ADVANCED MATHEMATICS
MASTER'S FINAL PROJECT

# MODERN PORTFOLIO OPTIMIZATION

*Author:*
Fernando Conde Montero

*Supervisor:*
Dr. José Manuel Corcuera Valverde

**Facultat de Matemàtiques i Informàtica**

September, 2022

# Abstract

The objective of this thesis is to survey some of the many models studied on modern portfolio theory, one of the main branches of quantitative finance. The first part of this work is dedicated to covering some of the main results on convex optimization with special emphasis on the Lagrangian and the Karush-Kuhn-Tucker optimality conditions. The second and third chapter are dedicated to two of the first and most important optimization models: the Markowitz model and the Capital Asset Pricing Model (CAPM). These two models are of paramount importance as they are the building blocks upon which later developments stand. However these models are quite static in the sense that they only allow for one period of time so, in the fourth chapter we introduce two multi-period models. For simplicity we will only contemplate the case with one risk-free asset and one risky asset, although the ideas there exposed allow the incorporation of many risky assets. So far, all models assumed that there was only one price at which assets are sold and bought. In the final chapter we will extend the notion of optimal portfolio to the context of financial market with two prices (the bid and ask price).

# Contents

# Chapter 1

# Introduction

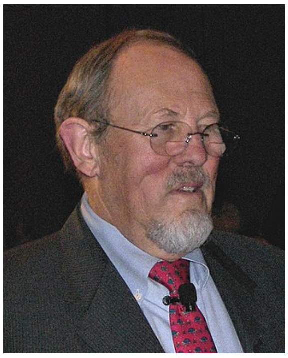
The problem of asset allocation can be stated as follows: given a fixed amount of money and a number of investment opportunities, what is the best way to distribute the money among these investments? This problem has been around, on one form or another, since the beginning of commerce. Zeno of Citium experimented first hand the dangers of not diversifying his assets when the ship carrying *all* of his cargo sank in a storm and he was left bankrupt. It was this life-changing event that pushed him away from commerce and onto the study of philosophy, so, in a sense, Stoicism has its origin in one bad investment.

Diversification of investments has always been common practice among investors (and the general population for that matter; we've all heard the popular saying "don't put all your eggs on one basket") since way before modern portfolio theory began. What lacked prior to Markowitz's seminal 1952 paper *Portfolio selection* ([9]) was an adequate theoretical framework that gave mathematical meaning to concepts, such as risk or asset correlation, which investors were already very familiar with and which allowed to analyze the effects of diversification and risk-return trade-offs on portfolios as a whole. Markowitz was the first to give a quantifiable measure of risk in the form of covariance among the assets that form a portfolio, thus taking the intuitive idea that diversification is desirable and giving it mathematical meaning. He devised the notion of optimal portfolio, efficient portfolio and efficient frontier, which are still used nowadays and which are all natural consequences his formulation of the asset allocation problem. Furthermore, he proved that any efficient portfolio can be obtained as a linear combination of the optimal portfolio and a risk-free asset (or any efficient portfolio when there is no risk-free asset). But, as revolutionary as this formulation was, it remained largely an academic curiosity known only to a few, due to a fundamental flaw that lies at its core: the model is very data exhaustive and very sensible to the *expected returns*. Although these parameters are mathematically well defined, they are famously difficult to estimate in rel life.

It wasn't until Sharpe et al. ([14]) introduced their capital asset pricing model (CAPM) and the notion of market equilibrium that researchers regained interest in portfolio theory. Sharpe argued that, assuming that the market has reached an equilibrium, there is no need to solve any optimization problem as the market portfolio should be observable from the capitalization of the companies that participate in the stock market. Sharpe also introduced a fundamental concept, still much used nowadays, referred to as an asset's (or portfolio's) beta, which represent the degree of correlation with the market as a whole. From here, he deduced the existence of systemic risk, a type of risk which can't be diversified away, and thus improved upon Markowitz's formulation.

Both of these models are what we call single-period models, as you start setting up the problem and only take into account what happens after a fixed time horizon with no information on what happens in between or afterwards. In the fourth chapter of this thesis we will explore one of the natural extensions of these models: multi-period models. In particular we will work with the binomial model and the trinomial model. The former was devised by Cox, Ross and Rubinstein and even though its formulation is quite simple, the results obtained from it are far from trivial.

(a)



(b)

Figure 1.1: Harry Markowitz (left), William Sharpe (right) and Merton Miller were awarded the 1990 Nobel prize in economics "for their pioneering work in the theory of financial economics".

This models has the property of being complete while the trinomial, although sharing a very similar formulation, is not. We will explore how this difference affects the optimization problem and implement a solution method known as the martingale method.

In the final chapter of this thesis we will serve as an introduction to conic finance theory. This theory, also known as two-price theory, tries to better reflect the behaviour of markets by renouncing the law of one price. This law is nothing more than the abstraction that investors buy and sell stock at the same price, but in any real market one always observes two prices, that is, the price at which the market is willing to buy (*bid*) and the price at which the market is willing to sell (*ask*)[1]. Once we have seen the basicis of conic finance theory we will analyze portfolio optimization from this optics for two cases: the long-only portfolio and the long-short portfolio. However, portfolio optimization in the context of conic finance is still in great measure an open problem.

[1] How is the law of one price reasonable then? Because the bid and the ask price are usually not far apart from one another.

# Chapter 2

# Convex optimization

The objective of this work is to study portfolio optimization in different frameworks, so we will start by covering the main tools and results in convex optimization in order to later apply them in different scenarios. This chapter is mainly based on [2].

## 2.1 Definitions and terminology

As in any subject, before we can start stating theorems and their proofs we must first define the mathematical objects that these are about.

**Definition 2.1.1.** We say that a set $C \subseteq \mathbb{R}^n$ is *affine* if $\forall x, y \in C$ and $\lambda \in \mathbb{R}$ we have that

$$\lambda x + (1 - \lambda)y \in C$$

**Definition 2.1.2.** We say that a set $C \subseteq \mathbb{R}^n$ is *convex* if $\forall x, y \in C$ and $\lambda \in [0, 1]$ we have that

$$\lambda x + (1 - \lambda)y \in C$$

In other words, the set $C$ is convex if for any 2 points in $C$ the straight line connecting both points is also fully contained in $C$. Notice that affinity implies convexity but not the other way round.

**Definition 2.1.3.** A set $C \subseteq \mathbb{R}^n$ is called a cone if $\forall x \in C$ and $\theta \geq 0$ we have that

$$\theta x \in C$$

Combining the two previous definitions we now define convex cones.

**Definition 2.1.4.** A convex cone is a set $C \subseteq \mathbb{R}^n$ such that $\forall x_1, x_2 \in C$ and $\theta_1, \theta_2 \geq 0$ we have that

$$\theta_1 x_1 + \theta_2 x_2 \in C$$

**Example.**

- *The empty set $\emptyset$, a single point $\{x_0\}$ and $\mathbb{R}^n$ are all affine (and thus also convex) subsets of $\mathbb{R}^n$.*
- *Line segments and rays are convex but not affine.*

**Definition 2.1.5.** Given $k$ points $x_1, ..., x_k$ we define an *affine combination* as any point of the form

$$y = \sum_{i=1}^{k} \theta_i x_i$$

Where $\theta_i \in \mathbb{R}$ for $i = 1, ..., k$ and $\sum_{i=1}^{k} \theta_i = 1$.

Any affine set contains all affine combinations of its points. If we substitute the condition that $\theta_i \in \mathbb{R}$ for $i = 1, ..., k$ for $\theta_i \in \mathbb{R}^+$ for $i = 1, ..., k$ in the previous definition, we have a *convex combination* instead.

**Definition 2.1.6.** Given $C \subseteq \mathbb{R}^n$, we call *affine hull* of $C$ to the set

$$\textbf{aff}\, C = \{\sum_{i=1}^{k} \theta_i x_i | x_1, ..., x_k \in C, \sum_{i=1}^{k} \theta_i = 1\}$$

The affine hull of a set is the set of all affine combinations of points belonging to that set. It's the smallest affine set that contains that set. We can similarly define the convex hull as the set of all convex combinations of points in a given set (or equivalently as the smallest convex set that contains a given set).

**Definition 2.1.7.** Given $C \subseteq \mathbb{R}^n$, we call *convex hull* of $C$ to the set

$$\textbf{conv}\, C = \{\sum_{i=1}^{k} \theta_i x_i | x_i \in C, \theta_i \geq 0, i = 1, ..., k, \sum_{i=1}^{k} \theta_i = 1\}$$

We can now use the affine hull to define the *relative interior* of a set; a refinement of the concept of interior of a set which will be useful later on.

**Definition 2.1.8.** Given $C \subseteq \mathbb{R}^n$, we call *relative interior* of $C$ to the set

$$\textbf{relint}\, C = \{x \in C | B(x, r) \cap \textbf{aff}\, C \text{ for some } r > 0\}$$

where $B(x, r)$ denotes the ball of center $x$ and radius $r$.

**Definition 2.1.9.** A function $f : \Omega \subseteq \mathbb{R}^n \to \mathbb{R}$ is said to be convex if:

1. $\Omega$ is convex.

2. $\forall x, y \in \Omega$ and $\lambda \in [0, 1]$ we have

$$f(\lambda x + (1 - \lambda)y) \leq \lambda f(x) + (1 - \lambda) f(y)$$

Intuitively, if for any 2 points $x, y \in \Omega$ the line connecting them is always above $f$, then $f$ is convex. In addition, a function $f$ is said to be *concave* if $-f$ is convex.

**Definition 2.1.10.** We call subgradient of $f$ at $x_0$ $(\partial f(x_0))$ to every vector $\gamma \in \mathbb{R}^n$ such that

$$f(x) \geq f(x_0) + \gamma^T (x - x_0)$$

**Definition 2.1.11.** we call subdifferential of $f$ at $x_0$ to the set of all subgradients of $f$ at $x_0$.

This notion of subdifferential allows us to generalize the concept of derivative to convex functions which are not necessarily differentiable. Note that if $f$ is differentiable at $x_0$ then we have that $\partial f(x_0) = \nabla f(x_0)$. We can employ the subdifferential to characterize the minimum of a convex function with the following theorem:

**Theorem 2.1.1.** .
*Let f be a proper*[1] *convex function and* $x_0 \in \mathbb{R}^n$. *We have that* $x_0$ *minimizes* $f \iff 0 \in \partial f(x_0)$.

*Proof.*
In fact, $0 \in \partial f(x_0) \iff f(x) \geq f(x_0) + 0(x - x_0) \quad \forall x \in \mathbb{R}^n$. This, in turn, only holds true if $x_0$ does indeed minimize $f$. □

[1] By proper we mean that the function never takes on the value $-\infty$ and is not identically equal to $+\infty$.

**Definition 2.1.12.** We use the notation

$$\begin{aligned} \text{minimize:} \quad & f_0(x) && (2.1) \\ \text{subject to:} \quad & f_i(x) \leq 0 \quad \forall i = 1, \ldots, m && (2.2) \\ & h_j(x) = 0 \quad \forall j = 1, \ldots, p && (2.3) \end{aligned}$$

to describe the optimization problem in which we wish to minimize an objective function $f_0(x)$, subject to some inequality constraints given by $f_i(x)$ and some equality constraints given by $h_i(x)$.

If the objective function and all the constraint functions are convex then we call this a convex optimization problem. The domain of the problem is $\mathbf{D} = \bigcap_{i=1}^{m} \textbf{dom}\ f_i \cap \bigcap_{j=1}^{p} \textbf{dom}\ h_j$ and a point $x \in \mathbf{D}$ is said to be feasible. Finally, the optimal value $p^*$ of the problem is

$$p^* = \inf_{x \in \mathbf{D}} \{f_0(x) | f_i(x) \leq 0, i = 1, \ldots, m, \quad h_j(x) = 0, j = 1, \ldots p\}$$

.

## 2.2 The Dual and the Lagrangian

In this section we will cover one of the most important concepts we will encounter in this chapter: the dual. Consider an optimization problem as the one in definition 2.1.12, with no assumptions made on the nature of the objective function or the constraints. The basic idea is to transform our original optimization problem (the primal) into another one (its dual) which is easier to solve. The first step in this direction is defining the Lagrangian.

**Definition 2.2.1.** The Lagrangian associated to an optimization problem like the one in definition 2.1.12 is a function $L : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^p \to \mathbb{R}$ defined by

$$L(x, \lambda, \nu) = f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{j=1}^{p} \nu_i h_i(x) \qquad (2.4)$$

Notice that $L$ is a weighted sum of the objective function and the constraint functions. The parameters $\lambda_i$ and $\nu_i$ are the well known *Lagrange multipliers* and the vectors $\lambda, \nu$ are the dual variables associated to the problem. Now, the Lagrange dual function, or simply the dual, is the smallest value of the Lagrangian over $x$, that is,

$$g(\lambda, \nu) = \inf_{x \in \mathbf{D}} L(x, \lambda, \nu) \qquad (2.5)$$

If $\lambda \geq 0$, since $f_i(x) \leq 0$ for $i = 1, \ldots, m$ and $h_i(x) = 0$ for $i = 1, ..., p$, we have that $g(\lambda, \nu)$ is a lower bound on $p^*$ which depends on $(\lambda, \nu)$. This begs the question "what is the best lower bound we can obtain from the dual?" which leads to the associated dual optimization problem:

$$\begin{aligned} \text{maximize:} \quad & g(\lambda, \nu) && (2.6) \\ \text{subject to:} \quad & \lambda_i \geq 0 \quad \forall i = 1, \ldots, m && (2.7) \end{aligned}$$

Notice that the dual optimization problem is convex regardless of the convexity, or lack thereof, of the primal. This is due to the fact that $g(\lambda, \nu)$ is defined as the pointwise infimum of a family of affine functions of $(\lambda, \nu)$ and therefore is concave even when the primal is not.

Another matter that comes naturally to mind is under what conditions, if any, do we have equality between the optimal solutions to the dual ($d^*$) and the primal ($p^*$). When this equality holds we have *strong duality*. We will introduce Slater's condition as a sufficient condition for strong duality. To this end we must now introduce and prove the hyperplane separation theorem which will be instrumental in proving Slater's condition.

## 2.3 Hyperplane separation theorem

We will now enunciate and prove the hyperplane separation theorem, which in turn we will use to prove Slater's condition. In order to prove it we will need the following propositions.

**Proposition 2.3.1.** *Let $S \subseteq \mathbb{R}^n$ be a closed, non-empty convex set and let $y \notin S$. Then $\exists!\ \bar{x} \in S$ which is closest to $y$ i.e. $||y-\bar{x}|| \leq ||y-x||$, $\forall x \in S$. Furthermore, $\bar{x}$ is a minimizing point $\iff (y-\bar{x})^T(x-\bar{x}) \leq 0$.*

*Proof.*
**Existence.**
As $S \neq \emptyset$, we know that there exists $\bar{x} \in S$ such that the minimum distance from $y$ to $S$ is less than or equal to $||y-\bar{x}||$. If we now define the set $\hat{S}$ as

$$\hat{S} := S \cap \{x : ||y-x|| \leq ||y-\bar{x}||\}$$

we have that $\hat{S}$ is closed and bounded and, since norm is a continuous function, we conclude by the Weierstrass theorem that there exists a minimum point $\bar{x} \in S$ such that $||y-\bar{x}|| = Inf\{||y-x||,\ x \in S\}$.
**Uniqueness.**
Let $\hat{x} \in S$ such that $||y-\hat{x}|| = ||y-\bar{x}|| = \alpha$. As $S$ is convex, we have that $\frac{\hat{x}+\bar{x}}{2} \in S$. But,

$$\left\| y - \frac{\hat{x}-\bar{x}}{2} \right\| \leq \frac{1}{2}||y-\hat{x}|| + \frac{1}{2}||y-\bar{x}|| = \alpha$$

Therefore, $||y-\hat{x}|| = \mu||y-\bar{x}||$, for some $\mu$. Now $||\mu|| = 1$. If $\mu = -1$, then $(y-\hat{x}) = -(y-\bar{x})$ which implies that $y = \frac{\hat{x}+\bar{x}}{2} \in S$ but $y \notin S$ by assumption so we have a contradiction. Thus, $\mu = 1$ and $\hat{x} = \bar{x}$ and the minimizing point is unique.

Finally for the last part of the proof, assume that $(y-\bar{x})^T(x-\bar{x}) \leq 0$ for all $x \in S$. Thus,

$$\begin{aligned} ||y-x||^2 &= ||y-\bar{x}+\bar{x}-x||^2 \\ &= ||y-\bar{x}||^2 + ||\bar{x}-x||^2 + 2(y-\bar{x})^T(\bar{x}-x) \end{aligned}$$

□

Therefore, as $||\bar{x}-x||^2 \geq 0$ and $(y-\bar{x})^T(\bar{x}-x) \geq 0$ we conclude that $||y-x||^2 \geq ||y-\bar{x}||^2$ for all $x \in S$ and $\bar{x}$ is indeed a minimizing point
Conversely, assume $\hat{x}$ is a minimizing point. Therefore $||y-x||^2 \geq ||y-\bar{x}||^2$ for all $x \in S$. Since $S$ is a convex set we have that $\lambda x + (1-\lambda)\bar{x} = \bar{x} + \lambda(x-\bar{x}) \in S$ for all $x \in S$, $\lambda \in [0,1]$. Thus,

$$||y-\bar{x}-\lambda(x-\bar{x})||^2 \geq ||y-\bar{x}||^2$$

and

$$||y-\bar{x}-\lambda(x-\bar{x})||^2 = ||y-\bar{x}||^2 + \lambda^2||x-\bar{x}||^2 - 2\lambda(y-\bar{x})^T(x-\bar{x})$$

from which

$$\begin{gathered} ||y-\bar{x}||^2 + \lambda^2||x-\bar{x}||^2 - 2\lambda(y-\bar{x})^T(x-\bar{x}) \geq ||y-\bar{x}||^2 \\ 2\lambda(y-\bar{x})^T(x-\bar{x}) \leq \lambda^2||x-\bar{x}|| \\ (y-\bar{x})^T(x-\bar{x}) \leq \frac{\lambda}{2}||x-\bar{x}|| \end{gathered}$$

As this expression holds for all $\lambda \in [0,1]$, taking $\lambda = 0$ we obtain:

$$(y-\bar{x})^T(x-\bar{x}) \leq 0$$

**Proposition 2.3.2.** *Let $S \subseteq \mathbb{R}^n$ be a closed, non-empty convex set and let $y \notin S$. Then there exists a hyperplane which separates $y$ and $S$, that is, $\forall x \in S\ \exists a \in \mathbb{R}^n, a \neq 0$ and $\alpha \in \mathbb{R}$ such that:*

$$a^T y > \alpha, \quad \textit{and} \quad a^T x \leq \alpha$$

*Proof.*
Since $S$ is a closed, non empty convex set and $y \notin S$ we know by the previous proposition that there exists $\bar{x} \in S$ such that $\forall x \in S$ the following inequality holds true:

$$(x-\bar{x})^T(y-\bar{x}) \leq 0$$

Therefore if we take

$$||y-\bar{x}||^2 = (y-\bar{x})^T(y-\bar{x}) = y^T(y-\bar{x}) - \bar{x}^T(y-\bar{x})$$

and combine it with the previous expression we obtain

$$||y-\bar{x}||^2 \leq (y-x)^T(y-\bar{x})$$

Thus, if we choose $a = y - \bar{x}$ we obtain $a^T y \geq a^T x + ||y-\bar{x}||^2$. Finally, we take $\alpha = \sup_{x\in S} a^T x$ and we are done. □

**Proposition 2.3.3.** *Let $S \subseteq \mathbb{R}^n$ be a convex set and let $\bar{x}$ be a point at the boundary of S. The S has a supporting hyperplane at $\bar{x}$, that is, $\exists a \in \mathbb{R}^n, a \neq 0$ such that $\forall x \in cl(S)$ we have that*

$$a^T(x-\bar{x}) \leq 0$$

*Proof.*
Since $\bar{x} \in \partial S$ we know that there exists a sequence $\{y_k\}$ such that $\forall k : y^k \notin cl(S)$ and $y^k \to \bar{x}$. Now, applying the previous theorem we know that to each $y_k$ we can associate a certain $a_k \neq 0$ such that $\forall x \in cl(S)$ we have that

$$a_k^T(x-y_k) < 0$$

Now, given that $a_k \neq 0, \forall k$ we can take without loss of generality that $||a_k|| = 1, \forall k$. Thus, the whole sequence is contained in a compact set and we can take a sub-sequence which converges to $a$ such that $||a|| = 1$. Finally, considering this sub-sequence and passing to the limit in the previous expression we obtain that $\forall x \in cl(S)$ we have that

$$a^T(x-\bar{x}) \leq 0$$

and we are done. □

**Corollary 2.3.4.** *Let $S \subseteq \mathbb{R}^n$ be a non-empty convex set and $\notin S$.Then $\exists a \in \mathbb{R}^n, a \neq 0$ such that $\forall x \in cl(S)$ we have that*

$$a^T(x-\bar{x}) \leq 0$$

*Proof.*
If $\bar{x} \notin cl(S)$ the result follows from Proposition 2.3.2 and if $\bar{x} \in \partial S$ it follows from Proposition 2.3.3. □

Taking into account these propositions we can now finally state and prove the hyperplane separation theorem; a very useful result which we will use later on.

**Theorem 2.3.5. *Hyperplane separation theorem***
*Let $S_1$ and $S_2$ be 2 non-empty convex sets with $S_1 \cap S_2 = \emptyset$. Then there exists a hyperplane which separates these sets, that is, $\exists a \in \mathbb{R}^n$ and $b \in \mathbb{R}$ such that $\forall x_1 \in S_1$ and $\forall x_2 \in S_2$ we have that*

$$a^T x_1 \leq b$$

$$a^T x_2 \geq b$$

*. The hyperplane $\{x|a^T x = b\}$ is the aforementioned separating hyperplane.*

*Proof.*
Let $S = \{x | x = x_1 - x_2, x_1 \in S_1, x_2 \in S_2\}$. Notice that S is convex and that $0 \notin S$. By applying Corollary 2.3.4 we know that $\exists a \in \mathbb{R}^n, a \neq 0$ such that $\forall x \in S$ we have that

$$a^T x \leq 0$$

Therefore we have that $\forall x_1 \in S_1$ and $\forall x_2 \in x_2$

$$a^T(x_1 - x_2) \leq 0$$

and we are done. □

## 2.4 Strong duality

As mentioned in section 2.2, we have strong duality when the equality

$$d^* = p^*$$

holds true. This is generally not the case, but there exists a sufficient condition which, if the primal satisfies, guarantees strong duality.

**Theorem 2.4.1.** ***Slater's condition.***
*Let the the following be a convex optimization problem:*

$$\begin{aligned} \textit{minimize:} \quad & f_0(x) && (2.8)\\ \textit{subject to:} \quad & f_i(x) \leq 0 \quad \forall i = 1, \dots, m && (2.9)\\ & h_j(x) = 0 \quad \forall j = 1, \dots, p && (2.10) \end{aligned}$$

*If* $\exists\, x^* \in$ ***relint D*** *such that*

$$\begin{aligned} f_i(x^*) < 0 \quad \forall i = 1, \dots, m \\ h_j(x) = 0 \quad \forall j = 1, \dots, p \end{aligned} \qquad (2.12)$$

*then we have strong duality.*

*Proof.*
We consider an optimization problem:

$$\begin{aligned} \text{minimize:} \quad & f_0(x) && (2.13)\\ \text{subject to:} \quad & f_i(x) \leq 0 \quad \forall i = 1, \dots, m && (2.14)\\ & Ax = b && (2.15) \end{aligned}$$

and assume that $f_i$ are convex $\forall i = 0, \dots, m$ and that Slater's condition is satisfied. Notice that we have expressed the equality constraints in matrix form. We make two more assumptions to further simplify things:

1. **D** has non empty interior (hence, **relint D**=**int D**)
2. Rank(A)=p

We also assume without loss of generality that $p^*$ is finite. Now we define two convex sets $A$ and $B$ as follows:

$$\begin{aligned} &A = \{(u, v, t) | \exists x \in \mathbf{D}, \quad f_i(x) \leq u_i, \forall i = 1, \dots, m, \quad h_j(x) = v_i \forall j = 1, \dots, p, \quad f_0(x) \leq t\} \\ &B = \{(0, 0, s) \in \mathbb{R}^m \times \mathbb{R}^p \times \mathbb{R} | s < p^*\} \end{aligned} \qquad (2.16)$$

Where set $A$ is convex as long as the underlying problem is convex and set $B$ is a ray and therefore convex. Now suppose we take $(u, v, t) \in A \cap B$. Since $(u, v, t) \in B$ we know that $u = v = 0$ and that $t < p^*$. On the other hand, as $(u, v, t) \in A$ we know that exists $x$ which satisfies all the constraints as well as $f_0(x) \leq t < p^*$ which is a contradiction because $p^*$ is the optimal value for the primal problem. Hence $A \cap B = \emptyset$.

We can now apply the hyperplane separation theorem (2.3.5) and conclude that $\exists(\tilde{\lambda}, \tilde{\nu}, \mu)$ and $\alpha$ such that:

$$(u, v, t) \in A \Rightarrow \tilde{\lambda}^T u + \tilde{\nu}^T v + \mu t \geq \alpha \qquad (2.17)$$

and

$$(u, v, t) \in B \Rightarrow \tilde{\lambda}^T u + \tilde{\nu}^T v + \mu t \leq \alpha \qquad (2.18)$$

From (2.17) we conclude that $\tilde{\lambda} \geq 0$ and $\mu \geq 0$, since otherwise $\tilde{\lambda}^T u + \mu t$ would be unbounded below over A and that would contradict (2.17). Condition (2.18) means that $\mu t \leq \alpha$ for all $t < p$ and hence $\mu p^* \leq \alpha$. From both conditions we conclude that $\forall x \in \mathbf{D}$

$$\sum_{i=1}^{m} \tilde{\lambda}_i f_i(x) + \tilde{\nu}^T(Ax - b) + \mu f_0(x) \geq \alpha \geq \mu p^* \tag{2.19}$$

Notice that assuming that $\mu > 0$ we can divide the previous expression by $\mu$ and we are left with a Lagrangian function on the left hand side:

$$L\left(x, \frac{\tilde{\lambda}}{\mu}, \frac{\tilde{\nu}}{\mu}\right) \geq p^* \tag{2.20}$$

Thus, if we minimize over $x$ we obtain that

$$g(\lambda, \nu) \geq p^* \tag{2.21}$$

where we have taken $\lambda = \tilde{\lambda}/\mu$ and $\nu = \tilde{\nu}/\mu$. As we know that $g(\lambda, \nu)$ is a lower bound of $p^*$ we conclude that in the previous expression we actually have equality. Hence, strong duality holds and the dual optimum is attained.

Finally, lets examine the case $\mu = 0$. In this case equation (2.19) takes the form

$$\sum_{i=1}^{m} \tilde{\lambda}_i f_i(x) + \tilde{\nu}^T(Ax - b) \geq 0 \tag{2.22}$$

Thus, applying the previous expression to a certain $\tilde{x} \in \textbf{relint}\,\mathbf{D}$ which satisfies Slater‘s condition we obtain

$$\sum_{i=1}^{m} \tilde{\lambda}_i f_i(\tilde{x}) \geq 0 \tag{2.23}$$

Since $f_i(\tilde{x}) < 0$ and $\tilde{\lambda}_i \geq 0$ we conclude that $\tilde{\lambda} = 0$. From $(\tilde{\lambda}, \tilde{\nu}, \mu) \neq 0$ and $\tilde{\lambda} = \mu = 0$ we conclude that $\tilde{\nu} \neq 0$. In turn, equation (2.19) implies that $\forall x \in \mathbf{D}$

$$\tilde{\nu}^T(Ax - b) \geq 0$$

But, as $\tilde{x}$ satisfies $\tilde{\nu}^T(A\tilde{x} - b) = 0$, and since $\tilde{x} \in \textbf{int}\,\mathbf{D}$, there are points in $\mathbf{D}$ with $\tilde{\nu}^T(Ax - b) < 0$ unless $A^T\tilde{\nu} = 0$. This contradicts our initial assumption that rank(A)=p. □

Now lets suppose we have strong duality for a given optimization problem and let $x^*$ and $(\lambda^*, \nu^*)$ be a primal optimal solution and a dual optimal solution respectively. Therefore we have that

$$\begin{aligned} f_0(x^*) &= g(\lambda^*, \nu^*) \\ &= \inf_{x \in \mathbf{D}} \left( f_0(x) + \sum_{i=1}^{m} \lambda_i^* f_i(x) + \sum_{i=1}^{p} \nu_i^* h_i(x) \right) \\ &\leq f_0(x^*) + \sum_{i=1}^{m} \lambda_i^* f_i(x^*) + \sum_{i=1}^{p} \nu_i^* h_i(x^*) \\ &\leq f_0(x^*) \end{aligned} \tag{2.24}$$

The first equality reflects the strong duality we have assumed; the second equality is just the definition of the dual. The first inequality follows from the fact that the infimum of the Lagrangian over $x$ is less or equal to $x^*$. The second inequality is due to $\lambda^* \geq 0$, $f_i(x^*) \leq 0$ for $i = 1, ..., m$ and the equality constraints being satisfied. Thus both inequalities in the previous expression are actually equalities. Several conclusions can be extracted from this relation. First notice that $x^*$ is a minimizer of the Lagrangian (though not the only one). We also have the very useful identity:

$$\sum_{i=1}^{m} \lambda_i^* f_i(x^*) = 0 \tag{2.25}$$

and since $\lambda_i, \geq 0$ and $f_i \leq 0$ $\forall i = 1, \dots, m$ it actually holds that

$$\lambda_i^* f_i(x^*) = 0, \quad \forall i = 1, \dots, m \tag{2.26}$$

Condition (2.26) is known as *complementary slackness* and holds when we have strong duality. We can equivalently express it as

$$\begin{cases} \lambda_i^* > 0 \quad \Rightarrow \quad \mathrm{f}_i(x^*) = 0 \\ \mathrm{f}_i(x^*) < 0 \quad \Rightarrow \quad \lambda_i^* = 0 \end{cases}$$

Which intuitively means that Lagrange multipliers are zero unless the associated constraint is active at the optimum.

## 2.5 Karush-Kuhn-Tucker conditions

Let us at this point make two assumptions:

1. Let the objective function as well as all the constraint functions be differentiable.
2. Let there be strong duality.

With these assumptions and collecting everything covered thus far, we're ready to enumerate the KKT conditions. Letting $x^*$ and $(\lambda^*, \nu^*)$ be the primal and dual solutions respectively we have:

1. **Stationarity.** As $x^*, \lambda^*, \nu^*$ minimize the Lagrangian we conclude that

$$0 \in \left( f_0(x^*) + \sum_{i=1}^{m} \lambda_i^* f_i(x^*) + \sum_{j=1}^{p} \nu_i^* h_i(x^*) \right)$$

   and, as we've assumed that all functions are differentiable, we can express the condition as

$$\nabla f_0(x^*) + \sum_{i=1}^{m} \lambda_i^* \nabla f_i(x^*) + \sum_{j=1}^{p} \nu_i^* \nabla h_i(x^*) = 0 \tag{2.27}$$

2. **Complementary slackness.** This condition relates each active constrain to its Lagrange multiplier and can be expressed as

$$\lambda_i^* f_i(x^*) = 0, \quad \forall i = 1, \dots, m \tag{2.28}$$

3. **Primal viability.** Which simply reflects the fact that any viable solution must satisfy the primal's restrictions

$$\begin{cases} f_i(x^*) \leq 0 \quad \text{for} \quad i = 1, ..., m \\ h_i(x^*) = 0 \quad \text{for} \quad i = 1, ..., p \end{cases} \tag{2.29}$$

4. **Dual viability.** Which guarantees that the Lagrange multipliers associated with the inequality constraints must be non-negative

$$\lambda_i \geq 0 \quad \text{for} \quad i = 1, ..., m \tag{2.30}$$

We will see that with the next theorem that any feasible convex optimization problem with differentiable objective and constraint functions and strong duality must satisfy these conditions. KKT conditions play an instrumental role in many optimization algorithms.

**Theorem 2.5.1.** *Given a convex optimization problem in the form of definition 2.1.12, if there is strong duality the KKT conditions are necessary and sufficient for there to exist an optimal solution.*

*Proof.*
**Necessity.**
Let $x^*, (\lambda^*, \nu^*)$ be the optimal solutions to the primal and dual respectively. As we have strong duality we know that

$$f_0(x^*) = g(\lambda^*, \nu^*) \tag{2.31}$$

$$= f_0(x^*) + \sum_{i=1}^{m} \lambda_i^* f_i(x^*) + \sum_{i=1}^{p} \nu_i^* h_i(x^*) \tag{2.32}$$

$$\leq f_0(x^*) + \sum_{i=1}^{m} \lambda_i^* f_i(x^*) + \sum_{i=1}^{p} \nu_i^* h_i(x^*) \tag{2.33}$$

$$\leq f_0(x^*) \tag{2.34}$$

Where, as we saw in section 2.4, the inequalities are actually equalities. From here we see that all the KKT conditions are fullfiled: $x^*$ minimizes the Lagrangian over $x$ so we know that the the stationarity condition is satisfied; complementary slackness, dual and primal feasibility must also hold true for there to be equality between the third and fourth line.

**Sufficiency.**
Let $x^*\lambda^*\nu^*$ be such that the KKT conditions are fulfilled. In such case we have that, since $\lambda \geq 0$, the Lagrangian is convex. This fact coupled with the stationarity condition allows us to conclude that $x^*$ minimizes the Lagrangian over $x$. It follows that

$$\begin{aligned} g(\lambda^*, \nu^*) &= L(x^*, \lambda^*, \nu^*) \\ &= f_0(x^*) + \sum_{i=1}^{m} \lambda_i f_i(x^*) + \sum_{j=1}^{p} \nu_i h_i(x^*) \qquad (2.35) \\ &= f_0(x^*) \end{aligned}$$

The first equality holds by definition and and the second one due to the complementary slackness condition by which we know that $\sum_{i=1}^{m} \lambda_i f_i(x^*) = 0$ and due to the primal viability condition by which we know that $h_i(x^*) = 0$ for all $i = 1, ..., p$. □

# Chapter 3

# The Markowitz Model

We begin our study of modern portfolio theory with the model that started it all: the Markowitz model. In his seminal 1952 paper titled *Portfolio Selection* [9] economist Harry Markowitz argued that expected returns should not be the only factor involved when deciding in what assets should one invest in and gave sound mathematical grounds to what diversifying means and why it is desirable. We will cover the main results in this framework as well as its advantages and shortcomings. This section is mainly based in [9] and [15].

## 3.1 Preliminary concepts, definitions and notation

In simple terms, the idea is that we have a given deterministic amount of money at time zero ($X_0$) which we wish to invest in $n$ different assets hoping to obtain the best performance. These $n$ different assets will form our portfolio. For simplicity we will consider that we can only invest in risky stocks and will leave out other assets as options, bonds, swaps, etc. We denote with $X_{0i}$ the amount of the total initial amount of money $X_0$ invested in the asset $i$. Thus we have:

$$X_0 = \sum_{i=1}^{n} X_{0i} \tag{3.1}$$

After one period of time the value of each asset is given by the random variable $X_{1i}$ and the value of the portfolio will be:

$$X_1 = \sum_{i=1}^{n} X_{1i} \tag{3.2}$$

Notice that each $X_{0i}$ can be positive or negative as long as $X_0 > 0$. It should be clear that $X_{0i} > 0$ means buying a certain amount of stocks which corresponds to the value of $X_{0i}$, but what about $X_{0i} < 0$? This case corresponds to what is known as *shorting* and for all intent and purpose should be regarded as the polar opposite of buying stocks (hence its name; buying stocks is also known as *longing*). One buys a stock hoping to make a profit out of the success of that company and one shorts a stock hoping to make a profit out of the downfall of a company[1].

We now define several important concepts that we will use through out this chapter.

**Definition 3.1.1.** The *total rate of return* of an asset i is given by

$$R_i = \frac{X_{1i}}{X_{0i}} \tag{3.3}$$

[1] An important thing to notice about shorting is that potential losses are unbounded from below. When we buy a share of a company the most we can potentially loose is what we originally invested, but theoretically there is no bound to what our profit can be if the company does well. In the other hand, when we short a share of a company the most we can hope to obtain as a profit is what we originally invested but potential losses, as mentioned, are unbounded from below. Thus, shorting is a inherently risky investing strategy but this risk is not reflected on the variance of the portfolio.

**Definition 3.1.2.** The *rate of return* of an asset i is given by

$$r_i := R_i - 1 = \frac{X_{1i} - X_{0i}}{X_{0i}} \tag{3.4}$$

**Definition 3.1.3.** The *expected rate of return*, or simply *expected return* of an asset i is given by

$$\bar{r_i} = E(r_i) \tag{3.5}$$

**Definition 3.1.4.** The *weights* of a portfolio are given by

$$\alpha_i = \frac{X_{0i}}{X_0} \tag{3.6}$$

and it follows that $\sum_{i=1}^{n} \alpha_i = 1$. Weights are very useful because they allow us to comfortably and compactly describe our portfolio.

Equipped with these notions we can now define the central concepts in the Markowitz model: the expected rate of return the variance of our portfolio. The rate of return of our portfolio can be expresses in terms of the wights and the rate of return of each asset as:

$$r = \sum_{i=1}^{n} \alpha_i r_i \tag{3.7}$$

It follows from here that the expected rate of return of our portfolio is:

$$\bar{r} = \sum_{i=1}^{n} \alpha_i \bar{r_i} \tag{3.8}$$

Any $\bar{r}$ that satisfies 3.8 is said to be *feasible*. The variance of the rate of return of each asset is given by

$$\sigma_i^2 = E(r_i^2) - \bar{r_i}^2 \tag{3.9}$$

And the covariance of each pair of assets by

$$\sigma_{ij} = E\{(r_i - \bar{r_i})(r_j - \bar{r_j})\} = E(r_i r_j) - \bar{r_i}\bar{r_j} \tag{3.10}$$

Thus, taking into account that $\sigma_{ii} = \sigma_i^2$, the variance of the rate of return of the portfolio can be expressed as:

$$\sigma^2 = \frac{1}{2} \sum_{i,j=1}^{n} \alpha_i \alpha_j \sigma_{ij} \tag{3.11}$$

The variance of the portfolio is the simplest measure of risk (there are other measures of risk based on utility functions which we will address in later sections, but for now we will limit ourselves with variance). It's easy to see why; the lower the $\sigma^2$, the thinner our distribution will be and the actual rate or return of our portfolio will have a higher probability of being closer to the expected rate of return. Notice that as the rate of return and variance of a portfolio is well defined, we can treat a whole portfolio as an asset for another portfolio.

## 3.2 Assumptions and formulation of the problem

In order to derive his model, Markowitz makes a series of assumptions about the nature of investors and markets. Firstly, we assume that:

- Investors are rational.
- Investors are risk averse.

Although everyone has, more or less, an idea of what being rational means, sometimes people struggle to agree on what the rational choice *is* when faced with a difficult decision. Thus, to avoid all and any possible ambiguities, in this context we will take it to mean that, when faced with having to choose between to assets of *equal variance*, investors will always choose the one with higher expected return. On the other hand, investors being risk averse means that they will always choose the asset with a lower variance when having to choose from assets of *equal expected return*. Hence, we conclude that investors should diversify their funds among all those securities which give maximum expected returns.

Markowitz rejects the hypothesis that investors should want to maximize expected returns regardless of risk because this leads, without exception, to investing everything in the asset with the highest expected return (as is clear from equation (3.8)). Thus, this hypothesis never leads to the creation of a diverse portfolio. With regards to the behaviour of the market we assume that:

- There is perfect competition[2]
- There is no privileged information.
- Stocks are arbitrarily divisible.
- There are no transaction costs.
- We only consider one period of time.
- Shorting is allowed in some assets as long as $\sum_{i=1}^{n} X_{0i} = X_0 > 0$

It follows from these assumptions that investors are searching for an optimal portfolio (i.e. the weights) which minimizes the variance for a certain desired expected return $\bar{r}$. Thus, we arrive to an optimization problem of the form:

$$\text{Minimize} \quad \sigma^2 = \frac{1}{2} \sum_{i,j=1}^{n} \alpha_i \alpha_j \sigma_{ij} \tag{3.12}$$

$$\text{subject to:} \quad \bar{r} = \sum_{i=1}^{n} \alpha_i \bar{r}_i = c, \qquad c \in \mathbb{R}_+ \tag{3.13}$$

$$\sum_{i=1}^{n} \alpha_i = 1 \tag{3.14}$$

We can transform this optimization problem into a linear program using the Lagrange multipliers and the KKT conditions. Firstly we must make each equality constraint equal zero and then, incorporate each of these constraints with their corresponding Lagrange multipliers $(\lambda_1, \lambda_2)$ into the Lagrangian associated to the problem. In this case the Lagrangian is of the form

$$L = \frac{1}{2} \sum_{i,j=1}^{n} \alpha_i \alpha_j \sigma_{ij} + \lambda_1 \left( \sum_{i=1}^{n} \alpha_i \bar{r}_i - c \right) + \lambda_2 \left( \sum_{i=1}^{n} \alpha_i - 1 \right) \tag{3.15}$$

[2]Which means that companies have no power to manipulate the price of stocks; it is determined purely by the interaction of supply and demand. It also implies that all stocks are perfectly liquid (i.e. at any point the bid and ask prices are always equal and there is always the possibility to buy or sell any desired amount of stocks at said price).

If we now take the derivative of $L$ with respect to each weight and equate them to zero we obtain a $n$ equations of the form

$$\sum_{j=1}^{n} \alpha_j \sigma_{ij} + \lambda_1 \bar{r_i} + \lambda_2 = 0 \tag{3.16}$$

for $i \in \{1, ..., n\}$. Each of these equations is linear $\alpha, \lambda_1$ and $\lambda_2$. These equations, coupled with constraints (3.13) and (3.14) yield the following system of $n+2$ equations:

$$\sum_{j=1}^{n} \alpha_j \sigma_{ij} + \lambda_1 \bar{r_i} + \lambda_2 = 0 \tag{3.17}$$

$$\sum_{i=1}^{n} \alpha_i \bar{r_i} = c \tag{3.18}$$

$$\sum_{i=1}^{n} \alpha_i = 1 \tag{3.19}$$

This is a linear program which we can easily solve. Any portfolio with $(\sigma, \bar{r})$ which satisfy 3.11 and 3.8 is said to be *feasible*. The set of all such portfolios is adeptly named the feasible set. For each fixed $\bar{r}$ the Markowitz problem produces the feasible portfolio with that expected return and minimum variance which lies in the frontier of the feasible set. If we now wish to find the portfolio with the smallest variance of all the minimum variance set we only need to eliminate condition 3.18 from our optimization problem. As we've eliminated one of the constraints, we now only need one Lagrange multiplier and the Lagrangian of the problem is given by

$$L = \frac{1}{2} \sum_{i,j=1}^{n} \alpha_i \alpha_j \sigma_{ij} + \lambda \left( \sum_{i=1}^{n} \alpha_i - 1 \right) \tag{3.20}$$

and proceeding as before, we arrive at a system of $n+1$ equations

$$\sum_{j=1}^{n} \alpha_j \sigma_{ij} + \lambda = 0 \tag{3.21}$$

$$\sum_{i=1}^{n} \alpha_i = 1 \tag{3.22}$$

We denote the minimum variance point with the coordinates $(\sigma^*, \bar{r}^*)$. The set made up of minimum variance portfolios with $\bar{r} > \bar{r}^*$ is known as the *efficient frontier*.

## 3.3 Efficient frontier

Portfolios that lie in the efficient frontier are known as efficient and the ones that don't are known as inefficient. The name stems from the fact that no rational investor would choose an inefficient portfolio over an efficient one. If shorting is allowed, the efficient frontier is unbounded from above and we can thus theoretically obtain a portfolio with an arbitrarily high expected return. Consider a portfolio formed by two risky assets with expected returns $\bar{r_1}, \bar{r_2}$ with $\bar{r_1} > \bar{r_2}$. Thus, the portfolio will have and expected return

$$\begin{aligned} \bar{r} &= \alpha \bar{r_1} + (1-\alpha)\bar{r_2} \\ &= \alpha(\bar{r_1} - \bar{r_2}) + \bar{r_2} \end{aligned} \tag{3.23}$$

Therefore, we can obtain any arbitrarily high $\bar{r}$ by investing increasing amounts in the asset with higher expected return and proportionally shorting the other asset (to ensure that $\alpha_1 + \alpha_2 = 0$). The variance of this portfolio is

$$\sigma^2 = \alpha^2 \sigma_1^2 + (1-\alpha)^2 \sigma_2^2 + 2\alpha(1-\alpha)\sigma_{12} \tag{3.24}$$

$$= \alpha^2(\sigma_1^2 + \sigma_2^2 - 2\sigma_{12}) + 2\alpha(\sigma_{12} - \sigma_2^2) + \sigma_2^2 \tag{3.25}$$

Now expressing the covariance as $\sigma_{12} = 2\rho\sigma_1\sigma_2$ we obtain

$$\sigma^2 = \alpha^2(\sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2) + 2\alpha(\rho\sigma_1\sigma_2 - \sigma_2^2) + \sigma_2^2 \tag{3.26}$$

Therefore, assuming that $\rho \neq 1$, it follows that

$$\lim_{\bar{r}\to\infty} \sigma^2 = \lim_{\alpha\to\infty} \sigma^2 \tag{3.27}$$

$$= \lim_{\alpha\to\infty} \alpha^2(\sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2) + 2\alpha(\rho\sigma_1\sigma_2 - \sigma_2^2) + \sigma_2^2 \tag{3.28}$$

$$= \infty \tag{3.29}$$

As noted in the previous section, investors can increase expected returns by taking on variance; shorting allows to do this ad infinitum.

We now present one of the main results of our model.

**Theorem 3.3.1** (Mutual fund separation theorem)**.** *Any efficient portfolio can be generated as a combination of two different efficient portfolios.*

*Proof.*
Let $w_1, w_2$ be two different portfolios which are a solution to the Markowitz problem with $\bar{r_1} \neq \bar{r_2}$ such that

$$w_1 = (\alpha_1, ..., \alpha_n, \lambda_1, \mu_1) \tag{3.30}$$

$$w_2 = (\beta_1, ..., \beta_n, \lambda_2, \mu_2) \tag{3.31}$$

from the linearity of the solution, we have that $\forall \alpha \in \mathbb{R}$, the combination

$$w_3 = \alpha w_1 + (1-\alpha) w_2 \tag{3.32}$$

is itself a solution to the Markowitz problem for the expected return $\bar{r} = \alpha\bar{r_1} + (1-\alpha)\bar{r_2}$.

We have that

$$\alpha w_1 = (\alpha\alpha_1, ..., \alpha\alpha_n, \alpha\lambda_1, \alpha\mu_1) \tag{3.33}$$

$$(1-\alpha)w_2 = ((1-\alpha)\beta_1, ..., (1-\alpha)\beta_n, (1-\alpha)\lambda_2, (1-\alpha)\mu_2) \tag{3.34}$$

Therefore, $w_3$ is of the form

$$w_3 = (\gamma_1, ..., \gamma_n, \lambda_3, \mu_3) \tag{3.35}$$

Where

$$\gamma_i = \alpha\alpha_i + (1-\alpha)\beta_i \tag{3.36}$$

$$\lambda_3 = \alpha\lambda_1 + (1-\alpha)\lambda_2 \tag{3.37}$$

$$\mu_3 = \alpha\mu_1 + (1-\alpha)\mu_2 \tag{3.38}$$

It follows that, $w_3$ is a solution to the linear system of equations 3.17-3.19. □

*Remark.* The efficient frontier can be generated from 2 distinct solutions to the Markowitz problem.

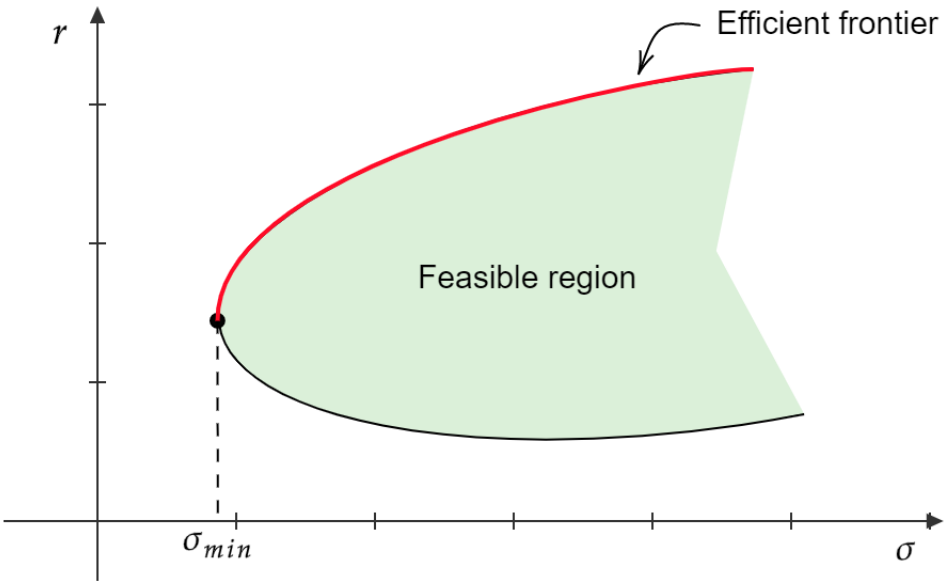



Figure 3.1: Sketch of the feasible region and its efficient frontier.

## 3.4 Incorporating a risk-free asset

In our previous analysis we assumed all $n$ assets to be risky ($\sigma_i^2 > 0, i \in \{1, ..., n\}$), we will now incorporate one more asset, a risk-free asset, and see how it alters the results. This type of asset usually means to place a certain amount of money in a bank account or use that money to buy bonds[3]. In any case, we will assume that the simple interest rate for this risk-free asset is $r_f$. Thus, if you invest an amount $x_0$ in this asset, after one period of time the payoff will be $x_1 = x_0(1+r_f)$. Notice that this asset, by definition, has $\sigma^2 = 0$ and $\bar{r} = r_f$. Therefore, the minimum variance portfolio would consist in investing all your funds in this one asset. Then why don't we do this? Because $r_f$ is usually quite small and not a great source of profit and perhaps we wish to obtain a higher return, greater than $r_f$. Instead, we will use this asset to lower the risk of our portfolio.

Consider now two portfolios, one made of only risky assets with weights $(\beta_1, ..., \beta_n)$, which we refer to as the *fund*, and another one which also includes the risk-free asset with weights $(\alpha_0, ..., \alpha_n)$. Here $\alpha_0$ corresponds to the risk-free asset. As before, we require that

$$\sum_{i=0}^{n} \alpha_i = \sum_{i=1}^{n} \beta_i = 1 \tag{3.39}$$

If we now take $1 - \alpha_0 = \sum_{i=1}^{n} \alpha_i$ we can rewrite the portfolio as

$$(\alpha_0, ..., \alpha_n) = (\alpha_0, (1 - \alpha_0)(\beta_1, ..., \beta_n)) \tag{3.40}$$

where

$$\beta_i = \frac{\alpha_i}{1 - \alpha_0} \tag{3.41}$$

[3]Naturally, no asset is truly risk-free as banks and bond issuers can go bankrupt, but this is a very unusual thing. Compared to stocks, the approximation that bank accounts and bonds are risk-free is good enough.

Notice that with this choice of $\beta_i$'s condition (3.39) holds. Thus we now have a portfolio formed by two assets, the risk- free one with weight $\alpha_0$ and the (risky) fund with weight $1-\alpha_0$. The fund has its own expected return and variance ($m$ and $\gamma^2$ respectively) which will be given by

$$m = \sum_{i=1}^{n} \beta_i \bar{r}_i \tag{3.42}$$

$$\gamma^2 = \sum_{i,j=1}^{n} \beta_i \beta_j \sigma_{ij} \tag{3.43}$$

Now, taking into account that the risk-free asset doesn't increase the variance, portfolio $(\alpha_0, ..., \alpha_n)$ has expected return and variance:

$$\bar{r} = \alpha_0 r_f + (1-\alpha_0)m \tag{3.44}$$

$$\sigma^2 = (1-\alpha_0)^2\gamma^2 \tag{3.45}$$

So, it is situated in the $\sigma - \bar{r}$ plane in the point:

$$(\sigma, \bar{r}) = (|1-\alpha_0|\gamma, \alpha_0 r_f + (1-\alpha_0)m) \tag{3.46}$$

For $\alpha_0 \leq 1$ we have that $|1-\alpha_0| = 1-\alpha_0$ and thus the point can be rewritten as:

$$(1-\alpha_0)(\gamma, m) + \alpha_0(0, r_f) \tag{3.47}$$

Therefore, depending on how much of the portfolio wants to be allocated on the risk free asset, $\alpha_0$ varies from 1 to 0 and the point traces out a line in the $(\sigma, r)$ plane that connects the risky portfolio $(\gamma, m)$ and the risk-free asset $(0, r_f)$. Thus, this line [4] is given by the equation

$$\bar{r} = \frac{m - r_f}{\gamma}\sigma + r_f \tag{3.48}$$

The slope must be positive, otherwise $r_f > m$ and there would be no point in investing any amount of money in the risky portfolio. Notice that the greater the slope, the more efficient our line is, as we get a higher expected return for the same variance, but we can't just choose any portfolio with high expected return, we need to choose an attainable portfolio from the feasible region. As before, rational investors should only consider portfolios that lie on the old efficient frontier and thus, we conclude that the optimal portfolio $M = (\gamma, m)$[5] which maximizes the slope is the one that yields a line tangent to the old efficient frontier.

**Theorem 3.4.1** (One-fund theorem). *When the possibility of incorporating a risk-free asset to a risky portfolio is allowed, the new efficient frontier is a line connecting the point $(0, r_f)$ to the unique portfolio $M$, which lies at the tangent point between the old efficient frontier and the new one. The fund $M = (\beta_1, ..., \beta_n)$ is given by*

$$\beta_i = \frac{v_i}{\sum_{j=1}^{n} v_j}, \quad i \in \{1, ..., n\} \tag{3.49}$$

*where $(v_1, ..., v_n)$ is the solution to the set of linear equations*

$$\sum_{j=1}^{n} v_j \sigma_{ij} = \bar{r}_i - r_f, \quad i \in \{1, ..., n\} \tag{3.50}$$

*Proof.*
As stated before, the portfolio $M$ is that which maximizes the slope of equation (3.48), that is, we have to find $(\beta_1, ..., \beta_n)$ such that it maximizes

$$f(\beta_1, ..., \beta_n) = \frac{m - r_f}{\gamma} \tag{3.51}$$

[4] Which we will later refer to as the *capital market line* in the context of CAPM.
[5] Known as the *market portfolio* in the context of CAPM.

Remember that $m$ and $\gamma$ are both functions of $(\beta_1, ..., \beta_n)$ as shown in expressions (3.42) and (3.43). Thus we wish to solve the following maximization problem

$$\text{Maximize} \quad f(\beta_1, ..., \beta_n) = \frac{\sum_{i=1}^n \beta_i \bar{r}_i - r_f}{\sqrt{\sum_{i,j=1}^n \beta_i \beta_j \sigma_{ij}}} \tag{3.52}$$

$$\text{subject to:} \qquad \sum_{i=1}^n \beta_i = 1 \tag{3.53}$$

As $\sum_{i=1}^n \beta_i = 1$, we can factor it out from the numerator and rewrite the previous expression as

$$f(\beta_1, ..., \beta_n) = \frac{\sum_{i=1}^n \beta_i (\bar{r}_i - r_f)}{\sqrt{\sum_{i,j=1}^n \beta_i \beta_j \sigma_{ij}}} \tag{3.54}$$

Notice that if doesn't make any difference changing $\beta_i$ for $c\beta_i$, with $c > 0$, as this constant would cancel out in the numerator and the denominator. Thus, we can ignore the constraint and deal with it later by normalizing and there is no need to to use Lagrange multipliers. Taking derivatives with respect to each $\beta_i$ and equating them to zero yields a system of $n$ linear equations

$$\sum_{i=1}^n v_j \sigma_{ji} = \bar{r}_i - r_f, \quad i \in \{1, ..., n\} \tag{3.55}$$

The new variable is defined by $v_i = c\beta_i$ where, in turn, $c$ is an unknown constant given by

$$c = \frac{\sum_{i=1}^n \beta_i (\bar{r}_i - r_f)}{\sum_{i,j=1}^n \beta_i \beta_j \sigma_{ij}} \tag{3.56}$$

and the $\beta_i$ are from the optimal solution. Finally, we normalize each $v_i$ dividing by $\sum_{i=1}^n v_i$. □

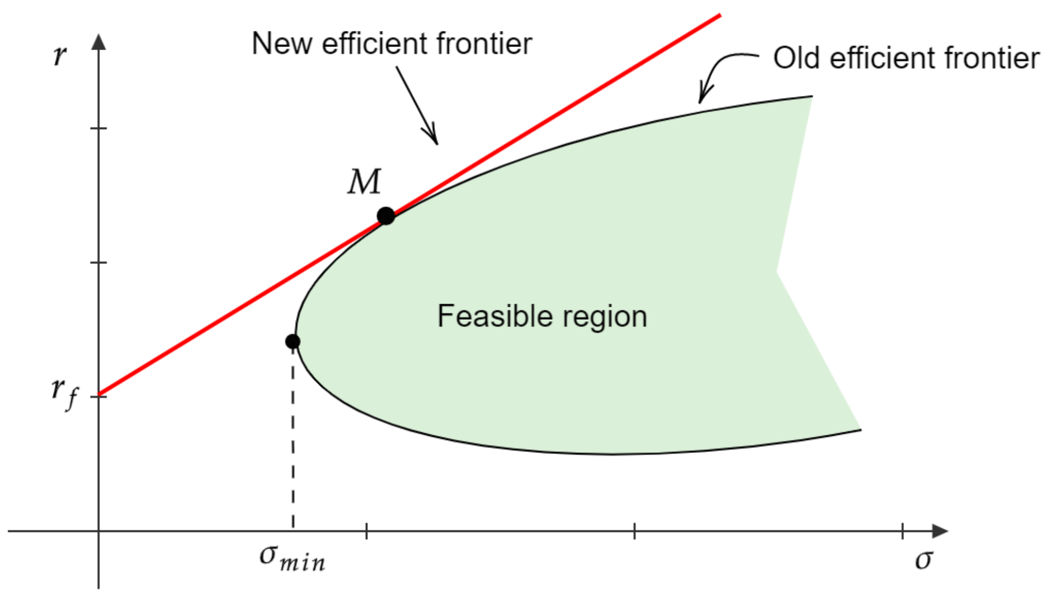



Figure 3.2: Sketch of the new efficient frontier after a risk-free asset is incorporated.

## 3.5 Limitations of the Markowitz model

For many years, this model was little more than an academic curiosity, scarcely used by investors and hedge funds, because of one fatal flaw that lies at its heart: the model is *very* sensible to small changes in expected returns. Throughout this chapter we've explained the origin, tools and main results in the Markowitz model but we've carefully left out all the limitations that this model has, which we will now discuss in this section.

Firstly, in this framework it is assumed that investors have some beliefs about the future performance of certain assets (i.e. expected returns) which in turn play a key role in the choice of the portfolio. Nevertheless, there is absolutely no discussion about how these beliefs are formed. In addition, regardless of how these beliefs are formed, the model requires that the investor has beliefs for *all* assets which make up the portfolio. If the portfolio is made up of a big number of assets then this requirement ceases to be reasonable; if the portfolio is made up of a small number of assets, even though the requirement is full filled, this defeats the purpose because diversification will be very limited.

Covariances among assets are fairly straightforward to estimate, but there is no straight forward way to estimate expected returns (at least in this framework). So, we have a model which is very sensible to changes in the expected returns, coupled with the fact that these are very difficult to estimate. In practice, this result in portfolios which are not very diverse, concentrating most of the funds in just a few assets. When shorting is allowed, portfolios obtained usually have massive shorting positions in certain assets which no rational investor would condone. This is a consequence of the unboundedness from above of the efficient frontier mentioned in section 3.3.

Many of the limitations of this model stem form its very simple nature; it is a single period, two factor model with one of the factors being very difficult to estimate and the other one being the simplest measure of risk. In subsequent sections we explore how some of these limitations have been dealt with over the years in order to construct new models based on the mean-variance analysis explained in this chapter.

# Chapter 4

# Capital asset pricing model (CAPM)

This model builds on the the ideas developed by Markowitz and introduces important new concepts such as *equilibrium*, *systemic risk* and an asset's *beta*. It was independently developed by Sharpe, Treynor, Lintner and Mossin. This model tries to estimate the appropriate expected returns for each asset. We will work assuming the existence of a risk-free asset for the sake of simplicity, but let it be noted that Black developed in 1972 a version of the CAPM which doesn't require this assumption and which has proven to be perform better when tested. This chapter is based on [14], [15], [6] and [7].

## 4.1 Utility functions

In the previous chapter, we explained what risk- aversion is using an example. In this section we will (briefly) formalize the mathematics behind this concept. Firstly, let $\Psi$ be some non-empty set which contains all possible choices an economic agent can make. Naturally, when presented with two choices, $x, y \in \Psi$, agents might have some *preferences*, which we formalize as

**Definition 4.1.1.** A *preference order* on $\Psi$ is a binary relation, represented by the symbol $\succ$, which satisfies the following properties:

1. Asymmetry: If $x \succ y$[1], then $y \not\succ x$.
2. Negative transitivity: if $x \succ y$ and $z \in \Psi$, then either $x \succ z$ or $z \succ y$ or both must be true.

In other words, negative transitivity means that if there is a clear preference between two choices and a third choice becomes available, there will still be a choice that is most preferable and one which is least preferable. Once we have a preference order on $\Psi$, a corresponding *weak preference* order ($\succcurlyeq$) and an *indifference* order ($\sim$) respectively defined as:

1. $x \succcurlyeq y \iff y \not\succ x$
2. $x \sim y \iff x \succcurlyeq y$ and $y \succcurlyeq x$

Weak preference is characterized by the following two properties:

1. *Completeness:* $\forall x, y \in \Psi$ either $y \succcurlyeq x$ or $x \succcurlyeq y$ or both hold. This means that it is always possible to determine whether one choice is preferred over another.
2. *Transitivity*: if $x \succcurlyeq y$ and $y \succcurlyeq z$ then $x \succcurlyeq z$. Transitivity implies that choices can be ordered in a hierarchy from best to worst, allowing for ties.

We now wish to quantify these preferences in some way. To this end we employ numerical representations, known as utility functions, and defined as:

[1] This should be read as "x is strictly preferred to y".

**Definition 4.1.2.** Given a preference order $\succ$, a *utility function* is a function $U : \Psi \to \mathbb{R}$ such that

$$x \succ y \iff U(x) > U(y) \tag{4.1}$$

This definition implies that

$$x \succcurlyeq y \iff U(x) \geq U(y) \tag{4.2}$$

Utility functions are not unique. This can be easily seen as follows: let $U$ be a utility function and take a function $f$ which is strictly increasing; if we define a a new function $\tilde{U}(x) := f(U(x))$, then $\tilde{U}$ will also be a utility function, as it satisfies the property stated above.

Utility functions are normally categorized in two families: *ordinal utility functions* and *cardinal utility functions.* In the former, the only relevant information is that, for example $U(x) > U(y)$ while in the latter, the difference between $U(x)$ and $U(y)$ is quantifiable and allows to reflect the *intensity* of a given preference. When dealing with cardinal utility functions relations as

$$U(x) - U(z) = 2(U(y) - U(z))$$

make sense and are meaningful. In this case, the previous equation states that an agent prefers $x$ over $z$ twice as much as he prefers $y$ to $z$.

**Definition 4.1.3.** Let $\succ$ be a preference relation and let $\Phi \subseteq \Psi$. We say $\Phi$ is *order dense* if $\forall x, y \in \Psi$ such that $x \succ y$ there exists some $z \in \Phi$ such that $x \succcurlyeq z \succcurlyeq y$.

Equipped with these notions, we now enunciate a theorem which characterizes those preference relations for which existence of a utility function is guaranteed.

**Theorem 4.1.1.** *Given a preference relation $\succ$ on a set $\Psi$, a necessary and sufficient condition for the existence of a utility function is the existence of a subset $\Phi \subseteq \Psi$ which is countable and order dense.*

*Proof.*
The proof for this theorem can be found in page 46 of [7]. □

Utility functions are used in many different areas of knowledge, and their characteristics may vary, but in the framework of quantitative finance they are generally taken to be:

- Strictly increasing
- Strictly concave
- Differentiable

and with derivative that vanishes at infinity and which explodes at 0. They can be functions of the form $u = u(\sigma, r)$ or $u = u(V)$ where $V$ is the value of a given portfolio. Some examples of readily used utility functions in finance include:

- Exponential utility: $u(V) = \frac{1-a^{-aV}}{a}$ for $a > 0$
- Logarithmic utility: $u(V) = \log V$

Exponential utility belongs to the family of *constant absolute risk aversion* utility functions while logarithmic utility belongs to the family of *hyperbolic absolute risk aversion* utility functions.

## 4.2 Geometric approach to the CAPM

In the in this section we will derive the main result of the CAPM, known as the beta formula. Firstly we will derive it using a more geometric (and perhaps intuitive) approach and afterwards in section 4.6, we will derive it, in a more technical fashion, as a necessary condition for there to be an *equilibrium.*

If we make the same assumptions listed in section 3.2 and in addition assume that every investor can lend/borrow unlimited quantities from a risk-free asset, we conclude that everyone is solving different Markowitz problems which in turn will produce portfolios on the efficient frontier. Given the *one fund theorem*, we know that the efficient frontier is a line that connects the risk free asset to the *market portfolio* which is tangent to the old efficient frontier. The efficient frontier, which we now rename as the *capital market line*, is given by

$$\bar{r} = r_f + \frac{\bar{r}_M - r_f}{\sigma_M}\sigma \tag{4.3}$$

The thing is that if everyone is essentially solving the same optimization problem, with the same information and in the same market, there should be some other way, different from the method explained in section 3.4, to arrive at this market portfolio. If we assume that the market ha reached an equilibrium, then the market portfolio will be given by the distribution of the market's capitalization. That is, the weights of each asset $i$ will be given by

$$w_i = \frac{V_i}{\sum_{j=1}^n V_k} \tag{4.4}$$

where $V_i$ is the capitalization of asset $i$ (i.e. the total value of asset $i$, that is, the number of total available shares of that asset times the value of each share) and the denominator is the total capitalization of the market.

Lets take a portfolio formed by an inefficient fund $i$ and the market portfolio $M$ with weights $(\alpha, 1-\alpha)$, $\alpha \in [0,1]$. Fund $i$ is assumed inefficient so our portfolio lies ina point somewhere in the feasible region but not in the efficient frontier. Its coordinates in the $(\sigma, \bar{r})$ plane are

$$\bar{r}(\alpha) = \alpha\bar{r}_i + (1-\alpha)\bar{r}_M \tag{4.5}$$
$$= \alpha(\bar{r}_i - \bar{r}_M) + \bar{r}_M \tag{4.6}$$

and

$$\sigma(\alpha) = \sqrt{\alpha^2\sigma_i^2 + (1-\alpha)^2\sigma_M^2 + 2\alpha(1-\alpha)\sigma_{Mi}} \tag{4.7}$$
$$= \sqrt{\alpha^2(\sigma_i^2 + \sigma_M^2 - 2\sigma_{Mi}) + 2\alpha(\sigma_{Mi} - \sigma_M^2) + \sigma_M^2} \tag{4.8}$$

Notice that as $\alpha$ varies, the point moves in the plane tracing out a smooth curve parameterized by $\alpha$ which starts in the market portfolio for $\alpha = 0$ and finishes in the inefficient fund in $\alpha = 1$. The curve is contained in the feasible region and only touches the capital market line at $\alpha = 0$. Thus we conclude that at that point the curve is tangent to the capital market line and therefore must satisfy that

$$\left.\frac{d\bar{r}(\alpha)}{d\sigma(\alpha)}\right|_{\alpha=0} = \frac{\bar{r}_M - r_f}{\sigma_M} \tag{4.9}$$

Where the right hand side of the previous equation is the slope of the capital market line, as expressed in equation (4.3). If we apply the chain rule we can express the left hand side as

$$\frac{d\bar{r}(\alpha)}{d\sigma(\alpha)} = \frac{d\bar{r}(\alpha)/d\alpha}{d\sigma(\alpha)/d\alpha} \tag{4.10}$$

Where

$$\frac{d\bar{r}(\alpha)}{d\alpha} = \bar{r}_i - \bar{r}_M \tag{4.11}$$

and

$$\frac{d\sigma(\alpha)}{d\alpha} = \frac{\alpha(\sigma_i^2 + \sigma_M^2 - 2\sigma_{Mi}) + \sigma_{Mi} - \sigma_M^2}{\sqrt{\alpha^2(\sigma_i^2 + \sigma_M^2 - 2\sigma_{Mi}) + 2\alpha(\sigma_{Mi} - \sigma_M^2) + \sigma_M^2}} \tag{4.12}$$

Evaluating the previous derivatives at $\alpha = 0$ and combining them yields

$$\frac{d\bar{r}(\alpha)}{d\sigma(\alpha)} = \frac{\sigma_M(\bar{r}_i - \bar{r}_M)}{\sigma_{Mi} - \sigma_M^2} \tag{4.13}$$

Now, equating the previous result with equation (4.9) we obtain

$$\frac{\sigma_M(\bar{r}_i - \bar{r}_M)}{\sigma_{Mi} - \sigma_M^2} = \frac{\bar{r}_M - r_f}{\sigma_M} \tag{4.14}$$

Finally, we only have to solve for the *risk premium*[2], that is $\bar{r}_i - r_f$, which yields

$$\bar{r}_i - r_f = \underbrace{\frac{\sigma_{Mi}}{\sigma_M^2}}_{\beta_i}(\bar{r}_M - r_f) \tag{4.15}$$

This equation is the ubiquitous beta equation which relates the expected return of any given fund (efficient or inefficient) with that of the market as a whole. Notice a few things:

- If the market portfolio has a positive price, and our portfolio is positively (negatively) correlated with it, the risk premium will be positive (negative).
- If our portfolio is uncorrelated with the market, the beta formula states that your portfolio's expected return should be equal to the risk-free asset's return.

Notice also that a negative risk premium means that we would get a lower expected return for that asset than for the risk-free asset. So, why would anyone invest in such an asset? Because of the negative correlation with the rest of the market! One can use such an investment to hedge the market and therefore it can be thought of as an insurance; in normal conditions this asset will have a very low expected return but if there is some sudden, unexpected crisis that causes the market to fall, the expected return of this asset will rise. A typical example of such an investment is gold.

So far we have discussed the beta of portfolios, but we can also talk about the beta of an individual asset as in the following example.

**Example.** *Suppose a given stock has $\beta = 2$, a stock specific volatility (i.e variance) of 2%, that yesterdays closing price was of 100€and that today the market goes up by 1%. We can use the beta, for example, to estimate the probability of our asset's closing price being, say, at least 103€.*

*For simplicity we will start by considering that there is no risk free asset so equation (4.81) adopts the simpler version*

$$\begin{aligned} E(r_{stock}) &= \beta E(r_M) \\ &= 2 \cdot 0.01 \\ &= 0.02 \end{aligned}$$

*If we now assume that stock returns are normally distributed [3], we can calculate the probability of the stock's closing price being at least 103€ as*

$$\begin{aligned} P(r_{stock} \geq 0.03) &= 1 - P(r_{stock} \leq 0.03) \\ &= 1 - \Psi\left(\frac{0.03 - 0.02}{0.02}\right) = 1 - \Psi\,(0.5) \\ &= 0.31 \end{aligned}$$

[2] Also known as *expected excess rate of return.*

[3] Which we do, again, for the sake of simplicity as it has been observed that stock returns are generally not normally distributed.

## 4.3 Pricing formula

Equipped with the beta formula that allows us to calculate the expected returns, the next natural step is to see if we can find a formula for the price of a given asset expressed in terms of the $\beta$, the market portfolio and the risk free asset. We start by considering an asset with with price $P = X_0$ at $t = 0$ and payoff $Q = X_1$ at $t = 1$. By virtue of definitions 3.1.2 and 3.1.3, the expected rate of return is

$$\bar{r} = \frac{E(X_1) - X_0}{X_0} = \frac{\bar{Q} - P}{P} \tag{4.16}$$

If we now solve for P we can express the price as the discounted payoff (i.e. the present value) with discount rate $\bar{r}$

$$P = \frac{\bar{Q}}{1 + \bar{r}} \tag{4.17}$$

We can now use the beta formula to substitute $\bar{r} = r_f + \beta(r_M - r_f)$ into the previous expression

$$P = \frac{\bar{Q}}{1 + r_f + \beta(\bar{r}_M - r_f)} \tag{4.18}$$

We now look for an explicit expression for the beta. Thus we first simplify the covariance

$$Cov(r, r_M) = Cov\left(\frac{Q - P}{P}, r_M\right) = Cov\left(\frac{Q}{P} - 1, r_M\right) \tag{4.19}$$

$$= Cov\left(\frac{Q}{P}, r_M\right) \tag{4.20}$$

$$= \frac{1}{P} Cov(Q, r_M) \tag{4.21}$$

We can therefore express the $\beta$ as

$$\beta = \frac{Cov(Q, r_M)}{P\sigma_M^2} \tag{4.22}$$

Substituting this into equation (4.18) we obtain

$$P = \frac{\bar{Q}}{1 + r_f + \frac{Cov(Q, r_M)}{P\sigma_M^2}} \tag{4.23}$$

Finally, we only have to solve for the price $P$ and we obtain

$$P = \frac{\bar{Q}}{1 + r_f} - \frac{Cov(Q, r_M)(r_M - r_f)}{\sigma_M^2(1 + r_f)} \tag{4.24}$$

Notice that we have arrived to an expression for the price where the first term in the right hand side is simply the discounted payoff (with discount rate $r_f$) and the second term is a sort of correction to this discounted payoff due to the correlation of the asset with the market. This correction yields a lower price (higher) for assets that are positively (negatively) correlated with the market. For uncorrelated assets, the correction term vanishes and therefore the price should simply be the discounted payoff.

## 4.4 Estimating the market portfolios and the betas

A common misconception, due to how the media talks about the financial markets, is that certain indexes (as the IBEX 35 or the S&P 500) *are* the stock markets, when in reality they only represent a given number of companies available in the stock markets. Take for example the IBEX 35; this index is composed of the 35 most liquid[4] companies from the Spanish stock market and it's calculated weighing each company's capital value[5] against the total capital value of said 35 companies.

Constructing the actual market portfolio is unfeasible because of the large number of companies that participate in the market and indexes, such as the ones previously mentioned, function as estimates of market portfolios. One of the most common ways of estimating the beta consists in using the method of *historical returns*. For this we choose N points in time such as, for example, the end of the last 10 years and calculate the average for an asset $A$ and for the $S\&P$ index

$$\hat{\bar{r}}_A = \frac{1}{N}\sum_{k=1}^{N} r_{Ak} \tag{4.25}$$

$$\hat{\bar{r}}_{S\&P} = \frac{1}{N}\sum_{k=1}^{N} r_{S\&Pk} \tag{4.26}$$

We now use these historical returns to estimate the variance and covariance needed for the beta

$$Var(S\&P) = \frac{1}{N-1}\sum_{k=1}^{N}(r_{S\&Pk} - \hat{\bar{r}}_{S\&P})^2 \tag{4.27}$$

$$Cov(S\&P, A) = \frac{1}{N-1}\sum_{k=1}^{N}(r_{S\&Pk} - \hat{\bar{r}}_{S\&P})(r_{S\&Pk} - \hat{\bar{r}}_A) \tag{4.28}$$

And finally we have our estimate of the beta:

$$\hat{\beta}_A = \frac{\sum_{k=1}^{N}(r_{S\&Pk} - \hat{\bar{r}}_{S\&P})(r_{S\&Pk} - \hat{\bar{r}}_A)}{\sum_{k=1}^{N}(r_{S\&Pk} - \hat{\bar{r}}_{S\&P})^2} \tag{4.29}$$

## 4.5 Systemic risk

Finally, we now introduce another important concept that arises in the CAPM: systemic risk. Now, disregarding the expected values, we express the expected return of an asset $i$ using equation (4.15) as

$$r_i = r_f + \beta_i(r_M - r_f) + \epsilon_i \tag{4.30}$$

Where we have introduced a random variable error term, $\epsilon_i$, defined as

$$\epsilon_i = r_i - r_f - \beta_i(r_M - r_f) \tag{4.31}$$

We will now study the relevant properties of our error term. We first compute its expected value

$$\begin{aligned} E(\epsilon_i) &= E(r_i - r_f - \beta_i(r_M - r_f)) && (4.32)\\ &= E(r_i) - \underbrace{E(r_f + \beta_i(r_M - r_f))}_{E(r_i)} && (4.33)\\ &= 0 && (4.34) \end{aligned}$$

[4]Liquidity refers to the capacity a company has of instantly transforming its assets into actual money without a significant loss of their value.

[5]The number of stocks each company has times the value of the stock.

and now the covariance of the error term and the market portfolio

$$\begin{aligned}
Cov(\epsilon_i, r_M) &= Cov(r_i - r_f - \beta_i(r_M - r_f), r_M) && (4.35)\\
&= Cov(r_i, r_M) - Cov(r_f, r_M) - \beta_i Cov(r_M, r_M) + \beta_i Cov(r_f, r_M) && (4.36)\\
&= Cov(r_i, r_M) - \frac{Cov(r_i, r_M)}{Var(r_M)} Var(r_M) && (4.37)\\
&= 0 && (4.38)
\end{aligned}$$

Taking both of these results into account, if we now take the variance on both sides of (4.30) we obtain

$$Var(r_i) = \beta_i^2 Var(r_M) + Var(\epsilon_i) \tag{4.39}$$

This important result shows that the risk of any asset can be decomposed into two orthogonal components known as the *systematic risk* ($\beta_i^2 Var(r_M)$) and the *non-systematic risk* ($Var(\epsilon_i)$). Only the latter of these two can be diversified away. Systematic risk represents the risk associated to events that affect to the market as a whole and which can't be avoided as, for example, does the Ukraine war nowadays. Thus, in the CAPM we have an improvement with respect to the Markowitz model, as we've managed to refine the concept of risk to make it more realistic.

Finally, lets analyze the risk of an efficient portfolio with coordinates $(\sigma_p, \bar{r}_p)$. From equation (4.3), we can solve for $\sigma_p$ and obtain

$$\sigma_p = \frac{\bar{r}_p - r_f}{\bar{r}_M - r_f} \sigma_M \tag{4.40}$$

Notice that the numerator is precisely the risk premium, so we can use the beta formula from equation (4.81) which yields

$$\sigma_p = \beta_p \sigma_M \tag{4.41}$$

Therefore, any efficient portfolio *only* has systemic risk. This brings us again to the one-fund theorem; any efficient portfolio can be decomposed as a weighted sum of the risk-free asset and the market portfolio. The former is deterministic so it adds no risk while the later is pure systemic risk as $\beta_M = 1$.

## 4.6 The CAPM as an equilibrium model

We will now deduce the beta formula as a necessary condition for there to be an equilibrium in an exchange economy with two dates. In this case we will consider that there are $m$ rational investors who can invest in a stock market comprised of $n$ different assets whose payoff are random variables denoted by $X_1^j$. We can assume these to be linearly independent without loss of generality. We will also consider that there is one single consumption good, external to the financial market, which we take as a numeraire[6]. At $t = 0$ we assume for simplicity that each investor $i$ starts with nothing and is allowed to construct a portfolio $\theta_i = (\theta_i^1, ..., \theta_i^n)$ [7] as long as he doesn't run into debt. Thus, at $t = 1$ each agent has wealth:

$$c_i = e_i + \sum_{j=1}^{n} \theta_i^j X_1^j \tag{4.42}$$

Where $e_i$ is a random variable representing the value at $t = 1$ of the consumption good and which we can interpret as the payoff of an initial portfolio. Okay now let $C$ be a finite dimensional vector space spanned by the random variables $(X_1^1, .., X_1^n)$ we endow it with the inner product defined by:

$$\langle c, c' \rangle = E(cc') \qquad \forall c, c' \in C \tag{4.43}$$

[6] A numeraire is a standard by which value is computed; a tradable good in terms of whose price the relative price of other tradable goods are expressed.

[7] Notice that in this case $\theta_i^j$ denotes the number of assets of type $j$ in the portfolio $i$ and should not be confused with the weights.

This naturally induces a norm $||c||_2 = E(c^2)$. We now make the following assumptions:

1. $e_i \in C \quad \forall i \in C$.
2. $e = \sum_{i=1}^{m} e_i \neq$ constant almost surely.[8]
3. There exists a risk-free asset such that $X_1^1 = 1$.

As in the Markowitz model, we assume that investors have preferences when choosing elements of $C$ which are reflected on the utility functions. Naturally, this preference can be no other than risk aversion.

**Definition 4.6.1.** A utility function $U_i$ is defined as $U_i : C \to \mathbb{R}$ such that for any pair $(c, c') \in C^2$ which satisfies $E(c) = E(c')$, the inequality $Var(c) < Var(c')$ implies that $U_i(c) > U_i(c')$.

Thus, if the prices of the assets are denoted with $S \in \mathbb{R}^n$, each investor $i$ will choose a portfolio $\theta_i = (\theta_i^1, ..., \theta_i^n)$ which is a solution to the optimization problem:

$$\text{Maximize} \quad U_i\left(e_i + \sum_{j=1}^{n} \theta_i^j X_i^j\right) \tag{4.44}$$

$$\text{subject to:} \quad S \cdot \theta_i \leq 0 \tag{4.45}$$

Where the inequality represents the condition that investors can't run into debt when forming the portfolio at $t = 0$. This is because of the assumption that we start with nothing at $t = 0$ and therefore, if we want to avoid running into debt, the only available portfolios for us will be those with negative value. At this point we introduce one of the central concepts of the CAPM: *the equilibrium.*

**Definition 4.6.2.** We say that a set of prices $\bar{S} \in \mathbb{R}^n$ and a set of portfolios $\bar{\theta}_i$, $i = 1, ..., m$ are in equilibrium if the following are satisfied:

- Each Portfolio $\bar{\theta}_i$ is a solution to the previous optimization problem.
- The security market clears, i.e. $\sum_{i=1}^{m} \bar{\theta}_i = 0$

**Lemma 4.6.1.** *With these assumptions there exists an equilibrium.*

*Proof.*
The proof can be found in [5]. □

This proof of existence relies on the assumption that there exists a risk-free asset; if we eliminate this assumption we will be find that existence of equilibrium is not always guaranteed.

Notice that by assumption, maximizing the utility function is the same as minimizing the variance when expected returns are fixed, so, at an equilibrium $(\bar{S}, \bar{\theta}_i; i = 1, ..., m)$, each investor $i$ will choose a portfolio $\theta_i$ which is a solution to the optimization problem

$$\text{Minimize} \quad Var\left(e_i + \sum_{j=1}^{n} \theta_i^j X_i^j\right) \tag{4.46}$$

$$\text{subject to:} \quad \bar{S} \cdot \theta_i \leq 0 \tag{4.47}$$

$$E\big(\underbrace{e_i + \sum_{j=1}^{n} \theta_i^j X_i^j}_{c_i}\big) = E\big(\underbrace{e_i + \sum_{j=1}^{n} \bar{\theta_i}^j X_i^j}_{\bar{c_i}}\big) \tag{4.48}$$

[8]The quantity $e$ is known as the *aggregate wealth.*

Where $c_i$ is the wealth at $t = 1$ as defined in (4.42). We now consider a linear functional $\bar{\varphi}$ defined on $C$ as

$$\bar{\varphi}(z_i) = \bar{S} \cdot \theta_i \tag{4.49}$$

Where

$$z_i = \sum_{j=1}^{n} \theta_i^j X_1^j \tag{4.50}$$

Notice that if $z_i$ is the value of the portfolio at $t = 1$ then $\bar{\varphi}(z_i)$ is its price at $t = 0$. Invoking Riesz's representation theorem we conclude that there exists $\varphi \in C$ such that

$$\bar{\varphi}(z_i) = \langle \varphi, z_i \rangle \tag{4.51}$$

We can now employ this functional to express our optimization problem in a more compact way. Firstly, taking into account that $c_i = e_i + z_i$, we can rewrite the no debt condition as

$$\begin{aligned} \langle \varphi, z_i \rangle = \langle \varphi, c_i - e_i \rangle = \bar{S} \cdot \theta_i \leq 0 \\ \langle \varphi, c_i \rangle - \underbrace{\langle \varphi, e_i \rangle}_{a_0} \leq 0 \\ \langle \varphi, c_i \rangle - a_0 \leq 0 \end{aligned}$$

Similarly, we can express the equality constraint as

$$\langle 1, c_i \rangle = \underbrace{\langle 1, \bar{c}_i \rangle}_{a_1} \tag{4.52}$$

$$\langle 1, c_i \rangle - a_1 = 0 \tag{4.53}$$

Thus, expressing the optimization problem in terms of the scalar product yields

$$\begin{aligned} \text{Minimize} \quad & ||c_i||_2 && (4.54) \\ \text{subject to:} \quad & \langle \varphi, c_i \rangle - a_0 \leq 0 && (4.55) \\ & \langle 1, c_i \rangle - a_1 = 0 && (4.56) \end{aligned}$$

We now construct our Lagrangian as usual, taking two Lagrange multipliers for every $i$; $\mu_i$ for the inequality constraint (with $\mu_i \geq 0$ because of the KKT condition known as dual feasibility) and $\lambda_i \in \mathbb{R}$ for the equality constraint. Thus the Lagrangian now adopts the form

$$L = \sum_{i=1}^{n} E\left(c_i^2\right) - \sum_{i=1}^{n} \lambda_i E\left(2(c_i - \bar{c}_i)\right) + \sum_{i=1}^{n} \mu_i E(2\varphi_i(c_i - e_i)) \tag{4.57}$$

$$= E\left(\sum_{i=1}^{n} c_i^2 - \sum_{i=1}^{n} \lambda_i 2(c_i - \bar{c}_i) + \sum_{i=1}^{n} \mu_i 2\varphi_i(c_i - e_i)\right) \tag{4.58}$$

If we now take derivatives with respect to $c_i$ and equate to zero we obtain

$$c_i = \lambda_i - \mu_i \varphi \tag{4.59}$$

and as this is the solution and we're assuming the existence of equilibrium we conclude that $c_i = \bar{c}_i$. Thus

$$\bar{c}_i = \lambda_i - \mu_i \varphi \quad \text{a.s.} \tag{4.60}$$

and as $\bar{c}_i = e_i + \sum_{j=1}^{n} \bar{\theta}_i^{\,j} X_i^j$ is in equilibrium, we conclude that there exist $\mu \geq 0$ and $\lambda \in \mathbb{R}$ such that

$$\sum_{i=1}^{m} \bar{c}_i = e = \lambda - \mu\varphi \quad \text{a.s.} \tag{4.61}$$

Recall that, by assumption, $e$ is non constant almost surely, and $\mu \geq 0$. Thus, for all $i$ there exists $a_i \geq 0$ and $b_i \in \mathbb{R}$ such that

$$\bar{c}_i = a_i e + b_i \tag{4.62}$$

This is another way of expressing the *one fund theorem* from section 3.4 which tells us that any portfolio can be created as a combination of the market portfolio and the risk-free asset. We can rewrite (4.61) as

$$\varphi = -\underbrace{\frac{1}{\mu}}_{a} e + \underbrace{\frac{\lambda}{\mu}}_{b} \tag{4.63}$$

Hence, our functional $\bar{\varphi}(z_i)$ now takes the form

$$\bar{\varphi}(z_i) = \langle \varphi, z_i \rangle \tag{4.64}$$

$$\bar{\varphi}(z_i) = \langle -ae + b, z_i \rangle = -a\langle e, z_i \rangle + b\langle 1, z_i \rangle \tag{4.65}$$

$$\bar{\varphi}(z_i) = -a\mathrm{Cov}(e, z_i) + bE(z_i) \tag{4.66}$$

We can now derive from this expression the equilibrium price for each asset as

$$\bar{\varphi}(z_i) = -a\mathrm{Cov}(e, z_i) + bE(z_i) = \bar{S} \cdot \theta_i$$

$$-a\mathrm{Cov}\left(e, \sum_{j=1}^{n} \theta_i^j X_1^j\right) + bE\left(\sum_{j=1}^{n} \theta_i^j X_1^j\right) = \sum_{j=1}^{n} \bar{S}^j \theta_i^j$$

$$\sum_{j=1}^{n} \left(-a\mathrm{Cov}(e, X_1^j) + bE(X_1^j)\right) \theta_i^j = \sum_{j=1}^{n} \bar{S}^j \theta_i^j$$

Which implies

$$\bar{S}^j = -a\mathrm{Cov}(e, X_1^j) + bE(X_1^j) \tag{4.67}$$

If we now take, for example, $b = \frac{1}{1+r}$ the previous formula becomes

$$\bar{S}^j = -a\mathrm{Cov}(e, X_1^j) + \frac{E(X_1^j)}{1+r} \tag{4.68}$$

from whence we deduce that the price of an asset that has positive correlation with the aggregate wealth is lower than its discounted expected return. We now give an alternative definition for the market portfolio.

**Definition 4.6.3.** We call *market portfolio* to $M \in \mathbb{R}^n$ such that

$$\sum_{j=1}^{n} M^j X_1^j = e \tag{4.69}$$

i.e. the portfolio whose value at $t = 1$ equals that of the aggregate wealth.

If we now express a portfolios return as

$$r_{\theta_i} = \sum_{j=1}^{n} \frac{\theta_i^j X_1^j}{\bar{S} \cdot \theta_i} \qquad \bar{S} \cdot \theta_i \neq 0 \tag{4.70}$$

Then for the market portfolio we simply have

$$r_M = \frac{e}{\bar{S} \cdot M} \tag{4.71}$$

At this point, collecting all the results we can finally deduce the famous beta formula which relates the expected return for assets with systematic risk of the market as a whole. We will relax the notation a bit and drop the sub indices $i$. We begin with formulas (4.49) and (4.66) with $b = \frac{1}{1+r}$:

$$\bar{S}\theta = -a\mathrm{Cov}(e, z) + \frac{E(z)}{1+r} \tag{4.72}$$

$$1 = -a\mathrm{Cov}\left(e, \frac{z}{\bar{S}\theta}\right) + \frac{E\left(\frac{z}{\bar{S}\theta}\right)}{1+r} \tag{4.73}$$

$$1 = -a\mathrm{Cov}\left(e, r_\theta\right) + \frac{E\left(r_\theta\right)}{1+r} \tag{4.74}$$

Now, solving for $E(r_\theta)$ we obtain

$$E(r_\theta) = (1+r)(1 + a\text{Cov}(e, r_\theta)) \tag{4.75}$$

Applying this result to the market portfolio we obtain

$$E(r_M) = (1+r)(1 + a\text{Cov}(e, r_M)) \tag{4.76}$$

and equating these two formulas we obtain

$$\frac{E(r_\theta) - (1+r)}{a\text{Cov}(e, r_\theta)} = \frac{E(r_M) - (1+r)}{a\text{Cov}(e, r_M)} \tag{4.77}$$

$$E(r_\theta) - (1+r) = (E(r_M) - (1+r)) \frac{\text{Cov}(e, r_\theta)}{\text{Cov}(e, r_M)} \tag{4.78}$$

$$= (E(r_M) - (1+r)) \frac{\text{Cov}(r_M, r_\theta)}{\text{Var}(r_M)} \tag{4.79}$$

We define the *beta* of a portfolio with respect to the market as:

$$\beta_\theta = \frac{\text{Cov}(r_M, r_\theta)}{\text{Var}(r_M)} \tag{4.80}$$

And finally obtain the beta formula:

$$E(r_\theta) - (1+r) = \beta_\theta \big(E(r_M) - (1+r)\big) \tag{4.81}$$

The left hand side of this equation is known as the *risk premium* of the portfolio.

## 4.7 Limitations of the CAPM

Although the CAPM presents some improvements with respect to the Markowitz model, it still has severe limitations both in the theoretical and in the practical plane. For example Richard Roll's critique centers around two problems he perceives in the CAPM. The first part of his critique is more philosophical in nature and is known as the *mean-variance tautology*[9]. This refers to the fact that mean-variance efficiency of the market is equivalent to the beta formula holding and thus, given an approximation for the market portfolio, testing the beta formula is equivalent to testing the mean-variance efficiency of said portfolio[10].

Secondly, Roll claims that the market portfolio is unobservable; not only should it include every possible asset available in the stock market, but also assets which are not equity, as real estate, precious metals, and basically anything worth anything that is available to be bought. This is, for obvious reasons, not feasible. Although in our case, as we are limiting ourselves to the study of financial markets, indexes such as the $S\&P500$ serve as fairly decent approximations.

Combining both points of the critique, he claims that the CAPM can't be empirically tested as its validity is equivalent to the market being mean-variance efficient with respect to all investment opportunities, but it's not possible to check if any portfolio is efficient without checking every investment opportunity available. In essence, for Roll the problem is that the model is, as the Markowitz model, too data intensive.

[9] A tautology is a statement that is true by virtue of its logical form alone. Examples of this include the phrase "either it will rain tomorrow or it won't rain" or the mathematical statement $1 = 1$.

[10] A proof of this equivalence can be found on [11]

# Chapter 5

# Multi-period models

Thus far, the market models we've considered (Markowitz model and CAPM) only allowed for one given period of time. In this chapter we will consider market models characterized by an arbitrary number of periods of time. We will introduce two similar models, the binomial and the trionamial market model, and explain the fundamental difference between them: *completeness*. We will explain the notion of market completeness and later see how this affects our optimization problem. Finally we will review one possible solution method known as "the martingale method" and work out an explicit example in both cases for a logarithmic utility function. This chapter is based mainly on the works by W. J. Rungaldier in [12] and Rutkowski in [13].

## 5.1 Problem definition and market completeness

The two key point around what everything else orbits in this chapter are arguably martingales and market completeness. Firstly, given a finite probability space $(\Omega, \mathcal{F}, P)$ with $\mathcal{F} = \mathcal{P}(\Omega)$, $P(\{\omega\}) > 0$ for all $\omega \in \Omega$ and given a filtration $(\mathcal{F}_n)_{0\leq n\leq N}$ such that $\mathcal{F}_0 = \{\emptyset, \Omega\}$ we have the following definitions:

**Definition 5.1.1.** A sequence of random variables $X = (X_n)_{0\leq n\leq N}$ is said to be *adapted* if $X_n$ is $\mathcal{F}_n$-measurable, $0 \leq n \leq N$.

**Definition 5.1.2.** A sequence of random variables $X = (X_n)_{0\leq n\leq N}$ is said to be *predictable* if $X_n$ is $\mathcal{F}_{n-1}$-measurable, $0 \leq n \leq N$.

**Definition 5.1.3.** An adapted sequence $(M_n)_{0\leq n\leq N}$ is said to be a *martingale* if

$$E(M_{n+1}|\mathcal{F}_n) = M_n$$

**Definition 5.1.4.** A risk-free asset is one whose price evolves according to

$$B_{t+1} = (1 + r)B_t \tag{5.1}$$

where $r \in \mathbb{R}^+$ is known as the short rate of interest[1].

Recall at this point that a probability measure $Q$ is equivalent to $P$ ($Q \sim P$) if they have the same null-sets. Thus we arrive at the following definition.

**Definition 5.1.5.** A market model is said to be *complete* if there exits a unique equivalent martingale measure (normally called the *risk-neutral* measure[2]) under which the expected values of the discounted price of assets are martingales, that is, $\exists! Q \sim P$ such that

$$E_Q\left(\frac{S^i_{t+1}}{B_{t+1}}|\mathcal{F}_t\right) = \frac{S^i_t}{B_t}, \qquad \forall t \in [0, T] \cap \mathbb{N} \tag{5.2}$$

[1] Generally the short rate of interest depends on time, but for our purpose we will take it without loss of generality as constant to simplify the computations that will follow.

[2] In some cases it also called the *equilibrium measure.*

where $B_t$ and $S_t$ are the price of the risk-free asset and the risky asset respectively. This can be equivalently expressed as

$$E_Q(S^i_{t+1}|\mathcal{F}_t) = \frac{B_{t+1}}{B_t} S^i_t = (1+r)S^i_t \tag{5.3}$$

If there are more than one equivalent martingale measures we say that market model is *incomplete*. The notion of completeness comes from a classic and well known problem in finance of hedging a contingent claim (a buy/sell option for example). In a complete market any derivative is replicable i.e. there exists and admissible strategy such that its final value is equal to the claim's payoff. In the context of no-arbitrage theory, the notion of market completeness is directly related to the existence (or lack thereof) of arbitrage opportunities through the *second fundamental theorem of asset pricing*. For more information in this subject we refer to [3] and [7].

We will consider an hyperbolic absolute risk aversion utility function of our portfolio's wealth given by

$$u(V) = \log V \tag{5.4}$$

and for the investment criteria we will consider the maximization of expected utility from terminal wealth:

$$\begin{cases} \underset{\alpha}{\text{Maximize}} \quad E[u(V_T^\alpha)] \\ \text{Subject to:} \begin{cases} V_0 = v \\ \alpha : \text{self-financing and predictable} \end{cases} \end{cases}$$

Where $\alpha$ denotes the optimal portfolio that will maximize the terminal wealth's expected utility. Self-financing refers here to the fact that any changes done in the portfolio must be carried out by selling other assets in the portfolio and not by injecting more money into it. We now formalize this idea in the following definition.

**Definition 5.1.6.** Given a portfolio $\alpha_t = (\alpha^0_t, \alpha^1_t)$ formed of one risk-free asset (whose price evolves according to $B_{t+1} = (1+r)B_t$) and one risky asset (whose price evolves according to $S_{t+1} = S_t\xi_{t+1}$) we say that it is *self-financing* if

$$V_{t+1} - V_t = \alpha^0_{t+1}\Delta B_t + \alpha^1_{t+1}\Delta S_t \tag{5.5}$$

or equivalently

$$\alpha^0_t B_t + \alpha^1_t S_t = \alpha^0_{t+1} B_t + \alpha^1_{t+1} S_t \tag{5.6}$$

Other investment criteria are also possible as the maximization of the expected utility from a consumption process or the maximization of expected utility from consumption *and* terminal wealth, but we do not consider these here.

## 5.2 The binomial market model

We shall consider a market model with $T$ periods of time, that is, $t \in [0,T] \cap \mathbb{N}$ and composed of one risk-free asset and one risky asset. After each period, there is a probability $p$ that the price of the risky asset will increase at a fixed rate $u$ and a probability $(1-p)$ that it will decrease at rate $d$. Therefore the price evolve as

$$S_{t+1} = S_t\xi_{t+1} \tag{5.7}$$

where $\xi_t$ are independent and identically distributed and such that

$$\xi_t = \begin{cases} u & \text{with probability} \quad p \\ d & \text{with probability} \quad 1-p \end{cases} \tag{5.8}$$

Recall now that given a finite probability space $(\Omega, \mathcal{F}, P)$, a Bernoulli process $X = \{X_t\}_{0\le t\le T}$ with parameter $p$ is a discrete random process, where $X_t$ are independent and identically distributed and there are only two possible outcomes such that

$$P(X_t = 1) = 1 - P(X_t = 0) = p \tag{5.9}$$

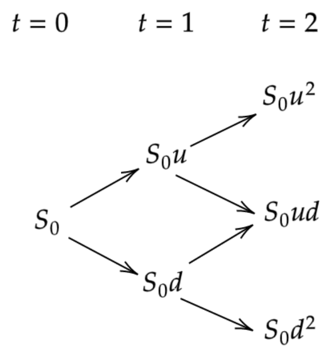



Figure 5.1: Binomial price tree for the risky asset.

Given such a Bernoulli process, we can also define a Bernoulli counting process $N = \{N_t\}_{0\leq t\leq T}$ as

$$N_t := \sum_{i=1}^{t} X_i \tag{5.10}$$

We can now naturally express our $\xi_t$ in terms of the Bernoulli process $X_t$ as

$$\xi_t = uX_t + d(1 - X_t) \tag{5.11}$$

And express the price of the risky asset in terms of the counting process $N_t$

$$S_t = S_0 u^{N_t} d^{t-N_t} \tag{5.12}$$

Notice that for every $t \in [0, T] \cap \mathbb{N}$, the random variable $N_t$ follows a binomial distribution with parameters $p$ and $t$. Therefore, the probability distribution of our risky asset's price $S_t$ is given by

$$P(S_t = S_0 u^n d^{t-n}) = \binom{t}{n} p^n (1-p)^{t-n} \tag{5.13}$$

Now the only thing we have left is to compute the equivalent martingale measure $q$. We define

$$q := Q(X_{t+1}|\mathcal{F}_t) \tag{5.14}$$

Now, from the martingale condition from equation (5.3) in definition 5.1.5 we have

$$E_Q(S_{t+1}|\mathcal{F}_t) = (1+r)S_t \tag{5.15}$$

$$E_Q(S_t\xi_{t+1}|\mathcal{F}_t) = (1+r)S_t \tag{5.16}$$

$$S_t E_Q(uX_{t+1} + d(1 - X_{t+1})|\mathcal{F}_t) = (1+r)S_t \tag{5.17}$$

$$qu + d(1-q) = 1 + r \tag{5.18}$$

And thus, solving for $q$ yields

$$q = \frac{1+r-d}{u-d} \tag{5.19}$$

Notice that $q \in [0, 1]$ and is unique as long a we have $d < 1 + r < u$[3] Finally, we wish to find an expression for the Radon-Nikodym derivative which will be useful later. To this end, recall the sample space $\Omega$ of the binomial model and taking $\omega \in \Omega$ with $\sum_{i=1}^{T} \omega_i = n$ we can write

[3]This is perfectly reasonable; if $1 + r > u$ there would be no point in investing in the risky asset and if $1 + r < d$ there would be arbitrage which is not allowed.

$$\begin{cases} P(\omega) = p^n(1-p)^{T-n} \\ Q(\omega) = q^n(1-q)^{T-n} \end{cases} \tag{5.20}$$

From where we can compute the Radon-Nikodym derivative on $\Omega$

$$L(\omega) = \frac{Q(\omega)}{P(\omega)} = \left(\frac{q}{p}\right)^n \left(\frac{1-q}{1-p}\right)^{T-n} \tag{5.21}$$

## 5.3 The trinomial market model

In this market model we again consider that it has $T$ periods of time and is composed of one risk-free asset and one risky asset whose price evolves obeying equation (5.7) but with the fundamental difference that $\xi_t$ is now given by

$$\xi_t = \begin{cases} u & \text{with probability} \quad p_1 \\ m & \text{with probability} \quad p_2 \\ d & \text{with probability} \quad p_3 \end{cases} \tag{5.22}$$

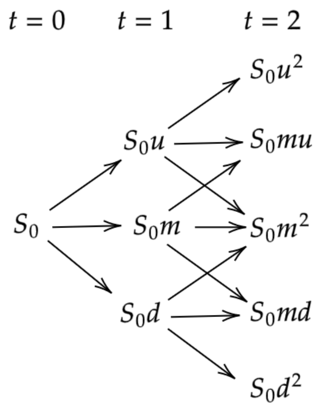



Figure 5.2: Trinomial price tree for the risky asset. For the sake of clearness this sketch represents the case where $u, d$ are such that $ud = m^2$.

We wish to have a representation for $\xi_t$ analogous to that expressed in equation (5.11), so we consider a discrete random process $X = \{X_t\}_{0\le t\le T}$, with the random variables $X_t$ independent and identically distributed, characterized by

$$\begin{cases} P(X_t = 1) = p_1 \\ P(X_t = 2) = p_2 \\ P(X_t = 3) = p_3 \end{cases} \tag{5.23}$$

for all $t \in [0, T] \cap \mathbb{N}$ and with $p_1 + p_2 + p_3 = 1$. In this case the sample space is $\Omega = \{\omega\}$ with $\omega$ of the form

$$\omega = \underbrace{(1, 2, 3, 1, 1, 2, 1, 3, ...)}_{\text{T times}}$$

We can express the probabilities in the measure $P$ as

$$P(\omega) = p_1^{n_1} p_2^{n_2} p_3^{n_3} \tag{5.24}$$

where

$$n_i = \sum_{t=1}^{T} 1_{\{\omega_t = i\}}, \quad i = 1, 2, 3 \tag{5.25}$$

We now define three counting processes analogous to that expressed in equation (5.10)

$$N_t^i := \sum_{s=1}^{t} 1_{\{X_s = i\}}, \quad \begin{cases} i = 1, 2, 3 \\ t \in [0, T] \cap \mathbb{N} \end{cases} \tag{5.26}$$

and taking $t = n_1 + n_2 + n_3$ we have

$$P(N_t^1 = n_1, N_t^2 = n_2, N_t^3 = n_3) = \frac{t!}{n_1! n_2! n_3!} p_1^{n_1} p_2^{n_2} p_3^{n_3} \tag{5.27}$$

Therefore, we now have the representation

$$\xi_t = u 1_{\{X_t = 1\}} + m 1_{\{X_t = 2\}} + d 1_{\{X_t = 3\}} \tag{5.28}$$

We can also express the price of the risky asset in terms of the counting processes from equation (5.26) as

$$S_t = S_0 u^{N_t^1} m^{N_t^2} d^{N_t^3} \tag{5.29}$$

The probability distribution of our risky asset's price is now given by

$$P(S_t = S_0 u^{n_1} m^{n_2} d^{n_3}) = \frac{t!}{n_1! n_2! n_3!} p_1^{n_1} p_2^{n_2} p_3^{n_3} \tag{5.30}$$

where again we must have that $t = n_1 + n_2 + n_3$.

We now wish to compute the equivalent martingale measures for this model which, unlike the previous one, is not complete. Recall that, given $Q \sim P$ on $\Omega$ such that condition (5.3) holds, the fact that $X_t$ are independent and identically distributed under $P$ does not imply that they are independent under $Q$. Therefore we define

$$q_i(t+1) := Q(X_{t+1} = i | \mathcal{F}_t); \quad i = 1, 2, 3 \tag{5.31}$$

Employing this notation, condition (5.3) adopts the form

$$(q_1(t+1)u + q_2(t+1)m + q_3(t+1)d)S_t = (1+r)S_t \tag{5.32}$$

Which, when paired with the fact that $q_1(t+1) + q_2(t+1) + q_3(t+1) = 1$, yields an under-determined system of equations with 3 unknowns and only 2 equations. Hence we conclude that the trinomial market model is incomplete. Notice that as the coefficients $u, m, d$ are independent of $t$ and $\mathcal{F}_t$, we have that the solution will also be independent of these and therefore $(q_1(t), q_2(t), q_3(t)) \equiv (q_1, q_2, q_3)$. These infinite solutions lie on a segment whose vertices depend on the value of $m$. To compute these extremal points we first check whether

$$m \begin{cases} \geq 1 + r \\ < 1 + r \end{cases}$$

Then the first vertex is then given by

$$(q_1^0, q_2^0, q_3^0) = \begin{cases} \left(0, \frac{1+r-d}{m-d}, \frac{m-(1+r)}{m-d}\right) & \text{for} \quad m \geq 1 + r \\ \left(\frac{1+r-m}{u-m}, \frac{u-(1+r)}{u-m}, 0\right) & \text{for} \quad m < 1 + r \end{cases} \tag{5.33}$$

where in each case, now that one of the probabilities is zero, we have a binomial model and the remaining probabilities are $(q, 1-q)$ where $q$ is as in equation (5.19) when the appropriate changes are made. The other vertex is then given by the analogous

$$(q_1^1, q_2^1, q_3^1) = \left(\frac{1+r-d}{u-d}, 0, \frac{u-(1+r)}{u-d}\right) \tag{5.34}$$

These probability measures are not equivalent to $(p_1, p_2, p_3)$ as the latter are all strictly positive. The equivalent measures are instead the convex combinations of these extremal measures i.e. the points that lie in the segment:

$$(q_1^\gamma, q_2^\gamma, q_3^\gamma) = \gamma(q_1^0, q_2^0, q_3^0) + (1-\gamma)(q_1^1, q_2^1, q_3^1), \qquad \gamma \in (0,1) \tag{5.35}$$

Now, the set of equivalent martingale measures is also a bounded convex set on $\Omega$ but finding its vertices is more complicated so we will limit ourselves to the case of $T = 2$ to illustrate the procedure. In such a case we have $\Omega = \{\omega\}$ with $\omega = (\omega_1, \omega_2)$ and $\omega_1, \omega_2 \in \{u, m, d\}$. Let's adopt the notation:

$$\begin{cases} \omega_t^1 = u \\ \omega_t^2 = m \\ \omega_t^3 = d \end{cases}$$

For a generic martingale measure $Q$ on $\Omega$ we have the expression

$$Q(\omega_1^i, \omega_2^j) = Q^1(\omega_1^i)Q^2(\omega_2^j|\omega_1^i) = Q^1(\omega_1^i)Q^2(\omega_2^j), \qquad i, j \in \{1,2,3\} \tag{5.36}$$

Where the conditional probability in the second period $Q^2$ is reduced to marginal probability due to the aforementioned independence of $(q_1, q_2, q_3)$ from $t$ and $\mathcal{F}_t$. We will now work with the specific case $m < 1 + r$. Letting $Q^{1,0}$ be the extremal probability measure, among all of the measures $Q^1$, that correspond to

$$(q_1^0, q_2^0, q_3^0) = \left(\frac{1+r-m}{u-m}, \frac{u-(1+r)}{u-m}, 0\right) \tag{5.37}$$

$Q^{1,1}$ be the other extremal probability measure corresponding to

$$(q_1^1, q_2^1, q_3^1) = \left(\frac{1+r-d}{u-d}, 0, \frac{u-(1+r)}{u-d}\right) \tag{5.38}$$

And proceeding in an analogous fashion for $Q^{2,0}$ and $Q^{2,1}$, we have that $Q^1$ and $Q^2$ are respectively convex combinations of these extremal measures. The factor $\gamma$ in these convex combinations may in theory depend on time and on $\omega$ itself (in a predictive way at least i.e. $\gamma = \gamma_t = \gamma(t; \omega_1, ..., \omega_{t-1})$), but in this case we can safely let gamma depend only on time because of the fact that $(q_1, q_2, q_3)$ are independent from $t$ and $\mathcal{F}_t$. Therefore we have

$$\begin{cases} Q^1(\omega_1^i) = \gamma_1 Q^{1,0}(\omega_1^i) + (1-\gamma_1)Q^{1,1}(\omega_1^i) \\ Q^2(\omega_2^j) = \gamma_2 Q^{2,0}(\omega_2^j) + (1-\gamma_2)Q^{2,1}(\omega_2^j) \end{cases} \tag{5.39}$$

Hence, we now arrive at

$$\begin{aligned} Q(\omega) &= Q(\omega_1^i, \omega_2^j) \\ &= Q^1(\omega_1^i)Q^2(\omega_2^j) \\ &= \gamma_1\gamma_2 Q^{1,0}(\omega_1^i)Q^{2,0}(\omega_2^j) + \\ &\quad + (1-\gamma_1)\gamma_2 Q^{1,1}(\omega_1^i)Q^{2,0}(\omega_2^j) + \gamma_1(1-\gamma_2)Q^{1,0}(\omega_1^i)Q^{2,1}(\omega_2^j) + \\ &\quad + (1-\gamma_1)(1-\gamma_2)Q^{1,1}(\omega_1^i)Q^{2,1}(\omega_2^j) \end{aligned} \tag{5.40}$$

and we end up with four extremal measures:

$$\begin{cases} \bar{Q}^1 = Q^{1,0}Q^{2,0} \\ \bar{Q}^2 = Q^{1,1}Q^{2,0} \\ \bar{Q}^3 = Q^{1,0}Q^{2,1} \\ \bar{Q}^4 = Q^{1,1}Q^{2,1} \end{cases}$$

Finally, we arrive at the following expression for the Radon-Nikodym derivatives

$$L^j(\omega) = \frac{\bar{Q}^j(\omega)}{P(\omega)} \tag{5.41}$$

Where $j = 1, ..., 4$ and $P(\omega) = P(\omega_1^i, \omega_2^j) = p_i p_j$.

## 5.4 The martingale method

Although this method has its origin in the hedging of a contingent claim problem, it is essentially a martingale representation problem. The aim is to determine the optimal reachable portfolio value $V_T^*$ and then find a portfolio such that the value of the portfolio is always a martingale.It consists on the following 3 steps:

1. Determine the set of reachable values for the wealth $V_T$ in the period $T$.
2. Determine the optimal reachable wealth $V_T^*$ in this set.
3. Determine the self financing strategy $\alpha^*$ whose terminal wealth is the same as the optimal value from the previous step, i.e. $V_T^\alpha = V_T^*$.

### 5.4.1 Complete market case

**Step 1**

The set of reachable portfolios is formed by those for which their discounted wealth's expected value equals to $v$, i.e.

$$\mathcal{V}_v = \{V : E_Q(B_T^{-1}V) = v\} \tag{5.42}$$

**Step 2**

We have to solve the problem of finding $V^*$ such that

$$E(u(V^*)) \geq E(u(V)), \quad \forall V \in \mathcal{V}_v \tag{5.43}$$

The condition $V_0 = v$ is equivalent to $E_Q(B_T^{-1}V) = v$ and thus the the problem can be reformulated as the following optimization problem

$$\text{Maximize} \quad E(u(V)) \tag{5.44}$$

$$\text{subject to:} \quad E_Q(B_T^{-1}V) = v \tag{5.45}$$

To solve this problem we employ Lagrange multipliers once again. Letting $L := \frac{dQ}{dP}$ be the Radon-Nikodym derivative and $\lambda$ be the Lagrange multiplier, the Lagrangian now adopts the form

$$L = E(u(V)) - \lambda(E_Q(B_T^{-1}V) - v) \tag{5.46}$$

$$= E[u(V) - \lambda L B_T^{-1}V] - \lambda v \tag{5.47}$$

Now, we simply take the derivative of the Lagrangian with respect to the wealth $V$ and equate it to zero to obtain

$$u'(V) = \lambda B_T^{-1}L \tag{5.48}$$

Recalling the properties of utility functions from section 4.1, we know that the inverse of $u'(\cdot)$ exists and we denote it by $I(\cdot) = (u'(\cdot))^{-1}$. Employing this function we solve for $V$ in the previous expression to obtain

$$V = I(\lambda B_T^{-1}L) \tag{5.49}$$

We can now rewrite constraint (5.47) in terms of $I$

$$E_Q[B_T^{-1}I(\lambda B_T^{-1}L)] = v \tag{5.50}$$

which is the same as

$$v = E[LB_T^{-1}I(\lambda B_T^{-1}L)] := V(\lambda) \tag{5.51}$$

This implies that, if $V(\cdot)$ is invertible, $\lambda = V^{-1}(v)$ and thus we obtain

$$V^* = I(V^{-1}(v)B_T^{-1}L) \tag{5.52}$$

As in our case we're working with the log-utility function ($U(V) = \log V$), we have that $I(x) = \frac{1}{x}$ and therefore constraint (5.47) becomes

$$v = E[LB_T^{-1}I(\lambda B_T^{-1}L)] = E[LB_T^{-1}I(\lambda B_T L^{-1})] = \frac{1}{\lambda} = V(\lambda) \tag{5.53}$$

This implies that $\lambda = \frac{1}{v}$ and thus we have that the optimal wealth is

$$V^* = I(\lambda B_T^{-1}L) \tag{5.54}$$
$$= \frac{vB_T}{L} \tag{5.55}$$
$$= v\left(\frac{q}{p}\right)^{-N_T}\left(\frac{1-q}{1-p}\right)^{N_T-T} \tag{5.56}$$

Where in the last step we have substituted the value already computed for $L$ from equation (5.21), $N_T \sim b(T, p)$ is the Bernoulli counting process that records the total number of "up-movements" and we have taken $B_T = 1$ in order to simplify the last step of the method. We can also now compute an explicit value for the expected utility from terminal wealth:

$$E[u(V^*)] = E[\log(V^*)] \tag{5.57}$$
$$= E\left[\log\left(v\left(\frac{q}{p}\right)^{-N_T}\left(\frac{1-q}{1-p}\right)^{N_T-T}\right)\right] \tag{5.58}$$
$$= \log v - \log\left(\frac{q}{p}\right)E(N_T) - \log\left(\frac{1-q}{1-p}\right)(T - E(N_T)) \tag{5.59}$$
$$= \log v - pT\log\left(\frac{q}{p}\right) - (1-p)T\log\left(\frac{1-q}{1-p}\right) \tag{5.60}$$

Finally, we check that this value for $V^*$ is indeed optimal with the help of the convex dual. This only works if we're working with a complete market model. To this end we employ the Legendre-Fenchel transform for which we give the definition below.

**Definition 5.4.1.** Given a function $U(x)$, its convex dual $\tilde{U}(y)$ is given by its Legendre-Fenchel transform:

$$\tilde{U}(y) := \max_x\{U(x) - xy\} \tag{5.61}$$

In our case, employing function $I$ defined in the previous page, we have

$$\tilde{U}(y) := \max_x\{U(x) - xy\} = U(I(y)) - yI(y) \tag{5.62}$$

From here it follows that

$$U(I(y)) - yI(y) \geq U(x) - xy, \quad \forall x \tag{5.63}$$

Recall now that

$$\begin{cases} V^* = I(\lambda B_T^{-1}L) \quad \text{with} \quad \lambda \quad \text{such that} \\ E(LB_T^{-1}V^*) = E(LB_T^{-1}I(\lambda B_T^{-1}L)) = v \end{cases}$$

If we now put $x = V, y = \lambda B_T^{-1}L$ with $\lambda \geq 0$. Thus, we obtain

$$U(V^*) - \lambda B_T^{-1}LV^* \geq U(V) - \lambda B_T^{-1}LV \tag{5.64}$$

Now, taking expectations on both sides yields

$$E[U(V^*)] - \lambda v \geq E[U(V)] - \lambda v \tag{5.65}$$

from which we conclude that

$$E[U(V^*)] \geq E[U(V)] \tag{5.66}$$

Which of course holds $\forall V$ that satisfies constraint (5.47).

**Step 3**

Finally, once obtained the optimal reachable wealth $V^*$ we will compute the optimal portfolio $\alpha_t^*$ that leads to this optimal wealth. As we're considering a binomial market model with one risk-free asset and one risky asset, our portfolio $\alpha_t$ will be a two-dimensional vector of the form

$$\alpha_t = (\alpha_t^0, \alpha_t^1) \tag{5.67}$$

where the first and second components stands for the number of units of the risk-free asset and the risky asset respectively. We will use backwards recursion to exploit the fact that the portfolio $\alpha_t$ is predictable (i.e. $\mathcal{F}_{T-1}$-measurable) and compute its components at each period from $t = T-1$ to $t = 0$.

We thus start at period $t = T-1$ and we have to determine the values for $\alpha_T^0, \alpha_T^1$. We want the final value of our portfolio to be the optimal reachable value computed in the previous step regardless on weather the price increases or decreases, that is

$$\alpha_T^1 S_T + \alpha_T^0 = V^* \tag{5.68}$$

Letting $N_{T-1} = n < T$ and recalling expressions (5.7) and (5.8) we arrive at the following system of equations which $\alpha_T^0, \alpha_T^1$ must verify

$$\alpha_T^1 S_{T-1} u + \alpha_T^0 = v \left(\frac{p}{q}\right)^{n+1} \left(\frac{1-p}{1-q}\right)^{T-n-1} \tag{5.69}$$

$$\alpha_T^1 S_{T-1} d + \alpha_T^0 = v \left(\frac{p}{q}\right)^{n} \left(\frac{1-p}{1-q}\right)^{T-n} \tag{5.70}$$

Where these equations represents the case where the price goes up and down respectively. From here we obtain

$$\alpha_T^1 S_{T-1}(u-d) = v \left(\frac{p}{q}\right)^{n} \left(\frac{1-p}{1-q}\right)^{T-n-1} \left(\frac{p}{q} - \frac{1-p}{1-q}\right) \tag{5.71}$$

and finally

$$\begin{cases} \alpha_T^1 = \frac{\left(\frac{p}{q}\right)^n \left(\frac{1-p}{1-q}\right)^{T-n-1}(p-q)}{S_{T-1}(u-d)q(1-q)} \\ \alpha_T^0 = \frac{\left(\frac{p}{q}\right)^n \left(\frac{1-p}{1-q}\right)^{T-n-1}\left(u(1-p)q - d(1-q)p\right)}{(u-d)q(1-q)} \end{cases} \tag{5.72}$$

Now, for the previous period $t = T-1$ and recalling the self financing condition we have that the optimal reachable wealth will be

$$V_{T-1}^* = \alpha_{T-1}^1 S_{T-1} + \alpha_{T-1}^0 \tag{5.73}$$

$$= \alpha_T^1 S_{T-1} + \alpha_T^0 \tag{5.74}$$

$$= \left(\frac{p}{q}\right)^{n} \left(\frac{1-p}{1-q}\right)^{T-n-1} \underbrace{\frac{\left(u(1-p)q - d(1-q)p + (p-q)\right)}{(u-d)q(1-q)}}_{=1} \tag{5.75}$$

$$= \left(\frac{p}{q}\right)^{n} \left(\frac{1-p}{1-q}\right)^{T-n-1} \tag{5.76}$$

Where we we have simplified the factor using the fact that $q = \frac{1-d}{u-d}$. As $V_{T-1}^*$ has the same structure as $V_T^*$, the calculations needed for the period $t = T-2$ are exactly the same as the ones here presented, and the same applies for each period until we reach $t = 0$. Therefore, we conclude that for a generic period $t \leq T$ with $N_t = n \leq t$ equation (5.68) becomes

$$\alpha_t^1 S_T + \alpha_t^0 = V_t^* \tag{5.77}$$

with

$$V_t^* = v \left(\frac{p}{q}\right)^{n} \left(\frac{1-p}{1-q}\right)^{t-n} \tag{5.78}$$

from where we obtain our portfolio's components

$$\begin{cases} \alpha_t^1 = \frac{\left(\frac{p}{q}\right)^n \left(\frac{1-p}{1-q}\right)^{t-n-1}(p-q)}{S_0 u^n d^{t-n-1}(u-d)q(1-q)} \\ \alpha_t^0 = \frac{\left(\frac{p}{q}\right)^n \left(\frac{1-p}{1-q}\right)^{t-n-1}\left(u(1-p)q-d(1-q)p\right)}{(u-d)q(1-q)} \end{cases} \tag{5.79}$$

Finally we express our portfolio in terms of its weights:

$$\begin{cases} \pi_t^0 &= \frac{\alpha_t^0}{V_t^*} = \frac{u(1-p)q-d(1-q)p}{(u-d)q(1-q)} \\ \pi_t^1 &= \frac{\alpha_t^1 S_t}{V_t^*} = \frac{p-q}{(u-d)q(1-q)} \end{cases} \tag{5.80}$$

Notice that they are independent of $t$ and of $S_t$. This however does not mean that the portfolio remains constant in time; as the price of the risky asset is changing every period, the amount invested in each asset must also change in each period so that the weights remain constant.

### 5.4.2 Incomplete market case

**Step 1**

In this case, as the market is incomplete, we have that the set of all martingale measures forms a bounded convex set with a finite number of vertices $J$. Therefore we have a $J$ extremal measures $Q^j, j = 1, ..., J$ and the set of reachable portfolio values is

$$\mathcal{V}_v = \{V : E_{Q^j}(B_T^{-1}V) = v \quad \text{for} \quad j = 1, ..., J\} \tag{5.81}$$

**Step 2**

Again we have to solve the problem of finding $V^*$ such that

$$E(u(V^*)) \geq E(u(V)), \quad \forall V \in \mathcal{V}_v \tag{5.82}$$

This time, as we have $J$ extremal measures $Q^j$ our problem's constraint adopts the form

$$E_Q(B_T^{-1}V) = v, \quad \forall Q \tag{5.83}$$

Where $Q$ is any martingale measure. As we already established that these martingale measures are convex combinations of the extremal measures $Q^j$, this constraint is equivalent to the following system of $J$ constraints

$$E_{Q^j}(B_T^{-1}V) = v, \quad \text{for} \quad j = 1, ..., J \tag{5.84}$$

Therefore, our optimization problem now adopts the form

$$\text{Maximize} \quad E(u(V)) \tag{5.85}$$
$$\text{subject to:} \quad E_{Q^j}(B_T^{-1}V) = v, \quad \text{for} \quad j = 1, ..., J \tag{5.86}$$

Letting $L^j := \frac{dQ^j}{dP}$, for $j = 1, \ldots, J$ be the Radon-Nikodym derivatives and $\lambda_j$ be the Lagrange multiplier corresponding to the *j-th* constraibt, the Lagrangian now adopts the form

$$L = E(u(V)) - \sum_{j=1}^{J} \lambda_j (E_{Q^j}(B_T^{-1}V) - v) \tag{5.87}$$
$$= E[u(V) - \sum_{j=1}^{J} \lambda_j L^j B_T^{-1} V] - \sum_{j=1}^{J} \lambda_j v \tag{5.88}$$

Now, we simply take the derivative of the Lagrangian with respect to the wealth $V$ and equate it to zero to obtain

$$u'(V) = \sum_{j=1}^{J} \lambda_j B_T^{-1} L^j \tag{5.89}$$

Which implies

$$V = I\left(\sum_{j=1}^{J}\lambda_j B_T^{-1}L^j\right) \tag{5.90}$$

Where recall that $I(\cdot) = (u'(\cdot))^{-1}$. We can now express the $J$ constraints from (5.88) as

$$E\Bigg[L^j B_T^{-1} I\left(\sum_{j=1}^{J}\lambda_j B_T^{-1}L^j\right)\Bigg] = v, \quad \text{for} \quad j = 1, ..., J \tag{5.91}$$

Finally, we compute the optimal wealth for the case $T = 2$ (which we already explored in section (5.3) and we assume again that $B_t = 1$. In this case we had four extremal measures $\bar{Q}^j$, $j = 1, ..., 4$. and the optimal reachable wealth is

$$V^* = \frac{1}{\lambda_1 L^1 + \lambda_2 L^2 + \lambda_3 L^3 + \lambda_4 L^4} \tag{5.92}$$

Here the Lagrange multipliers $\lambda_1, ..., \lambda_4$ are computed solving the following system of equations

$$\begin{aligned} v &= E\left(\frac{L^j}{\lambda_1 L^1 + \lambda_2 L^2 + \lambda_3 L^3 + \lambda_4 L^4}\right) \\ &= E\left(\frac{\bar{Q}^j}{\lambda_1 \bar{Q}^1 + \lambda_2 \bar{Q}^2 + \lambda_3 \bar{Q}^3 + \lambda_4 \bar{Q}^4}\right), \quad \text{for} \quad j = 1, ..., 4 \end{aligned} \tag{5.93}$$

**Step 3**

We finally only have left to compute the optimal portfolio $\alpha_t^*$. As the market is incomplete, not all values for the terminal wealth can be obtained via a self-financing portfolio. Nevertheless, the set of constraints (5.86) imply that $E_Q({}_T^{-1}V^*) = v$ for any equivalent martingale measure $Q$. Therefore $V^*$ is actually replicable with a self-financing portfolio which starts from the initial wealth $V_0 = v$. To determine the composition of this portfolio (in the case $T = 2$) we impose that

$$\alpha_2^1 S_2 + \alpha_2^0 = V^* \tag{5.94}$$

holds true for two of the three possible outcomes and the third one will automatically satisfy it too. We finally arrive to a point where we are in the same situation as in the binomial model and the calculations needed to compute the optimal portfolio are the same.

Other solution methods exist as, for example, dynamic programming. This method, when used in continuous time, leads to the Hamilton-Jacobi-Bellman equation. For more details on this subject check [16] and [4].

# Chapter 6

# Optimization in the framework of conic finance

So far, we've always assumed (though not explicitly) the *law of one price* i.e. that there is only one price at which we are able to buy and sell assets at will. Although this is a useful abstraction which allows us to derive many important results, real markets actually have two prices, that is, the price at which the market is willing to buy (*bid*) and the price at which the market is willing to sell (*ask*). In this chapter we will give an introduction to the main concepts of this two-price theory (referred to as conic finance theory for reasons that will become clear shortly) and then apply them to study portfolio theory from this optics. This final chapter is based on Madam and Schoutens' work on [8].

## 6.1 Introduction to conic finance

In complete markets, such as the binomial model from the previous chapter, the risk associated to any derivative can be completely eliminated, as we know that given any derivative we can always construct a self-financing portfolio that will hedge it. One of the foundations of conic finance theory is that risk elimination is typically unattainable and that we must tolerate some level of risk exposure. To that end we will start by defining *coherent risk measures.*

**Definition 6.1.1.** Let $X$ be a random amount of money that *you*[1] have to pay. A *risk measure* is any functional $\rho$ that assigns a real number to $X$. We say a risk measure is *coherent* if it satisfies the following properties. Let $X, Y$ be random variables and $c \in \mathbb{R}$. Then,

1. Translativity: $\rho(X + c) = \rho(X) + c$
2. Sub-additivity: $\rho(X + Y) \leq \rho(X) + \rho(Y)$
3. Positive homogeneity: $\rho(cX) = c\rho(X)$
4. Monotonicity: if $P(X \leq Y) = 1$, then $\rho(X) \leq \rho(Y)$

*Remark.* In this context, constants such as $c$ in the previous definition should be regarded as monetary units.

*Remark.* Under monotonicity subadditivity is equivalent to convexity.

An example of such a coherent risk measure is $\rho(X)$ of the form

$$\rho(X) = \sup_{Q \in \mathcal{M}} E_Q(X) \tag{6.1}$$

Where $\mathcal{M}$ is non empty set of probability measures. Actually, Artzner et al. showed in [1] that any coherent risk measure on a finite set of states of nature is of this form.

[1] Hereafter *you* means the market.

As the name suggest, large values of $\rho(X)$ tell us that $X$ is a very risky investment. Suppose now $X$ represents a derivative‘s payoff, then potential large payouts need to be accounted for and you should regard the quantity $\rho(X)$ as the amount of cash that should be added as a "buffer" so that the risk of paying out these potential payoffs becomes acceptable.

Now, according to the translation invariance property of $\rho(X)$, if you receive $c$ to pay $X$ the new risk will be

$$\rho(X - c) = \rho(X) - c. \tag{6.2}$$

So, if you wish to eliminate the risk you should take $c = \rho(X)$. This could be though as the price *to buy* the payoff $X$ to the market. Thus, we conclude that the *ask* price of $X$ should be

$$\text{ask}(X) = \rho(X) \tag{6.3}$$

Suppose now the converse, that is that*you receive* $X$. Receiving $X$ is equivalent to paying $-X$. In such case, the risk of receiving $X$, which we denote by $\bar{\rho}$[2], will be

$$\bar{\rho}(X) = \rho(-X). \tag{6.4}$$

Now, it is easy to see that $\bar{\rho}(X)$ satisfies the same properties as $\rho(X)$, namely :

(i) If $X \leq Y$ then $\bar{\rho}(X) \geq \bar{\rho}(Y)$.

(ii) If $c$ is a deterministic amount of money then $\bar{\rho}(X + c) = \bar{\rho}(X) - c$.

(iii) $\bar{\rho}(X + Y) \leq \bar{\rho}(X) + \bar{\rho}(Y)$.

(iv) If $c \geq 0$, is a constant, $\bar{\rho}(cX) = c\bar{\rho}(X)$.

Proceeding as before, if you wish now to eliminate risk then $\bar{\rho}(X - c) = \bar{\rho}(X) + c = 0$, so $c = -\bar{\rho}(X) = -\rho(-X)$. In other words, you have to pay $-\rho(-X)$. This is the price to *sell* the payoff $X$ to the market. Thus, we conclude that the *bid* price of $X$ is given by

$$\text{bid}(X) = -\rho(-X) = -\text{ask}(-X). \tag{6.5}$$

**Definition 6.1.2.** A zero-cost cash flow is a payoff $X$ that is acceptable by markets a zero cost. That is

$$\bar{\rho}(X) \leq 0.$$

Notice that if $\bar{\rho}(X) \leq 0$ the market will pay something for $X$. In this case zero would be an acceptable price for $X$, though not the best bid price for $X$. If $X \geq 0$ it satisfies $\bar{\rho}(X) \leq 0$ since, by $(iv)$, $\bar{\rho}(0) = 0$, then by $(i)$ we obtain that $\bar{\rho}(X) \leq 0$. Therefore the set of zero-cost cash flow is a *cone* containing the non-negative random payoffs and hence the name Conic Finance.

[2]To clarify, $\rho(X)$ refers to the "risk of giving" and $\bar{\rho}(X)$ to the "risk of receiving". Both formulations are correct bu the distinction between them is not always made clear in the literature.

## 6.2 Portfolio managament in Conic Finance

According to Madam and Schoutens we consider two situations: the optimal conic portfolio when only buying stocks (longing) is allowed and when shorting is also allowed.

### 6.2.1 The optimal Conic Long-only Portfolios

We start by considering a single-period model. Let $R_i, i = 1, ..., n$ be the returns, over a given time horizon, of the corresponding assets $A^{(i)}$, $i = 1, ..., n$. Assume that their mean return (under the risk-neutral measure chosen by the market to price the assets $A^{(i)}$) is zero. Then, if we use

$$\rho(R_i) = \sup_{Q\in\mathcal{M}} \mathbb{E}_Q (R_i) \tag{6.6}$$

where $\mathcal{M}$ contains the risk-neutral measure, we have that

$$\text{bid}(R_i) = -\rho(-R_i) = \inf_{Q\in\mathcal{M}} \mathbb{E}_Q (R_i) \leq 0. \tag{6.7}$$

Then we can look for a portfolio $(a_i)_{1\leq i\leq n}$ where $a_i$ is the fraction of wealth invested in $A^{(i)}$. It is assumed that $a_i \geq 0$ (as we're considering long-only portfolios). The portfolio's return will be then given by

$$R_p = \sum_{i=1}^{n} a_i R_i, \tag{6.8}$$

Madam and Schoutens propose to maximize the diversity measure given by

$$\text{bid}(R_p) - \sum_{i=1}^{n} a_i \text{bid}(R_i) \tag{6.9}$$

subject to the constraints

$$\begin{cases} a_i \geq 0, \quad \text{for} \quad i = 1, ..., n \\ \sum_{i=1}^{n} a_i = 1. \end{cases} \tag{6.10}$$

To solve this problem, first notice that

$$\begin{aligned} \text{bid}(R_p) - \sum_{i=1}^{n} a_i \text{bid}(R_i) &= \sum_{i=1}^{n} a_i \text{ask}(-R_i) - \text{ask}(-R_p) \\ &= \mu_p - \tilde{c}(a) \end{aligned} \tag{6.11}$$

where we've taken

$$\mu_p; = \sum_{i=1}^{n} a_i \text{ask}(-R_i) \tag{6.12}$$

and

$$\tilde{c}(a) := \text{ask}\left(-\sum_{i=1}^{n} a_i R_i\right). \tag{6.13}$$

We can now define the efficient frontier as the minimum of $\tilde{c}(a)$ for a fixed $\mu_p$, and with the weights satisfying conditions (6.10). If we then graph$(\tilde{c}(a), \mu_p)$ we will be able to visualize the efficient frontier (see Figure 6.1). The optimal long-only portfolio is represented by the point in the efficient frontier where the tangent line has slope equal to one. In fact, by taking the derivative with respect to $\tilde{c}$ in (6.11)

$$\mu_p'(\tilde{c}) - 1 = 0. \tag{6.14}$$

Let $(\mu_p(\tilde{c}^*), \tilde{c}^*)$ be this point, then the tangent line is

$$\mu_p(\tilde{c}^*) - \mu_p(0) = \tilde{c}^*$$

and consequently $\mu_p(0)$ gives the maximum value of (6.11). The intersect of this line with the y-axis is the diversity gap $\text{bid}(R_p) - \sum_{i=1}^{n} a_i \text{bid}(R_i)$.

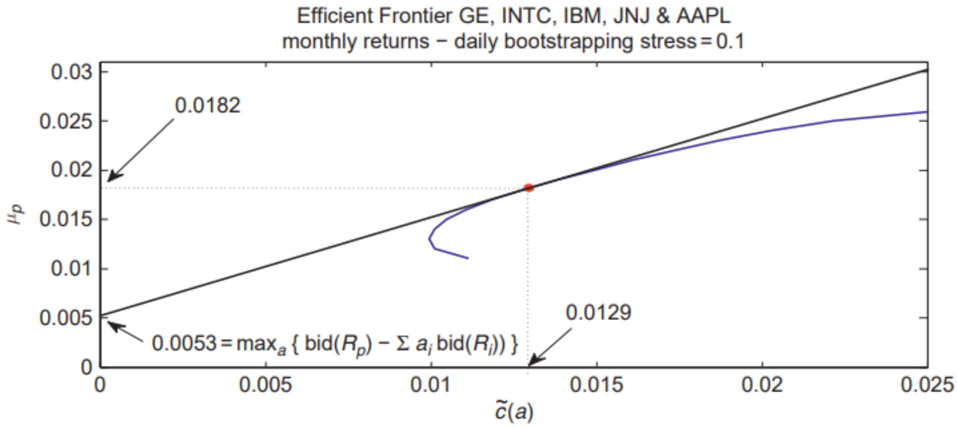



Figure 6.1: Example from [8] which portrays the efficient frontier for a long-only portfolio composed of stocks of General Electric, Intel, IBM, Johnson Johnson and Apple.

### 6.2.2 The optimal long-short portfolio with volatility constraint

In this case we allow for some shorting so not all the weight need to be positive. Thus, we eliminate the constraint $a_i \geq 0, i = 1, ..., n$. As the ask price optimization problem is well defined for a portfolio with a given mean, we can construct the efficient frontier as in the previous section but, without this constraint, the efficient frontier could have a slope greater than unit for all $\mu_p$ which would cause the unconstrained problem to not have a solution. In such a case Madam and Schoutens' propose bounding the volatility of the portfolio:

$$\sqrt{a^T \Sigma a} \leq \sigma^* \tag{6.15}$$

Where $\Sigma$ is the covariance matrix, $a$ the weight vector and $\sigma^*$ the target volatility. If the problem is well posed it is enough to maximize the bid price subject to $\sum_{i=1}^{n} a_i = 1$. An example of this situation can be seen in page 131 of [8]. Nevertheless the portfolio optimization in the context of Conic Finance is in my opinion still an open problem.

# Bibliography


[1] Artzner, P., Delbaen, F., Jean-Marc, E., and Heath, D. Coherent measures of risk. *Mathematical Finance 9* (07 1999), 203 – 228.

[2] Boyd, S., and Vandenberghe, L. *Convex optimization.* Cambridge university press, 2004.

[3] Capiński, M., and Kopp, E. *Discrete Models of Financial Markets.* Mastering Mathematical Finance. Cambridge University Press, 2012.

[4] Corcuera, J. M. *Quantitative Finance course notes.* 2022.

[5] Dana, R.-A. Existence, uniqueness and determinacy of equilibrium in capm with a riskless asset. *Journal of Mathematical Economics 32*, 2 (1999), 167–175.

[6] Dana, R.-A., and Jeanblanc-Picqué, M. *Financial markets in continuous time.* Springer, 2007.

[7] Hans, F., and Schied, A. *Stochastic finance an introduction in discrete time.* De Gruyter, 2016.

[8] Madan, D., and Schoutens, W. *Applied Conic Finance.* Cambridge University Press, 2016.

[9] Markowitz, H. Portfolio selection. *The Journal of Finance 7*, 1 (1952), 77.

[10] Markowitz, H. M. The early history of portfolio theory: 1600–1960. *Financial analysts journal 55*, 4 (1999), 5–16.

[11] Pollard, M. C. Proof of mean-variance and capm equivalence. Available at `http://www.matthewcpollard.com`.

[12] Runggaldier, W. *Portfolio optimization in discrete time.* Citeseer, 2006.

[13] Rutkowski M., M. M. *Martingale Methods in Financial Modelling*, second edition ed. 2005.

[14] Sharpe, W. F. Capital asset prices: A theory of market equilibrium under conditions of risk. *The Journal of Finance 19*, 3 (1964), 425.

[15] Sigman, K. Notes on financial engineering. Available at `http://www.columbia.edu/~ks20/FE-Notes/FE-Notes-Sigman.html`.

[16] Soner, H. M. *Stochastic optimal control in finance.* Scuola normale superiore, 2004.