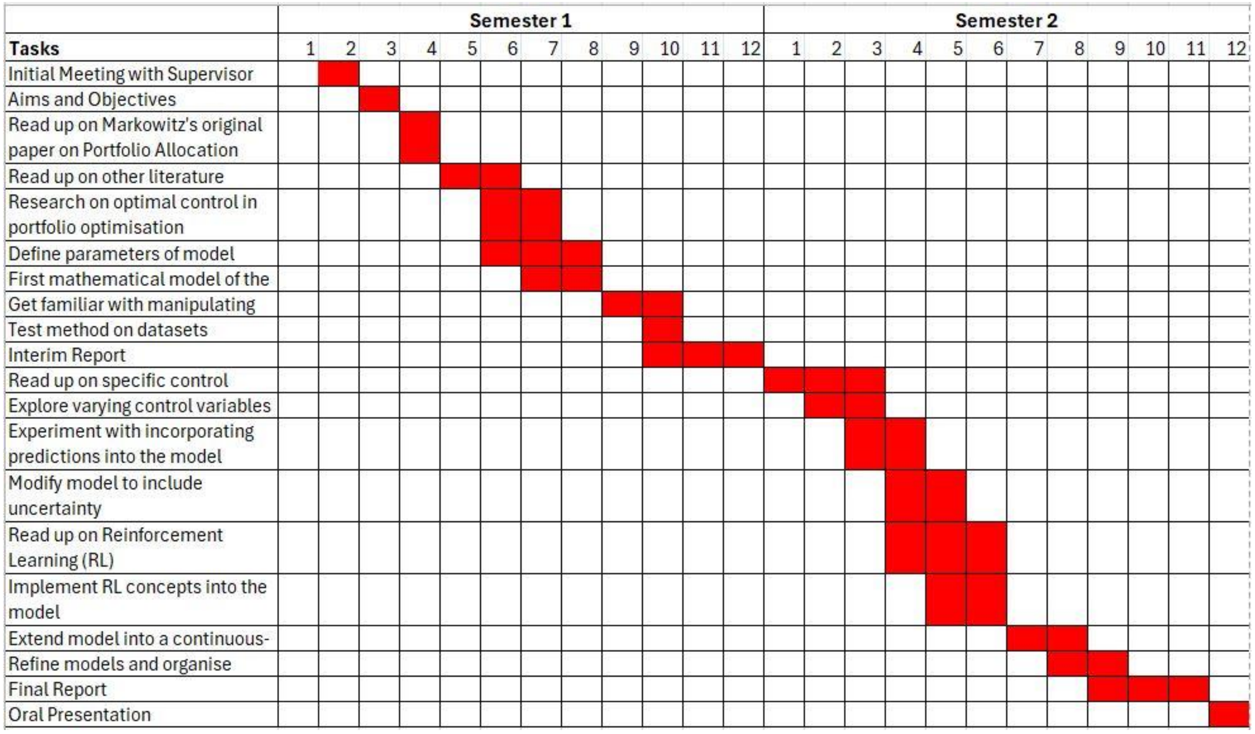
1. Review the literature on portfolio optimization focusing on discrete-time settings.
2. Explain and build a model of portfolio optimization using current methods.
3. Illustrate how portfolio optimization can be modelled as a state-spaced control problem under constraints in single and multi-period formulations.
4. Explore how using different control variables and varying them affect the system dynamics and how predictions can be incorporated into the model.

Advanced Objectives

5. Attempt to characterize the uncertainty of the system and incorporate it into the model.
6. Incorporate Reinforcement Learning as a control technique in portfolio optimization.

7. Illustrate at a basic level the extension of portfolio optimization to a continuous-time setting.

1.3 Project Management



	Semester 1												Semester 2											
Tasks	1	2	3	4	5	6	7	8	9	10	11	12	1	2	3	4	5	6	7	8	9	10	11	12
Initial Meeting with Supervisor		🟥																						
Aims and Objectives			🟥																					
Read up on Markowitz's original paper on Portfolio Allocation				🟥																				
Read up on other literature					🟥	🟥																		
Research on optimal control in portfolio optimisation						🟥	🟥																	
Define parameters of model						🟥	🟥	🟥																
First mathematical model of the							🟥	🟥																
Get familiar with manipulating									🟥	🟥														
Test method on datasets										🟥														
Interim Report										🟥	🟥	🟥												
Read up on specific control													🟥	🟥	🟥									
Explore varying control variables														🟥	🟥									
Experiment with incorporating predictions into the model															🟥	🟥								
Modify model to include uncertainty																🟥	🟥							
Read up on Reinforcement Learning (RL)																🟥	🟥	🟥						
Implement RL concepts into the model																	🟥	🟥						
Extend model into a continuous-																			🟥	🟥				
Refine models and organise																				🟥	🟥			
Final Report																					🟥	🟥	🟥	
Oral Presentation																								🟥

In terms of project management, the progress plan from the Gantt Chart above made in week 4 turned out to be pretty accurate to Semester 1 and so the plan was not changed for Semester 2.

During semester 2, the difference between the result for the efficient frontier of the tested data and the expected result was reconciled during the first week and the data simulated using Monte Carlo simulation now shows the expected result. From week 2 to week 4, study on convex optimization was embarked upon and control variables and incorporating predictions were implicit in solving the single period optimization problem. Difficulty arose starting in week 4 during the study of the multi-period models. Both the open-loop and MPC case turned out to be more difficult than expected and the mathematical formulation for that extended well into week 8 before the spring break which pushed back the schedule by 2-3 weeks. Reinforcement learning was incorporated during week 9 and 10. Uncertainty was incorporated implicitly in the MPC and

RL models. The dissertation was also continued during week 9 and was finally completed in week 10.

1.4 Overview of Report

Chapter 2 introduces the mean-variance framework pioneered by Markowitz which is accompanied by a short summary of Moden Portfolio Theory. It concludes with a review of the literature on portfolio optimization.

Chapter 3 explains the basic concepts and terminology used in the modelling of stocks or financial assets using probability theory. Efficiency defined by Markowitz is explained to characterize the performance of holding a particular portfolio of assets.

Chapter 4 explores Modern Portfolio Theory in more depth and used to build up a simple two-asset portfolio, and how the risk-free asset comes into play, and how we generalize to the multi-asset case. An analytical solution for the minimum-variance portfolio and the tangent portfolio is shown for the “Mag 7” stocks at the end.

Chapter 5 discusses portfolio optimization as a convex optimization problem in terms of an objective, and how preferences of the investor such as an expected return, a minimum risk, or real-world constraints such as turnover, trading limits or costs can be formulated as the constraints of the problem. The models are compared assuming perfect estimates of expected returns and covariance matrices.

Chapter 6 extends the multi-period case to incorporate control and feedback from actual market outcomes to form the closed-loop Model Predictive Control model. All the models are compared in a new realistic scenario where estimates of mean and risk are not known beforehand and are estimated with historical data.

Chapter 7 illustrates the formulation of the portfolio optimization with Reinforcement Learning and explains its similarity to state-spaced control and convex optimization.

Chapter 8 concludes with the summary of our findings and gives an evaluation of the different models and the practical concerns when using the models in a realistic setting.

Chapter 2 Literature Review

To see how optimization and control theory are closely related to portfolio optimization, we must first look at contemporary methods of how mathematics is applied to finance, and arguably the most prominent of which is portfolio theory. This concept was pioneered by Harry Markowitz in his doctoral dissertation "Portfolio Selection" in 1952 which revolutionised the way academics and practitioners think about investment management by providing the first rigorous framework of modelling a portfolio of financial assets, now more popularly known as mean-variance analysis or **Modern Portfolio Theory (MPT)**.

The theory centers around the assumption that a rational investor intends to maximise their return while minimising the risk (potential downsides or lost) of their investments. Although it is not without criticism, Warren Buffett, who is widely regarded as the best investor in history has criticized the way of thinking of the risk of a stock as volatility in its price movement (Buffett, 1993) but it is our current best model.

Markowitz (1952) mathematically formalised the idea that diversification reduces portfolio risk by combining assets with low or negative correlations. He developed the concept of the **efficient frontier**, a set of a portfolios that are optimal in terms of risk and return. The idea of a risk-free interest rate from (Fisher, 1930) was used by (Tobin, 1954) to introduce the **Capital Market Line (CML)**, the straight line connecting the risk-free rate to the tangency portfolio. He demonstrated how to achieve an optimal portfolio on the efficient frontier when combining it with a risk-free asset. William Sharpe extended upon Markowitz's and Tobin's work by assuming market equilibrium (Sharpe, 1964), and developed the **Capital Asset Pricing Model (CAPM)** as a model to explain how individual assets are priced in equilibrium relative to the market. The CAPM assumes that idiosyncratic risks – the component of an asset's total risk that is unique to that asset and not explained by the overall market can be diversified away in a well-diversified portfolio. Hence, only the β (beta) – the systematic risk of the individual asset relative to the market should be considered as the risk premium for holding that asset. However, many papers have since criticized the assumption that expected excess return of an asset is only determined by market covariance. Classically, Fama and French (1993) showed that a three-factor model which includes a firm's size (market capitalization) and a value premium (high book-to-market stocks) and later a five-factor model Fama and French (2015) better capture the sources of expected returns rather than just a single factor model of the CAPM. Papers such as Lo and Mackinlay (1988) and Barberis, Shleifer and Vishny (1998) even directly challenge the Efficient-Market Hypothesis which assumes that markets are fully informationally efficient.

Throughout most of the extensive history (Kolm, Tutuncu and Fabozzi, 2014) of the original Markowitz mean-variance problem, algorithms like quadratic programming provided exact optimal solutions. However, as practical constraints and more assets are considered, the computational complexity becomes NP-hard and so approximation techniques must be considered (Kalayci *et al*. 2019). Extensions to the single-period formulation include dynamic models such as those proposed by Calafiore (2008), who presents two multi-period strategies: an open-loop approach, where weights are fixed over the horizon, and a closed-loop or receding horizon strategy, where allocations are updated based on observed returns. The latter approach drawing analogies to feedback mechanisms used in optimal control.

Modern developments demonstrate dynamic decision-making in stochastic processes, setting the stage for portfolio strategies (Li, Uysal and Mulvey, 2022) that integrate principles from control theory to adapt more effectively to uncertainty in financial markets.

Chapter 3 Concepts and Terminology

3.1 Return

We denote a portfolio P with financial assets $A_i, i = 1, 2, \ldots, n$ where n is the number of assets in the portfolio. Using probability theory, the future return R_P of the portfolio over a certain period can be modelled as the expected return:

$$E(R_P) = \sum_{i=1}^{n} X_i \overline{R_\iota} \qquad \text{(Eq. 3.1)}$$

where $\overline{R_\iota}$ is the average return and X_i is the proportion of asset A_i in the portfolio P over the period respectively.

3.2 Risk

According to MPT, risk can be modelled using variance or standard deviation of the return of an asset defined by:

$$Var(R) = E((R - \bar{R})^2),\ or,\ \sigma_R = Var(R)^{\frac{1}{2}} \qquad \text{(Eq. 3.2)}$$

3.3 Efficiency

To start out, two fundamental assumptions are made:

1. According to the CAPM, individual assets are correctly priced based on their risk relative to the market.
2. We can estimate the expected returns and risk of stocks by their historical prices.

Then determining the best portfolio (the portfolio with the optimal weight allocations across the assets) out of all possible portfolios moves from picking the asset that we think will provide the highest future returns, to a conversation of comparing portfolios with varying risk and returns with each other. The following definitions from (Joshi, 2013) can then be laid out.

Definition 1: The set of all possible pairs of returns and standard deviations attainable from investing in a collection of assets is called the opportunity set.

Definition 2: A portfolio is efficient relative to a given opportunity set provided no other portfolio in that opportunity set

1. Has at least as much expected return and lower standard deviation, and
2. Has a higher return and an equal or smaller standard deviation

Definition 3: The subset of the opportunity set which is efficient is called the efficient frontier.

Efficiency is defined relative to the set of investment opportunities, changing the set of assets available to investors also changes the set of efficient portfolios.

Chapter 4 Modelling stocks with Modern Portfolio Theory

4.1 Two asset portfolio

We consider the simple case of the opportunity set consisting of two risky assets A and B, and attempt to construct a relationship between the risk and return of the set of portfolios of these 2 assets.

Assuming investment fractions X_A and X_B such that

$$X_A + X_B = 1$$

and from (Eq. 3.1) we get

$$R_P = X_A R_A + X_B R_B \tag{Eq. 4.1}$$

Then applying the linearity of expectations to (Eq. 4.1):

$$\begin{aligned} E(R_P) &= X_A E(R_A) + (1 - X_A)E(R_B) \\ &= X_A\big(E(R_A) - E(R_B)\big) + E(R_B) \end{aligned} \tag{Eq. 4.2}$$

and applying the property of variance to (Eq. 4.1):

$${\sigma_P}^2 = {X_A}^2{\sigma_A}^2 + (1 - X_A){\sigma_B}^2 + 2X_A(1 - X_A)\sigma_{AB} \tag{Eq. 4.3}$$

We can see then that the expected return is linear in X_A whilst the variance is quadratic and also depend on the correlation between the two assets.

We illustrate the parabola curve in the risk-return space using a numerical example for the assets A and B. Using the parameters for expected returns of 12 and 8, standard deviations of 20 and 15 for A and B respectively, and correlation of 0.3, the efficient frontier for this opportunity set was computed using (Eq. 4.3) with 100 weightings of X_A from 0 to 1 in increments of 0.01. The resulting plot from Python is shown in Fig. 4.1.

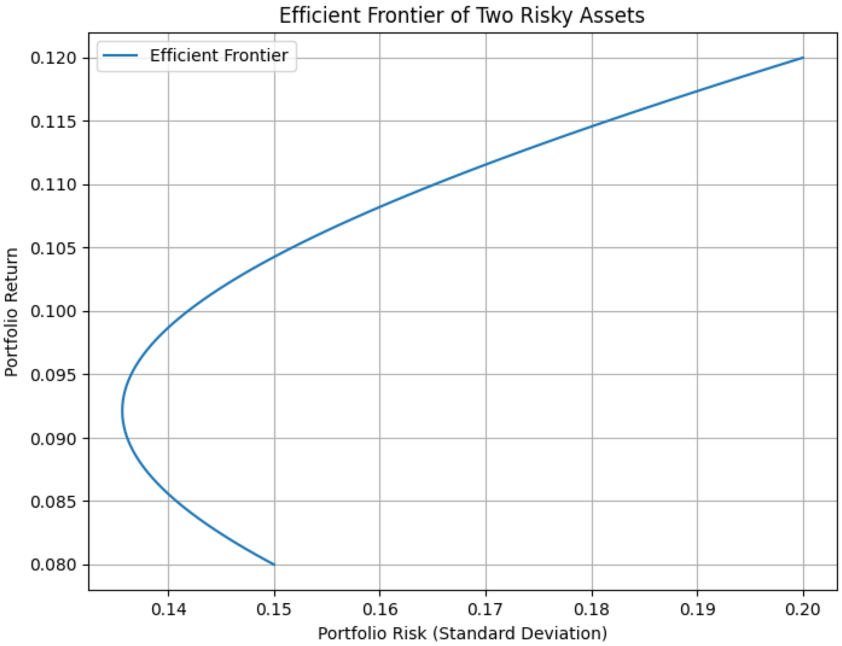


Fig. 4.1: Efficient frontier for two risky assets

4.2 Risk-free asset and the Tangent Portfolio

We introduce the risk-free asset with the definition from (Joshi, 2013) as follows:

> **Definition 4**: An asset whose return is known in advance is said to be risk-free. An asset f, is risk-free if and only if:
>
> 1. The variance of returns is zero
> 2. The standard deviation of returns is zero

Suppose that a portfolio P consists of 1 – y units of the risk-free asset f with return R_f, and y units of the risky asset (or a portfolio of risky assets) A with return R_A, then expected return of P is then

$$\overline{R_P} = (1-y)R_f + y\overline{R_A} \qquad \text{(Eq. 4.4)}$$

applying the property of variance, and since R_f is riskless, the risk of the portfolio would be

$$Var(R_P) = y^2 Var(R_A),$$

$$\sigma_P = |y|\sigma_A$$

Restricting y ≥ 0, we have

$$y = \frac{\sigma_P}{\sigma_A}$$

Substituting y into (Eq. 4.4), we have

$$\overline{R_P} = \frac{\overline{R_A} - R_f}{\sigma_A}\sigma_P + R_f$$

This shows that the new portfolio P, which is combination of a risk-free asset with a risky portfolio A produces a straight line for the opportunity set, which is called the Capital Market Line (Tobin, 1958). The gradient $\frac{\overline{R_A}-R_f}{\sigma_A}$ turns out to be the Sharpe ratio (Sharpe, 1964), which represents the ratio of return per unit of increase in risk that an investor undertakes.

The CML, the entire line through points (0, R_f) and $(\sigma_A, \overline{R_A})$ for a particular portfolio of risky assets and a risk-free asset is efficient. We state two theorems from (Joshi, 2013) omitting the proof.

> Theorem 1: If there is a risk-free asset, all efficient portfolios lie on a straight line in standard deviation/expected return space.

Even after discarding the risk-free asset, investing solely in a portfolio of risky assets A is itself efficient if the new portfolio P is efficient.

> Theorem 2: If P is efficient, then the portfolio A consisting of risky assets in P is efficient relative to investing solely in risky assets.

Reconciling the CML with the opportunity set for a portfolio of risky assets A, we summarize omitting the full proof from (Joshi, 2013) that

1. The efficient set of A is a hyperbola in risk/return space.
2. Combining the risk-free asset R_f with risky assets A produces a new portfolio P that is a straight line through points (0, R_f) and $(\sigma_A, \overline{R_A})$. And this whole line is also efficient.
3. The point of tangency of the efficient line P and the hyperbola efficient set of A is an efficient portfolio called the tangent portfolio.

4.3 Multi-asset portfolio

Generalising to the multi-asset case, A is now a portfolio of risky assets with return $\overline{R_A}$ and standard deviation σ_A given by

$$\overline{R_A} = \langle x, \bar{R} \rangle, and\ \sigma_A = (x^T C x)^{\frac{1}{2}} \qquad \text{(Eq. 4.5)}$$

where x is a vector or portfolio weights, C is the covariance matrix, and $\bar{R}$ is the vector of returns for the underlying assets, and notation $\langle a, b \rangle$ denotes the dot product of vectors a and b.

Let us now illustrate the efficient frontier in solely the risky case when holding a portfolio of 7 stocks. Raw public data of the closing prices for the 7 stocks of Apple, Amazon, Google, Meta, Microsoft, Nvidia and Tesla were analysed from 27th November 2023 to 22nd November 2024. The period was 252 days or roughly equivalent to a full year's worth of trading days.

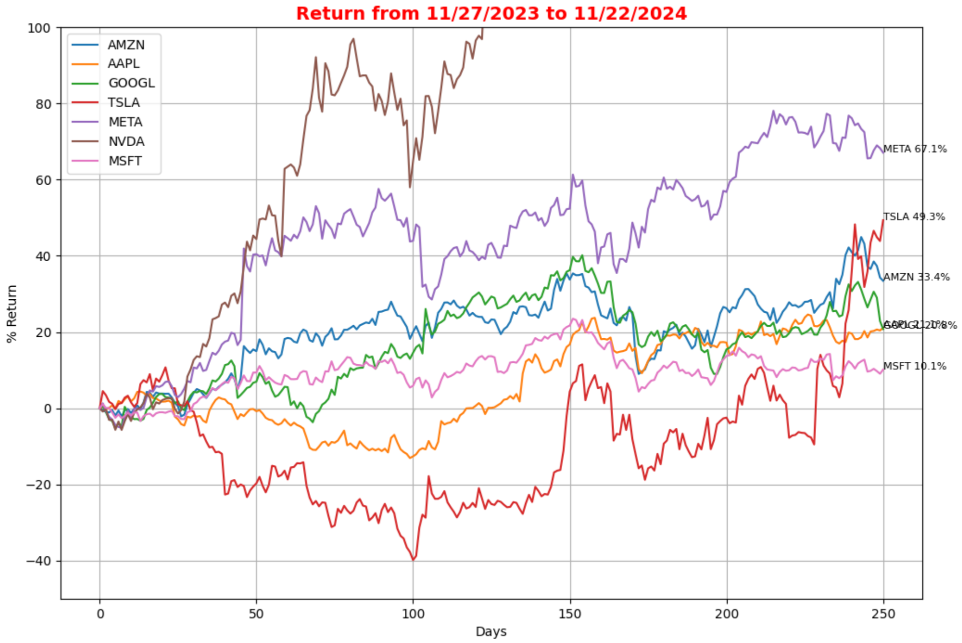


Fig. 4.2: The cumulative returns of the 7 tech stocks for a full year

From the daily closing prices, 3 statistics were computed for each stock. Simple returns from the start to end period, daily percentage returns, and the cumulative returns for the whole period. Cumulative returns were then plotted using Python to produce Fig. 4.2.

Using (Eq. 4.5), we compute expected returns and the standard deviations of the 7 stocks from data of daily returns. The covariance matrix can also be determined analytically from the data but here the *pandas* library from python was used. Then using Monte Carlo simulations for 10,000 random weights and allowing for short selling, we compute the opportunity set and plot it as shown in Fig. 4.3.

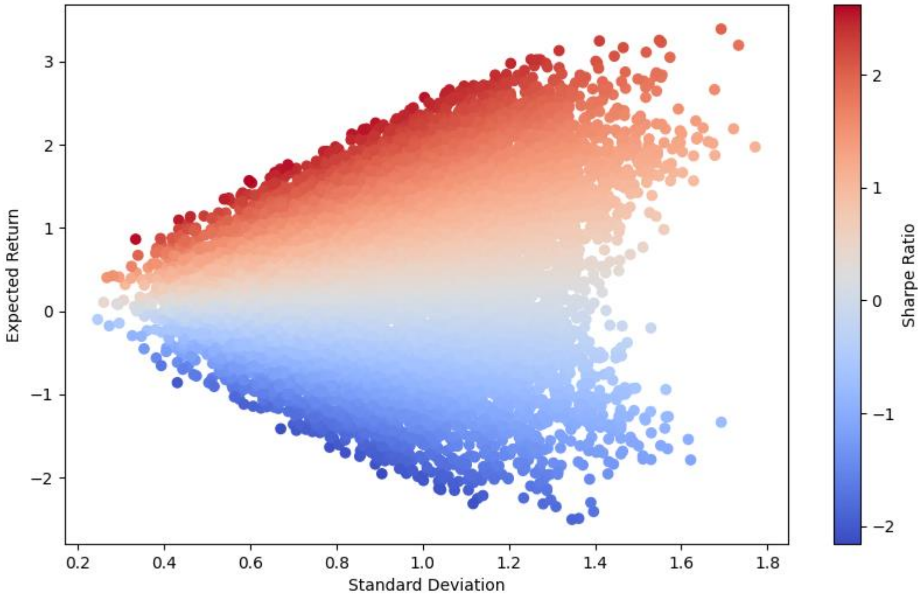


Fig. 4.3: The opportunity set of the “Magnificent 7” stocks including short selling

In practice, an investor would want to take into account the return of their portfolio in relation to the return of the risk-free asset. From section 4.2, we know that the tangent point between the CML and the opportunity set for a portfolio of risky assets gives us the tangent portfolio, the efficient portfolio where all the funds are invested in the risky assets and none in the risk-free asset.

Hence, the problem now reduces to maximizing the slope

$$\theta = \frac{\overline{R_A} - R_f}{\sigma_A}$$
$$= \frac{\langle x, \bar{R} \rangle - R_f}{(x^T C x)^{\frac{1}{2}}}$$

with the constraint $\sum_{i=1}^{n} x_i = 1$.

From (Joshi, 2013), the algorithm for computing the tangent portfolio weights of vector x is:

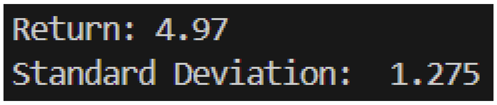
1. Let $\widetilde{R_\imath} = \overline{R_\imath} - R_f$
2. Solve $Cy = \tilde{R}$
3. Set $x_i = \frac{y_i}{\sum_{j=1}^{n} y_j}$

If an investor wants to determine the efficient portfolio with the minimal risk, and hence the minimal variance portfolio (MVP), we can vary the risk-free rate to get lower and lower, the slope of the CML gets steeper and steeper, and the tangent portfolio gets closer to the tip (i.e. the point of minimal variance). Omitting the full proof from (Joshi, 2013), it follows that the weights x of the MVP can be obtained by letting the risk-free rate tend to –∞, and we have

$$x = \frac{C^{-1}e}{\langle C^{-1}e, e\rangle}$$

where e is a vector of ones of size n.

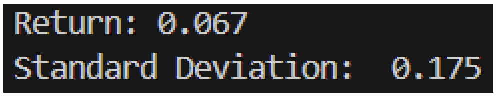
We use the algorithm to compute the tangent portfolio weights, and the equation for the MVP weights to compute the MVP weights. The expected return and standard deviations of the tangent portfolio of the “Magnificent 7” stocks are shown in Fig. 4.4 below.



```
Return: 4.97
Standard Deviation:  1.275
```

Fig. 4.4

while for the MVP they are shown in Fig. 4.5 below.



```
Return: 0.067
Standard Deviation:  0.175
```

Fig. 4.5

Assuming the November 2024 1-month Treasury Rate of 4.72% as the theoretical risk-free asset, the new efficient line of portfolios including the tangent portfolio produces a plot in Fig. 4.6.

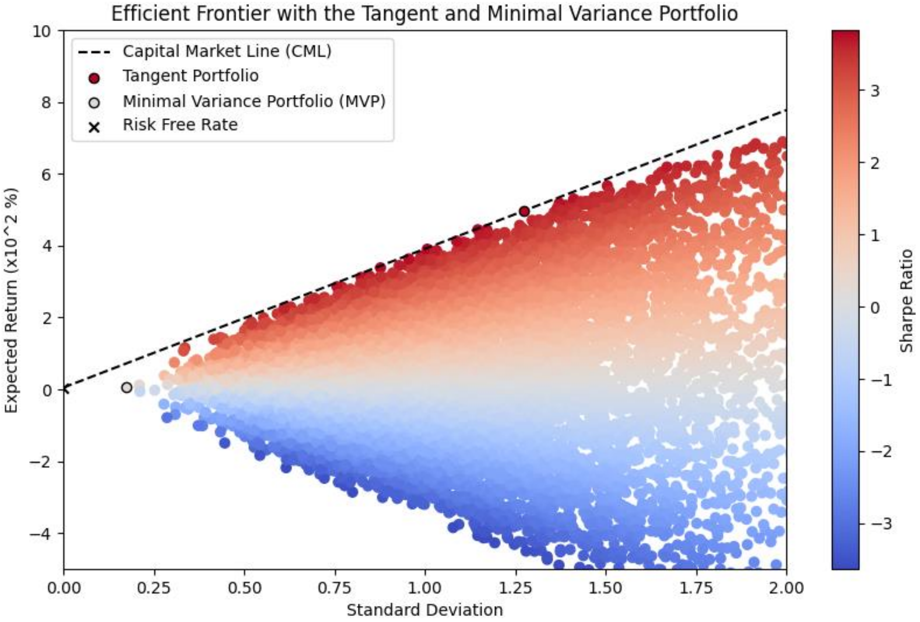


Fig. 4.6: The tangent portfolio and MVP of the 7 tech stocks including short selling

The CML crosses the opportunity set of risky stocks A at the tangent portfolio with inefficient portfolios lying under the curve. We note here that because the overall return of the "Magnificent 7" stocks during this period were so high, the risk-free rate of 0.0472 (4.72%) is almost insignificant even if a slightly different risk-free rate were used.

Chapter 5 Portfolio Optimization with Convex Optimization

5.1 Single Period Optimization (SPO)

We have found an analytical solution for the portfolio weights x_i for the MVP and the Tangent Portfolio. To find the optimal weights x_i given any minimum return or maximum risk for the portfolio, we can extend the problem to a convex optimisation problem where the requirements of the portfolio are formulated as the objective function and its constraints.

The two classic forms are:

1. Minimize risk (variance) given an expected return (mean)

$$\min_x x^T \Sigma x$$

$$s.t. \mu^T x \geq R,$$

$$1^T x = 1; x \geq 0$$

where $x \in \mathbb{R}^n$ is a weight vector of n assets, $\mu \in \mathbb{R}^n$ is the expected return vector and R is the minimum return.

2. Maximize expected return given a risk constraint

$$\max_x \mu^T x$$

$$s.t. x^T \Sigma x \leq \sigma_{max}^2,$$

$$1^T x = 1; x \geq 0$$

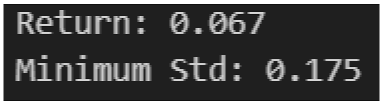
where σ_{max}^2 is the risk threshold.

As an illustration, the MVP can be formulated as a quadratic convex optimisation problem where the objective function is the variance of the portfolio, and the portfolio weights $\sum x_i = 1$ as an equality constraint. Formally:

$$min\ x^T \Sigma x$$

$$s.t. \mathbf{1}^T x = 1$$

A convex optimisation problem can be solved reliably using solvers. We use an open-source Python solver CVXPY which makes solving convex optimisation problems in standard form easy and straightforward. The results that we obtain are in Fig. 5.1 below:



```
Return: 0.067
Minimum Std: 0.175
```

Fig. 5.1

As expected, it gives us the same solution as the analytical solution in Fig. 4.5.

A few other formulations of SPO are as follows:

Mean-variance utility maximization – combines return and risk into a single objective function

$$\max_{x} \mu^T x - \gamma x^T \Sigma x$$

where γ is the risk aversion coefficient.

Robust portfolio optimization – addresses uncertainty in the estimates of mean and covariance

$$\max_{x} \min_{\mu, \Sigma \in \mathcal{U}} \mu^T x - \gamma x^T \Sigma x$$

where the problem is formulated as a worst-case optimization over an uncertainty set $\mathcal{U}$.

Conditional Value-at-Risk (CVaR) based optimization – focusing on downside risk

$$\max_{x} E[R(x)]$$

$$s.t. CVAR_{\alpha}(x) \leq \tau$$

$$\mathbf{1}^T x = 1, x \in \mathcal{W}$$

where $\mathcal{W}$ is the feasible set, τ is the risk threshold, $CVAR_{\alpha}(x)$ is the expected loss beyond a given quantile α.

To maintain tractability and simplicity, we will not consider the above cases and only consider linear and quadratic programs.

5.2 Multi Period Optimization (MPO)

A natural extension to SPO would be to consider multiple smaller periods within the overall period. The overall period is called the investment horizon, and the smaller periods are split between that investment horizon. For example, we can define the investment horizon in which we are considering to be the annual year. Then, we can define the period k to be a quarter (3 months) of a year. The investment horizon is then divided equally among $k = [1, 4]$ where $k - 1$ would be the start of the kth period and k would be the end of the kth period within the horizon.

We layout the problem and define a few notations from Calafiore and El Ghaoui (2014). The simple return of an investment in asset i over the k-th period from $(k-1)\Delta$ to $k\Delta$ is:

$$r_i(k) \doteq \frac{p_i(k) - p_i(k-1)}{p_i(k-1)}, i = 1, \dots, n; k = 1,2, \dots,$$

$$g_i(k) \doteq 1 + r_i(k), i = 1, \dots, n; k = 1,2, \dots,$$

where $g_i(k)$ is the corresponding gain.

Then,

$$\begin{aligned} g(k) &= 1 + r(k) \\ &= \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} + \begin{bmatrix} r_1(k) \\ \vdots \\ r_n(k) \end{bmatrix}, \\ G(k) &= diag\big(g(k)\big) \\ &= \begin{bmatrix} 1 + r_1(k) & \cdots & 1 \\ \vdots & \ddots & \vdots \\ 1 & \cdots & 1 + r_n(k) \end{bmatrix}, \\ \Phi(v,k) &\doteq G(k)G(k-1)\cdots G(v), \\ \Phi(k,k) &\doteq G(k) \end{aligned}$$

where $g(k)$ denotes the vector of gains of the assets over the k-th period, $r(k)$ being the vector of the assets' returns, $G(k)$ is the diagonal matrix of the elements of $g(k)$, and $\Phi(v,k)$ is the compounded gain matrix from the beginning of period v to the end of period k where $v \leq k$.

The investor's total wealth at time k is

$$w(k) = \sum_{i=1}^{n} x_i(k) = \mathbf{1}^T x(k) \qquad \text{(Eq. 5.1)}$$

where $x(k) = [x_1(k) \quad \cdots \quad x_n(k)]^T$ is the vector of portfolio weights at time k.

The random portfolio composition at time $k = 1, \dots, T$ is

$$x(k) = \Phi(1,k)x(0) + \sum_{j=1}^{k} \Phi(j,k)u(j-1) \qquad \text{(Eq. 5.2)}$$

where $u(j-1)$ is the portfolio inputs for assets $i = 1, \dots, n$ at the beginning of period $k = j - 1$.

From (Eq. 5.1), the total wealth at time k can be written compactly as

$$w(k) = 1^T x(k) = \phi(1,k)x(0) + w_k^T u \qquad \text{(Eq. 5.3)}$$

where $\phi(v,k) = 1^T \Phi(v,k)$ is the row vector of the compounded gain matrix $\Phi(v,k)$, and $w_k^T = [\phi(1,k) \quad \cdots \quad \phi(k-1,k) \quad \phi(k,k)]$ is the row vector of all the compounded gain matrices of period 1 to period k.

Incorporating $x(0)$ into the dynamics of the portfolio input u, we can simplify (Eq. 5.3) to

$$w(T) = {w_T}^T u \qquad \text{(Eq. 5.4)}$$

where

$$u = [w(0) \quad u(1) \quad \cdots \quad u(T-1)]^T \in \mathbb{R}^{nT},$$
$${w_T}^T = [\phi(1,T) \quad \phi(2,T) \quad \cdots \quad \phi(T,T)] \in \mathbb{R}^{1\times nT}$$

The objective function is defined as

$$J(T) = \sum_{k=1}^{T} \mathrm{Var}\big(\mathrm{w(k)}\big)$$

$$s.t. \sum_{i=1}^{n} w_i(0) = 1\,,$$

$$\sum_{i=1}^{n} u_i(k) = 0\,; k = 1, \dots, T-1$$

where $w(0) = 1$ is the fully invested constraint at the start of the investment horizon and $u(k) = 0$ is the self-financing constraint at each timestep k, and n is the number of assets.

The multi-period formulation can then be written as a single period minimization problem as

$$\min_{w(0),u(1),\dots,u(T-1)} Var\left(w(T)\right)$$

Then

$$\begin{aligned} & Var\big(w(T)\big) \\ & = Var({w_T}^T u) \\ & = u^T Cov(w_T) u \\ & = u^T Q u \end{aligned}$$

where $Q = Cov(w_T)$.

Using historical covariances as point estimates for each period $k = 1, 2 \dots, T$ and assuming that the probability distribution of asset returns between each period as independent and identically distributed (IID) random variables, the risk matrix Q can be estimated as a block-diagonal matrix of each period's covariance matrix with zero correlation with each other:

$$Q = \begin{bmatrix} \Sigma_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \Sigma_k \end{bmatrix}$$

where $\Sigma_k = Cov\big(g(k)\big)$ us the covariance matrix of gains at time k.

To summarize, we choose rebalancing vectors $w(0), u(1), \dots, u(T-1)$ that minimise the variance of terminal wealth $w(T) = {w_T}^T u$. With IID and zero serial correlation between periods assumptions, the risk matrix Q is a block diagonal matrix of each period's covariance matrix. The optimization problem is:

$$\min_{w(0),u(1),\dots,u(T-1)} u^T Q u \qquad \text{(Eq. 5.5)}$$

$$\begin{aligned} s.t.\, \mathbf{1}^T w(0) &= 1, \\ \mathbf{1}^T u(k) &= 0, k = 1, \dots, T-1, \\ Q &= Cov(w_T) \in \mathbb{R}_+^{nT \times nT}, \end{aligned}$$

$$w(0), u(k) \in \mathbb{R}^n; n = number\ of\ asset; T = number\ of\ periods$$

This is a quadratic convex optimization program that can be solved using standard convex optimization solvers.

Again, we illustrate the multi-period case with the MVP using $T = 2$.

From (Eq. 5.5), the problem is

$$\underset{w(0),u(1)}{min} \begin{bmatrix} w(0) \\ u(1) \end{bmatrix}^T Q \begin{bmatrix} w(0) \\ u(1) \end{bmatrix}$$

$$s.t. \mathbf{1}^T w(0) = 1,$$

$$\mathbf{1}^T u(1) = 0,$$

$$Q = \begin{bmatrix} \Sigma_1 & 0 \\ 0 & \Sigma_2 \end{bmatrix}, \Sigma_1 = Cov(R_1), \Sigma_2 = Cov(R_2), R_k = daily\ simple\ returns\ in\ period\ k$$

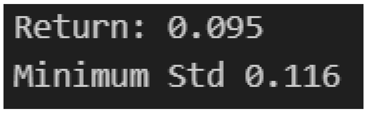
From (Eq. 5.4), we compute terminal wealth as

$$w(T) = \phi(1{,}2)w(0) + \phi(2{,}2)u(1),$$

$$\phi(1{,}2) = a_1 \odot a_2, \phi(2{,}2) = a_2,$$

where a_1 and a_2 are the vectors of simple returns over the period 1 and period 2 respectively, and $\odot$ is the element-wise multiplication operator.

The obtain results in Fig. 5.2 below:



```
Return: 0.095
Minimum Std 0.116
```

Fig. 5.2

With a lower risk of 11.6% and a return of 9.5%, it performs better than the single period in both metrics of risk of 17.5% and return of 6.7%. The returns between the SPO model and the MPO model over the investment horizon is shown in Fig. 5.3 below.

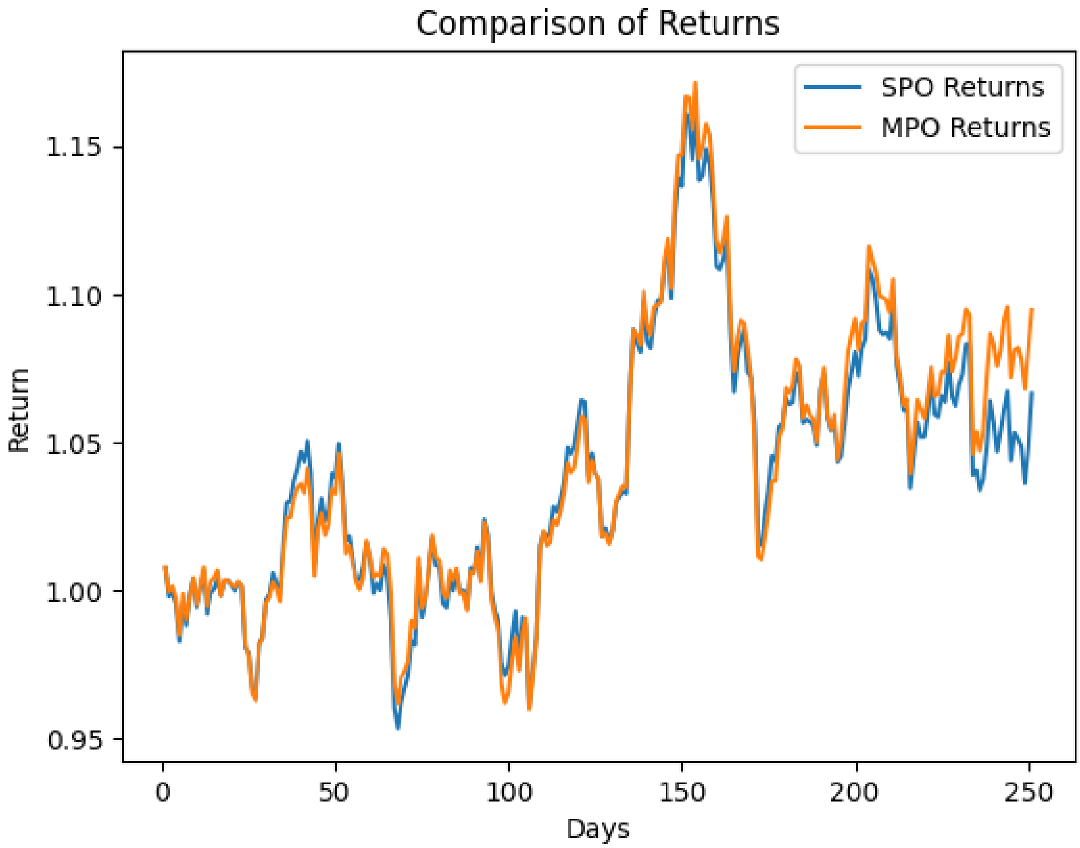


Fig. 5.3: Comparison of returns between the SPO and MPO models

5.3 Effect of length of periods in multi-period models

To make a comparison between multi-period models with shorter and longer periods within the same investment horizon, we compare $T = 2$ and $T = 4$ periods simulating two half periods and 4 quarter periods over a year respectively. And to force the control inputs $u(k) = [1, T - 1]$ to make trades, we add an expected return of 55.26% from the equal-weighted portfolio as the benchmark over the investment horizon. Assuming the same self-financing constraints as (Eq. 5.5), the only difference is added the target expected return formulated as the compounded terminal simple return across the periods:

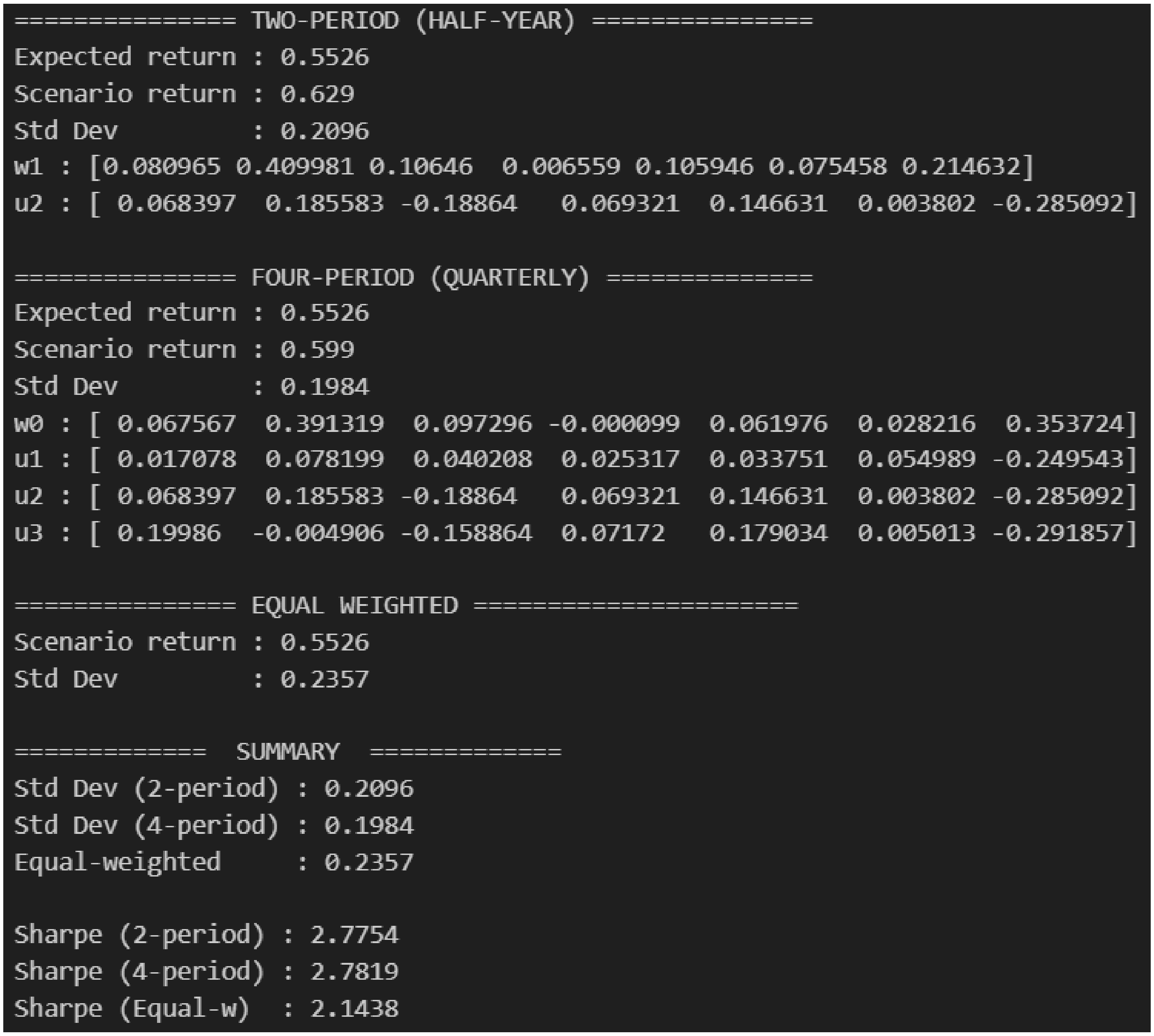
$$R^{term} = \sum_{k=0}^{T-1} \mu(k:)^T u(k)$$

where the cumulative mean after period k is $\mu(k:) := \sum_{j=k}^{T-1} \mu(j)$. Here, $R^{term} = 55.26\%$.

The optimization problem becomes

$$\begin{aligned} &\min_{u \in \mathbb{R}^{nT}} u^T Q u && \text{(Eq. 5.6)} \\ &s.t. \mathbf{1}^T w(0) = 1, \\ &\mathbf{1}^T u(k) = 0, k = 0, \dots, T-1, \\ &(\mu(0:)^T, \mu(1:)^T, \dots, \mu(T-1:)^T) u = \rho;\ \rho = 0.5526 \end{aligned}$$

Solving the problem across the full year data of the "Mag 7" stocks, we get the results as shown in Fig. 5.4 below.

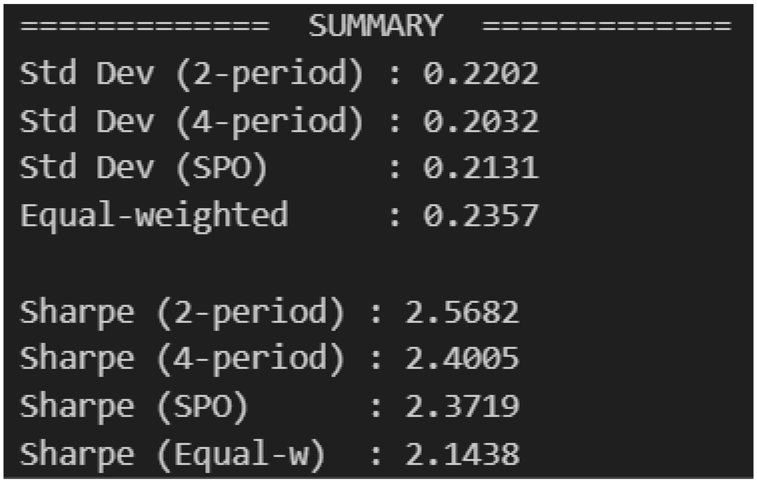


```
=============== TWO-PERIOD (HALF-YEAR) ===============
Expected return : 0.5526
Scenario return : 0.629
Std Dev         : 0.2096
w1 : [0.080965 0.409981 0.10646  0.006559 0.105946 0.075458 0.214632]
u2 : [ 0.068397  0.185583 -0.18864   0.069321  0.146631  0.003802 -0.285092]

=============== FOUR-PERIOD (QUARTERLY) ==============
Expected return : 0.5526
Scenario return : 0.599
Std Dev         : 0.1984
w0 : [ 0.067567  0.391319  0.097296 -0.000099  0.061976  0.028216  0.353724]
u1 : [ 0.017078  0.078199  0.040208  0.025317  0.033751  0.054989 -0.249543]
u2 : [ 0.068397  0.185583 -0.18864   0.069321  0.146631  0.003802 -0.285092]
u3 : [ 0.19986  -0.004906 -0.158864  0.07172   0.179034  0.005013 -0.291857]

=============== EQUAL WEIGHTED ======================
Scenario return : 0.5526
Std Dev         : 0.2357

============  SUMMARY  ============
Std Dev (2-period) : 0.2096
Std Dev (4-period) : 0.1984
Equal-weighted     : 0.2357

Sharpe (2-period) : 2.7754
Sharpe (4-period) : 2.7819
Sharpe (Equal-w)  : 2.1438
```

Fig. 5.4: Performance metrics of MPO models with perfect quarter estimates

The results show that the 2-period and 4-period MPO models both outperform the simple equal-weighted strategy in terms of Std. Dev. (risk) given an expected return using this specific scenario. The 4-period MPO model gives the lowest risk (19.84%) out of all the models. It should be noted here that the covariance estimates and mean vectors for each period are perfect as we are carrying out the experiment in the scenario using the same historical data as the period estimates. In this case, $\mu(0), \mu(1), \mu(2), \mu(k-1)$ and $\Sigma_0, \Sigma_1, \Sigma_2, \Sigma_{k-1}$ are known exactly for periods $k = [0, T-1]$ where $T = 4$ is the number of quarters in the year.

If say we only have a perfect estimate of the annual return μ and covariance matrix Σ, we compare the MPO models with the annual estimate scaled to the quarter, and we add in the SPO model with the annual estimate for comparison. Using the simple equal-weighted portfolio across the horizon as the benchmark, we obtain results as Fig. 5.5 below.



```
=============  SUMMARY  =============
Std Dev (2-period) : 0.2202
Std Dev (4-period) : 0.2032
Std Dev (SPO)      : 0.2131
Equal-weighted     : 0.2357

Sharpe (2-period) : 2.5682
Sharpe (4-period) : 2.4005
Sharpe (SPO)      : 2.3719
Sharpe (Equal-w)  : 2.1438
```

Fig. 5.5: Performance metrics of optimization models using annual estimates

Now, the 4-period MPO model still performs better than the 2-period MPO model looking purely at Std. Dev. with a value of 20.32% compared to 22.02%. However, on a Sharpe ratio basis which also considers the actual realized returns, the former performs worse with a value of 2.40 compared to 2.57 of the latter. The SPO model has a Std Dev. of 21.31% which is in the middle of both MPO models, but is outperformed by both the MPO models in terms of Sharpe ratio with a value of 2.37. All models perform better than the simple equal-weight strategy.

In practice however, we can never predict even the annual estimates accurately, the focus here is only to illustrate the comparison between models given good (in this case: perfect) estimates of the return μ and covariance matrix Σ of the period given everything else equal.

Within MPO models, it is not even clear whether having more periods within the investment horizon improves the model, as the 4-period model performed better in this scenario in terms of Std. Dev. but performed worse in terms of Sharpe ratio which considers realized returns. This shows that in practice, estimates of μ and Σ is by far the most dominant factor and is the deciding factor in how optimization models perform. A single good estimate across the horizon often works better in reality than trying to solve for the problem with inaccurate estimates for multiple periods.

A realistic scenario with practical considerations is shown in the next chapter.

Chapter 6 Model Predictive Control and practical concerns

The MPO problems we have solved so far is the open-loop case, where the portfolio weights are derived for every period and are followed through regardless of what the actual market outcomes were from previous periods. This usually produces suboptimal or even poor performance for the rest of the investment horizon. We modify the open-loop strategy to a closed-loop solution using Model Predictive Control (MPC) from control theory or also known as the receding horizon approach. At each time step, it predicts future asset returns and risks over the investment horizon and solves an optimal sequence of trades over the planning horizon but only executes the first decision. At the next time step, it updates its covariance and mean forecasts, takes into account evolving constraints like turnover and drawdown, then repeats the process again. This adapts to the market as new information becomes available.

The models so far have been tested on a single scenario using historical data from the same sample. This was useful to illustrate the differences between models given perfect estimates of future outcomes however past estimates often do not reflect the future. To make a fair comparison, we divide our yearly data on the "Mag 7" stocks into 4 quarters, where μ and Σ estimates are used from the first quarter Q1, and the optimization models are tested on unseen data from the remaining 3 quarters denoted as Q2, Q3, Q4.

We test all three models discussed so far: Single period optimization (SPO), Open-loop multi-period optimization (MPO) and now the Model Predictive Control model (MPC) over the same investment horizon of 3 quarters. As a benchmark, we solve the optimization problems with an expected minimum return of 32.21% which is the return from a simple equal-weighted portfolio over the 3 quarters. The MPC strategy is outlined below:

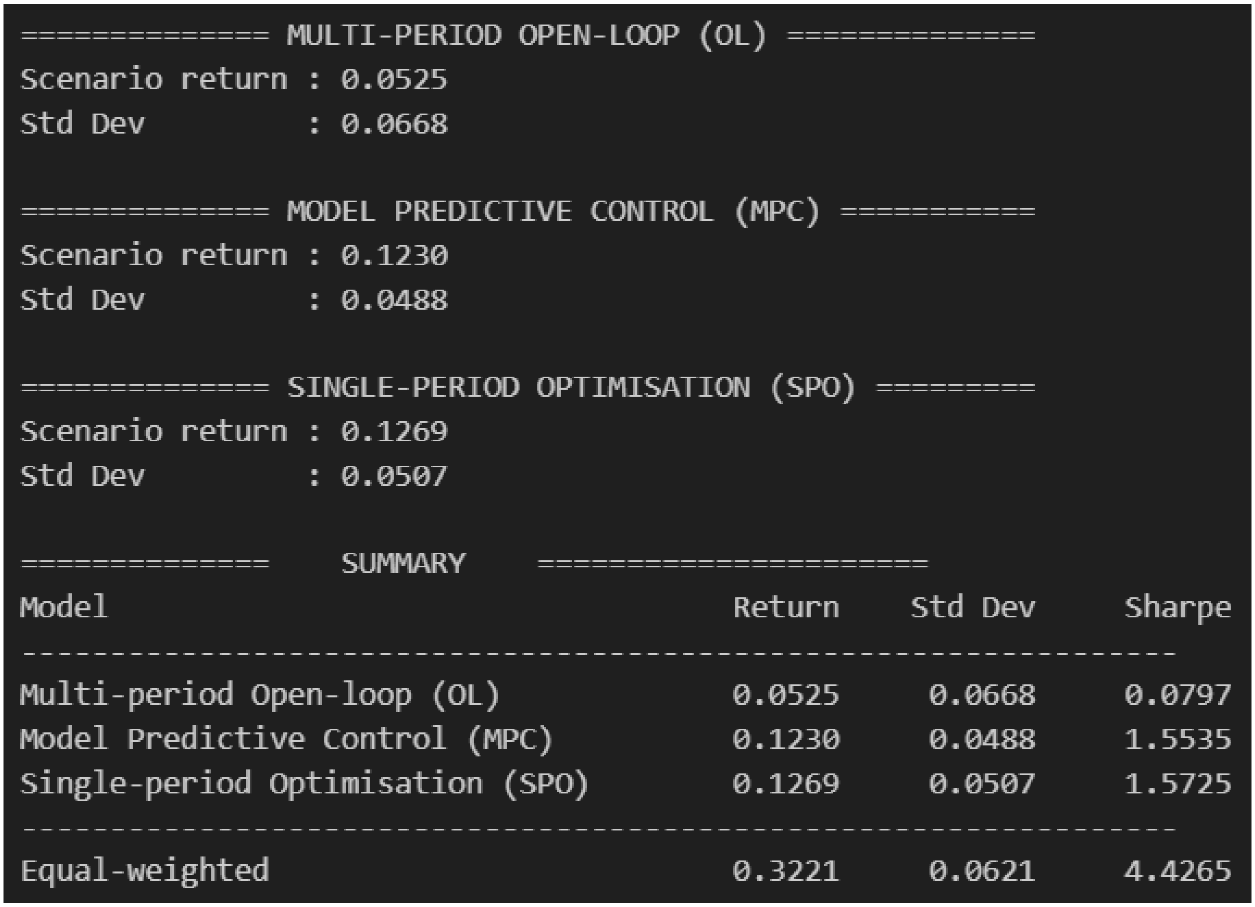
1) Solve the multi-period optimization problem over quarters Q2, Q3, Q4 using the block diagonal covariance matrix Q and mean vector of returns μ from Q1 scaled to 3 quarters as estimates.
2) Only execute the first solved weights for Q2.

3) At the end of Q2, calculate actual realised returns and re-solve the MPO problem for Q3 using updated estimates from Q2, execute the weights for the first period.
4) Repeat at the start of Q4.
5) Calculate end of horizon realised returns and risks.

The MPC is mathematically equivalent to (Eq. 5.6) with the only difference being the model is solved at the again every period t using updated μ estimates from the previous period. We write this as

$$\begin{aligned} \min_{u(t)\in\mathbb{R}^{n(T-t)}} \; & u(t)^T Q(t) u(t) \qquad \text{(Eq. 5.6)} \\ s.t. \; & \mathbf{1}^T w(t) = 1, \\ & \mathbf{1}^T u(k) = 0, k = t+1, \dots, T-1, \\ & (\mu(t:)^T, \mu(t+1:)^T, \dots, \mu(T-1:)^T) u(t) = \rho_t \end{aligned}$$

where the target return $\rho_t = 0.3221$



```
============= MULTI-PERIOD OPEN-LOOP (OL) =============
Scenario return : 0.0525
Std Dev         : 0.0668

============= MODEL PREDICTIVE CONTROL (MPC) ==========
Scenario return : 0.1230
Std Dev         : 0.0488

============= SINGLE-PERIOD OPTIMISATION (SPO) ========
Scenario return : 0.1269
Std Dev         : 0.0507

=============    SUMMARY    ======================
Model                                  Return    Std Dev     Sharpe
-----------------------------------------------------------------
Multi-period Open-loop (OL)            0.0525     0.0668     0.0797
Model Predictive Control (MPC)         0.1230     0.0488     1.5535
Single-period Optimisation (SPO)       0.1269     0.0507     1.5725
-----------------------------------------------------------------
Equal-weighted                         0.3221     0.0621     4.4265
```

Fig. 6.1: Performance metrics between models in a practical scenario

Fig. 6.1 above shows the results for the 3 models with the equal-weighted portfolio as the benchmark. All 3 models perform poorly compared to the equal-weighted portfolio. As expected, accurate estimates of future returns play a significant role in mean-variance frameworks.

Leaving that aside, we compare only the 3 models that used the same estimates. The multi-period open-loop model (MPO-OL) performed the worse with the highest risk at 6.68% and a mere return of 5.25%. MPO-OL optimizes the portfolio weights over all the periods at the start of the investment horizon and executes the remaining trades regardless of new market information or outcomes. The MPC and SPO models are comparable with the MPC model having slightly lower std. at 4.88% compared to 5.07% of the SPO model but the SPO model beating out the MPC model slightly with a return of 12.69% compared to 12.30%. In this case, the simple SPO model shows how it often performs better in reality compared to advance models such as the MPO models which often require accurate estimates over multiple periods to perform well. However, the MPC showcases how adapting to new market information from each quarter helped the model produce better results in the end by optimize the trade decisions accordingly.

For a practical implementation of how MPC works in trading, Boyd *et al*. (2017) and Li, Uysal and Mulvey (2022) show how real-world concerns like transaction costs and holding costs are incorporated in the model and how constraints like drawdown and turnover limits affect the problem.

Chapter 7 Reinforcement Learning

Reinforcement Learning (RL) forms one of the paradigms of machine learning and is suitable to be applied in portfolio optimization due to its core as a sequential decision-making framework under uncertainty. Compared to convex optimization models that optimize a static objective based on fixed inputs, RL models learn optimal investment strategies through interaction with its environment over time.

Castro, Tamar and Mannor (2012) shows an RL algorithm involving minimizing risk with a policy gradient approach in Markov Decision Processes. Prashanth and Ghavamzadeh (2014) illustrates actor-critic algorithms for risk-sensitive RL. Here, we develop a simple, online, deterministic policy that greedily minimizes a mean-variance loss using gradient descent.

A traditional RL framework is characterized by

1. States – represent the information available at a given time.
2. Actions – what the RL agent does to react to the state
3. Objective function – the objective that enables the agent to learn the policy that maximizes the cumulative expected reward over time. This is analogous to the objective function in standard optimization problems
4. Policy – a rule that maps states to actions such as a SoftMax or greedy policy in standard RL problems
5. Rewards – quantifies the quality of an action taken in a particular state

We now illustrate the formulation of portfolio optimization with RL using the “Mag 7” data similar to the previous chapter using the last 3 quarters of the year: Q2, Q3, Q4. Our goal is to learn the portfolio weights w for the beginning of each quarter that achieves the target return while keeping the portfolio variance low.

It does this through minimizing the mean-variance objective function that we define as the **loss function**:

$$L_t(w_t) = \frac{1}{2}({\mu_t}^T w_t - r^*)^2 + \frac{\lambda}{2} {w_t}^T \Sigma_t w_t$$

where r^* is the **target** return each day, the first term penalizing deviation from r^*, and the second term penalizing risk.

The **policy** $\boldsymbol{\pi(s_t)} \rightarrow a_t$ in which the RL agent uses to decide the action based on the state s_t is the online learning rule

$$\pi(s_t) = w_{t+1} = Normalize\big(w_t - \eta\nabla L_t(w_t)\big),$$

$$\nabla L_t(w_t) = ({\mu_t}^T w_t - r^*)\mu_t + \lambda\Sigma_t w_t$$

where $\nabla L_t(w_t))$ is the gradient descent of the loss function, η is the learning rate, and we normalize the update as the self-financing constraint on the portfolio weights. The mathematical convenience of having $^1/_2$ as the coefficient in the two terms of the loss function can also be seen when deriving the policy gradient update rule.

The portfolio weights at day 1 of Q2 are initialised to 0. At each timestep (day), the RL agent takes an action $a_t \in$ A

$$a_t = w_{t+1} - w_t$$

to update our weight estimates based on the covariance matrix and mean return vectors from the previous day. The RL agent then observes the reward R_t at time t to evaluate its action in that state s_t.

We define the **reward** $\boldsymbol{R_t}$ as the return minus the risk penalty from the previous day

$$R_t(w_t) = {\mu_t}^T w_t - \lambda {w_t}^T \Sigma_t w_t,$$
$$\mu_t = r_{t-1} - 1;\ \mu_t \in \mathbb{R}^n,$$
$$\Sigma_t = \mu_t {\mu_t}^T; \Sigma_t \in \mathbb{R}^{n\times n}$$

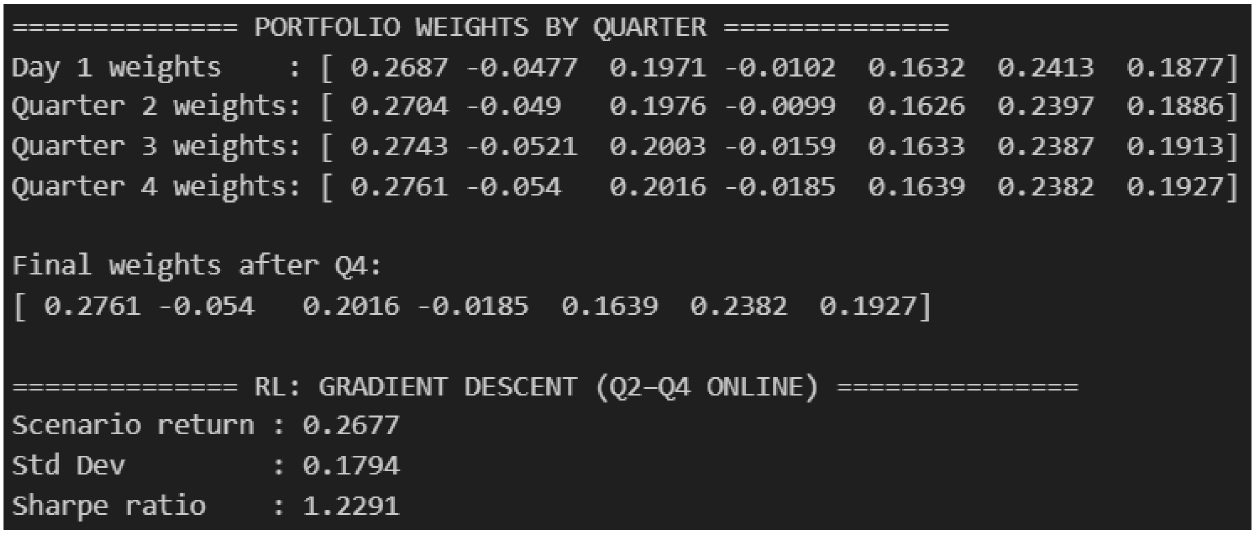
where μ_t is the return from the previous day and Σ_t is the crude 1-day covariance estimate from the previous day.

The **State space** $\boldsymbol{s_t} \in S$ from which our agent considers when taking an action is defined as

$$s_t = (w_t, r_{t-1})$$

where w_t is the portfolio weights at time t, r_{t-1} is the realized returns at time $t-1$ (used to estimate μ_t and Σ_t.

Setting $\lambda = 1$ for simplicity, $\eta = 0.1$ and the scaled daily return ${}^{32.21\%}\!/_{63} = \ 0.51\%$ of the equal-weighted portfolio as the daily target benchmark, we obtain results as in Fig. 7.1 below.



```
============= PORTFOLIO WEIGHTS BY QUARTER =============
Day 1 weights    : [ 0.2687 -0.0477  0.1971 -0.0102  0.1632  0.2413  0.1877]
Quarter 2 weights: [ 0.2704 -0.049   0.1976 -0.0099  0.1626  0.2397  0.1886]
Quarter 3 weights: [ 0.2743 -0.0521  0.2003 -0.0159  0.1633  0.2387  0.1913]
Quarter 4 weights: [ 0.2761 -0.054   0.2016 -0.0185  0.1639  0.2382  0.1927]

Final weights after Q4:
[ 0.2761 -0.054   0.2016 -0.0185  0.1639  0.2382  0.1927]

============= RL: GRADIENT DESCENT (Q2–Q4 ONLINE) ===============
Scenario return : 0.2677
Std Dev         : 0.1794
Sharpe ratio    : 1.2291
```

Fig. 7.1: Performance of RL model

It performs better than all 3 previous models: SPO, MPO, and MPC with a simple return across the three quarters of 26.77%. However, on a Sharpe ratio basis it performs worse than either the MPC or SPO model with a value of 1.23 compared to 1.55 and 1.57 of the 2 models respectively. And on a Std. Dev. basis it performs the worse out of all models with a value of 17.94%, compared with 5.07%, 6.68%, 4.88% for the SPO, MPO, and MPC models respectively as shown in Table 8.1 below.

Strategy	**Return (%)**	**Std Dev (%)**	**Sharpe Ratio**
SPO	12.69	5.07	1.57
MPO-OL	5.25	6.68	0.08
MPC	12.30	4.88	1.55
RL Online	26.77	17.94	1.23

Equal-weight	32.21	6.21	4.43

Table 8.1: Summary of performance of all 4 models against the benchmark EW

The equal-weighted portfolio outperforms all optimization models by far in this scenario. This highlights the drawdown of optimization models in practice as the return μ and covariance matrix Σ estimates are the deciding factor in performance and historical estimates would usually result in poor out of sample performance.

Chapter 8 Conclusions

The mathematical formulation of portfolio theory and portfolio optimization were derived from first principles. Different optimization models were developed and implemented using real data of the “Mag 7” stocks.

Optimization models worked better than the simple strategies such as holding an equal-weight portfolio provided that estimates of future returns and risks are accurate. In reality, historical estimates are often inaccurate and chapter 6 showed what the Markowitz mean-variance optimization framework could look like in practice.

Still, optimization models incorporating control such as the MPC model in Chapter 6 or the RL model in Chapter 7 showed more promising results and perhaps can be improved upon with better tuning of parameters, or better constraints to leverage evolving data at each control period.

Further experiments including real-world constraints such as trading costs or holding limits would make it more realistic and give a better indication of performance of the different models in practice. The models in this paper were only tested on a single scenario, including stochastic programming into the formulation of the problem and stochastic scenarios in the backtests would also show more realistic results of the models.

As estimation is the dominant factor in the mean-variance framework, estimation models of covariance and mean estimates incorporated in the strategies could further expand our understanding of the problem. Formulating portfolio optimization with different models would also be interesting, perhaps in a factor-based model such as the Fama and French (1993, 2015) models.

References

Barberis, N., Shleifer, A. and Vishny, R. (1998) ‘A Model of Investor Sentiment’, *Journal of Financial Economics*, 49(3), pp. 307-343. doi: https://doi.org/10.1016/S0304-405X(98)00027-0

Boyd, S. *et al.* (2017) ‘Multi-period Trading via Convex Optimization’, *Foundations and Trends in Optimization*, 3(1), pp. 1-76.

Buffett, W. (1993) *Warren Buffett’s Annual Letters to Berkshire Shareholders.* Available at: https://www.berkshirehathaway.com/letters/1993.html (Accessed: 6th December 2024).

Calafiore, G.C. and El Ghaoui, L. (2014) *Optimization Models*. Cambridge: Cambridge University Press.

Calafiore, G.C. (2008) ‘Multi-period Portfolio Optimization with Linear Control Policies’, *Automatica*, 44(10), pp. 2463-2473. doi: https://doi.org/10.1016/j.automatica.2008.02.007

Castro, D.D., Tamar, A. and Mannor, S. (2012) ‘Policy Gradients with Variance Related Risk Criteria’, *arXiv preprint arXiv:1206.6404*.

Fama, E.F., and French, K.R. (1993) ‘Common Risk Factors in the Returns on Stocks and Bonds’, *Journal of Financial Economics*, 33(1), pp. 3-56. doi: https://doi.org/10.1016/0304-405X(93)90023-5

Fama, E.F., and French, K.R. (2015) ‘A Five-factor Asset Pricing Model’, *Journal of Financial Economics*, 116(1), pp. 1-22. doi: https://doi.org/10.1016/j.jfineco.2014.10.010

Fisher, I. (1930) *The Theory of Interest, as determined by Impatience to Spend Income and Opportunity to Invest it*. N.Y.: Macmillan.

Joshi, M. S. and Paterson, J. M. (2013) *Introduction to Mathematical Portfolio Theory*. N.Y.: Cambridge University Press.

Kalayci, C.B., Ertenlice O. and Akbay M.A. (2019) ‘A Comprehensive Review of Deterministic Models and Applications for Mean-variance Portfolio Optimization’, *Expert Systems with Applications*, 125(2019), pp. 345-368.

Kolm, P.N., Tutuncu, R. and Fabozzi, F.J. (2014) ’60 Years of Portfolio Optimization: Practical challenges and current trends’, *European Journal of Operational Research*, 234(2), pp. 356-371. doi: https://doi.org/10.1016/j.ejor.2013.10.060

Li, X., Uysal, A.S. and Mulvey, J.M. (2022) ‘Multi-period Portfolio Optimization using Model Predictive Control with Mean-variance and Risk Parity Frameworks’, *European Journal of Operational Research*, 299(3), pp. 1158-1176. doi: https://doi.org/10.1016/j.ejor.2021.10.002

Lo, A.W. and Mackinlay, A.C. (1988) ‘Stock Market Prices Do Not Follow Random Walks: Evidence from a Simple Specification Test’, *The Review of Financial Studies*, 1(1), pp. 41-66. doi: https://doi.org/10.1093/rfs/1.1.41

Markowitz, H. (1952) ‘Portfolio Selection’, *Journal of Finance*, 7(1), pp. 77-91.

Prashanth, L.A. and Ghavamzadeh, M. (2014) ‘Actor-Critic Algorithms for Risk-Sensitive Reinforcement Learning’, *arXiv preprint arXiv:1403.6530.*

Sharpe, W. F. (1964) ‘Capital Asset Prices: A Theory of Market Equilibrium under Conditions or Risk’, *Journal of Finance*, 19(3), pp. 425-442.

Tobin, J. (1958) ‘Liquidity Preference as Behaviour Towards Risk’, *The Review of Economic Studies*, 25(2), pp. 65-86.