Market Risk and Portfolio Theory

(For 2021/2022 UCL FM, FRM & CM MSc students only—not for general circulation)

December 27, 2022

University College London

Master Level Course

These notes are based on the course module MATH0094 Market Risk and Portfolio Theory of the Financial Mathematics MSc program of University College London.

For any mistakes, please e-mail camilo.garcia@ucl.ac.uk.
Please check Moodle regularly for updates of these notes, as they will change during the term. I will correct mistakes and add material and exercises throughout.

Contents

0.1 Background . . . 7

1 Fundamentals of market theory **10**

1.1 A market model in discrete time . . . 10

1.1.1 Structure of the market . . . 11

1.2 Some examples of market models . . . 15

1.2.1 Independent and identically distributed returns . . . 15

1.2.2 Markovian models . . . 17

1.2.3 Non-Markov models . . . 18

1.3 Arbitrage and completeness . . . 19

1.4 Stochastic discount factor (SDF) . . . 21

1.5 Risk neutral probability measure . . . 23

1.6 Arbitrage-free pricing of financial products . . . 26

1.6.1 Models under $\mathbb{Q}$ . . . 28

1.7 A deeper look at one-period models in a finite probability space . . . 28

1.7.1 Understanding the SDF in terms of Arrow-Debreu securities . . . 28

1.7.2 Matrix interpretation . . . 29

1.7.3 Proof of the F.T.A.P. . . . 30

1.8 Exercises . . . 33

1.9 Summary . . . 35

2 Utility functions **37**

2.1 Utility functions to model behaviour . . . 37

2.2 Risk attitude and certainty equivalence . . . 39

2.3 (Arrow-Pratt) coefficients of risk aversion . . . 40

2.3.1 Absolute coefficient of risk aversion . . . 41

2.3.2 Relative risk aversion coefficient . . . 41

2.4 Notable examples of utility functions 42
2.4.1 Constant absolute risk aversion (CARA) 42
2.4.2 Constant relative risk aversion (CRRA) 43
2.4.3 Hyperbolic absolute risk-averse utility functions (HARA) 43
2.4.4 Quadratic utility 44
2.5 Utility functions in a multi-period setting 44
2.5.1 Time additive utility 45
2.6 Axiomatisation of utility maximisation(*) 45
2.7 Exercises 47
2.8 Summary 48

3 Portfolio choice **49**
3.1 Optimisation problems 49
3.1.1 The concave case 50
3.2 Lagrangian optimisation method for constrained optimisation 54
3.3 The investor's choice problem - one-period 56
3.3.1 Existence and uniqueness of the optimal investment problem 58
3.3.2 First order conditions 59
3.3.3 Examples: single risky asset 61
3.3.4 Example: CARA with multivariate Normal returns 63
3.4 Exercises - One period case 65
3.5 The multi-period case 67
3.5.1 The portfolio-consumption choice problem 67
3.5.2 Euler condition for maximality (First Order Conditions) 69
3.5.3 Solving the multi-period case 71
3.6 Other multi-period problems in mathematical finance 74
3.6.1 The pricing and hedging problem 74
3.6.2 Pure-investment problem with transaction costs(*) 75
3.7 Exercises - Multiperiod case 76
3.8 Summary 77

4 Risk measures **79**
4.1 Risk measures and their properties 80
4.2 Utility-based risk measures 82
4.2.1 Simple loss 82
4.2.2 Certainty equivalent 82

4.2.3 Shortfall risk . . . 83
4.3 Tail-based risk measures . . . 83
4.3.1 Value at risk . . . 86
4.3.2 Expected shortfall . . . 87
4.4 Summary of properties . . . 90
4.5 Robust representation(*) . . . 90
4.6 Conditional and dynamic risk measure . . . 91
4.7 Risk mitigation . . . 92
4.7.1 Collateral addition . . . 93
4.7.2 Hedging . . . 93
4.8 Exercises . . . 94
4.9 Notes . . . 95
4.10 Summary . . . 96

5 Equilibrium and pricing **97**
5.1 Introduction to equilibrium . . . 97
5.1.1 Setup of the equilibrium problem . . . 97
5.1.2 CAPM and its proof in the case of CARA utilities . . . 99
5.2 Exercises . . . 101

6 Performance measurement and efficient frontiers in a one-period market model **102**
6.1 Risk and return . . . 102
6.2 Preliminaries . . . 103
6.3 The calculus approach in the case with no risk-free assets . . . 103
6.4 Two-fund spanning . . . 104
6.5 Mean-standard deviation trade-off . . . 105
6.6 Global minimum mean–variance portfolio . . . 106
6.7 Mean–variance optimization with risk-free asset . . . 106
6.7.1 Tangency portfolio . . . 108
6.8 Two-fund spanning revisited . . . 108
6.9 Extension to other risk measures . . . 108
6.10 Risk allocation and performance comparison . . . 109
6.10.1 RORAC . . . 110
6.10.2 Risk allocation . . . 110
6.11 Exercises . . . 111

6.12 Summary 113

7 A brief overview on asset types (*) **114**

7.1 Equity 114

7.2 Fixed income 115

7.2.1 Zero-coupon bonds 115

7.2.2 Coupon paying bonds 116

7.2.3 Yields 116

7.2.4 Forward rates 117

7.3 Foreign Exchange (FX) 117

7.3.1 Forward exchange rate 117

7.4 Derivatives 118

7.5 Commodities and other asset classes 118

8 Statistics review **119**

8.1 Hypothesis testing 119

8.1.1 State the null hypothesis H_0 and additonal assumptions 119

8.1.2 Choose statistics and test parameters 120

8.1.3 Compute test statistic and p-value 121

8.1.4 Decide whether to reject the null assumption or not 122

8.2 Linear regression 122

8.2.1 Coefficient estimation 123

8.2.2 Generalized Least-Squares (GLS) 123

8.3 Model selection 124

9 Factor models **125**

9.1 Factor Models 125

9.1.1 Linear models 126

9.2 Beta pricing 126

9.2.1 Single-factor models with returns as factors 127

9.2.2 Fama-French models 128

9.3 Exercises 129

10 Risk measures in practice **130**

10.1 The measurement cycle 130

10.2 Risk factors and mapping of risks 131

10.3 Risk measure estimation 132
10.3.1 Explicit model assumption: The closed-form case 132
10.3.2 Explicit model assumption: The Monte Carlo simulation case 134
10.3.3 Historical simulation 134
10.4 Back-testing 136
10.4.1 Framework description 136
10.5 Back-testing covering for V@R 137
10.5.1 Test 1: Assuming independence of failures 137
10.6 Backtesting ES 138
10.6.1 Z test on the quantile space 138
10.6.2 Z test on conditional expectation 138
10.6.3 Ellicitability 139
10.7 Exercises 139

Introduction

0.1 Background

The fast rate of innovation that the world has lived since the industrial and the digital revolutions, has only been possible thanks to the existence of a robust financial sector. Without the capital to finance new companies and endeavours (from the construction of train lines to funding the research leading to the personal computer and mobile phones), many of those discoveries would have not been possible or would have taken much longer than they did.

Indeed, let us look at some essential roles of the financial system:

- It allows the transfer of economic resources through **time** and between different people/companies/countries (ex: bonds)
- It provides ways to **manage risk** (ex: insurance, options)
- It contributes to the flow of economic information and provides means to establish **prices** (ex: Stock Exchanges)
- It eases trade by providing ways to clear and liquidate payments (ex: FX, credit cards)
- It provides mechanisms to aggregate and divide resources. (ex: shares)

These goals are achieved by exchanges involving **financial institutions**, **financial instruments** and **financial markets**:

- Financial institutions are the agents whose main purpose is to provide financial services and are therefore the main interacting agents (ex: banks, insurers, regulators,...).
- Financial instruments are the assets that are exchanged. They belong to three main classes: debts, equity, derivatives.
- Financial markets are the "places" where those exchanges happen and the "rules" governing those exchanges [1].

[1] The"actual" way markets "exist" is very involved and has a lot to do with computers, communications, algorithms and regulations. For an interesting way of having a partial view of equity and derivative markets in the US, read Michael Lewis' Flash Boys

As pointed out above, risk is an intrinsic element of financial markets and, in fact, of every business activity. Let us remind ourselves what this notion means. The Oxford English Dictionary offers the following definition

Definition 0.1. By **risk** we understand the possibility of *financial loss* or *failure* as a *quantifiable* factor in *evaluating* the potential profit in a commercial enterprise or investment.

We are interested in understanding the interaction of market participants with risks. In particular, we will focus on market risk, that is

Definition 0.2. Market risk is the type of risk associated to the uncertainty of values of financial instruments in financial markets.

For example, a bank would qualify as market risk the following risks:

1. Default risk, interest rate risk, credit spread risk, equity risk, foreign exchange risk and commodities risk for trading book instruments; and
2. Foreign exchange risk and commodities risk for banking book instruments

To illustrate what market risk means, we can look at Figure 1: it highlights the fall in the FTSE100 index around Black Monday (19.10.1987), where around 27% of market value of this index was lost in one day. The index would take more than one year to retrieve its previous value.

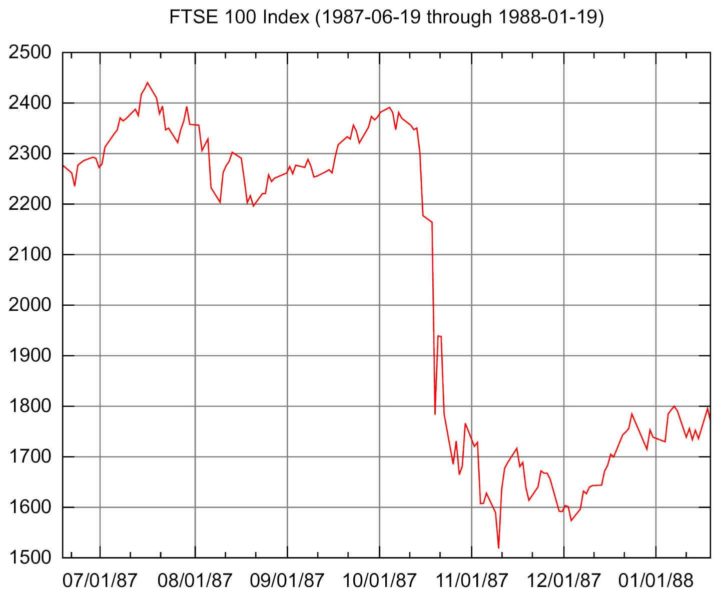

Figure 1: FTSE100 around the Black Monday. The index value fell nearly 27% in one day. Figure by Autopilot - Own work, CC BY-SA 3.0, https://commons.wikimedia.org/w/index.php?curid=9425883

We have explicitly asked in our definition for risks being *quantifiable*. We reserve another word *uncertainty* to potential losses that are not quantifiable. This is a very important property, because it means we can *measure* it. This allows us to define *risk measures*.

Another consequence of being able to quantify risks is that it allows for mathematical modelling: that is, to find a simplified representation of our object of study, in mathematical terms, with the aim of better understanding it and answer questions related to it.

In this course, we are interested in studying mathematical models of financial markets, with the aim to understand the effects of market risk, in particular those related to decision-making. To illustrate this idea, let us give some of examples of the main questions we will try to understand:

- How do we quantify risks? Can we do something about them?
- How to measure the performance of an investment? Are portfolios that are optimal in some performance sense?
- How to decide if we should access a financial instrument? For example, when should someone buy an insurance, or ask for a credit? More in general, is there a way to make optimal decisions?
- What should be the price of newly introduced financial assets?

Financial markets have evolved to be very complex entities comprising thousands of market participants of different types (banks, hedge funds, pension funds, insurers, ...), dealing with many types of instruments (stocks, bonds, FX, ...) through platforms where precise rules for matching operations and dealing with settlement requirements and regulations that can change from one country to the next. To obtain meaningful quantitative conclusions on the questions we raised, we introduce *market models*, a simplified **mathematical description of the participants** of the market, the **instruments** available and the **rules** under which they can trade them.

- We do not want to study specific market participants instruments and institutions. We rather create a (mathematical) abstract model that represents them all. In particular we do not want to explain *why* the values of a financial instrument are risky, but rather try to model their "unpredictable" changes. We are also interested in understanding the attitudes of participants.
- As with any model, we try to find a good balance between complexity and tractability: our purpose is to make simplifying assumptions that keep the results meaningful. In any case we must always remember that any results are obtained under the model, and have to be taken understanding its limitations. In many occasions the insights they produce are more important that their exact values.

Chapter 1

Fundamentals of market theory

All models are wrong, but some are useful

George Box

The financial markets generally are unpredictable. So that one has to have different scenarios... The idea that you can actually predict what's going to happen contradicts my way of looking at the market.

George Soros

In this chapter we consider the mathematical modelling of a financial market in discrete time with a finite horizon T. The market is composed of n assets that can be freely bought or sold. We model the asset total values as stochastic processes in a filtered probability space. One of the assets (identified by the index 0) is assumed to be locally risk-free.

1.1 A market model in discrete time

We assume that we are given a probability space $(\Omega, \mathbb{P}, \mathcal{F})$, where Ω is interpreted as the set of all possible outcomes in the market, $\mathcal{F}$ is the sigma algebra of all measurable events and $\mathbb{P}$ is the probability quantifying the likeliness of those events[1].

Let us introduce our characterisation of financial instruments in the market:

- We identify each asset by a number $0, 1, \ldots, n$. The number 0 will be reserved for a bank account asset (see Definition 1.8 below).

[1] See the lecture notes by Terence Tao on the subject (terrytao.wordpress.com/category/teaching/275a-probability-theory/) for a good refreshment in probability theory.

- Each asset $j = 0, \dots, n$ is completely characterised by a random process (a random vector in the finite horizon case) $\boldsymbol{S}^j = (S_0^j, \dots, S_T^j)$ that represents the *total market value of the given asset*. These values are expressed with respect to a common *numéraire* or currency. Moreover, we assume that the actions of a single agent on the market cannot by themselves change these values.
- We assume that investors can buy as many shares/instances of an asset as they want, in fractional quantities if they so desire. They can also short an asset as much as they want.

For convenience, we introduce also the following matrix-like notation: we write $\boldsymbol{S}_t = (S_t^0, \dots, S_t^n)^\top$, that is, a column vector with the random values at time t of each asset.

1.1.1 Structure of the market

An important part of a multi-period financial market is the information that is available at each time period. It then makes sense to add some structure to our probability space as to capture the flow of this information: we add a **filtration**, that is a sequence of sigma algebras $\{\mathcal{F}_t\}_{t\in\mathcal{I}}$, indexed by $\mathcal{I}$ (typically either $\mathbb{R}_+$, $\mathbb{N}$ or a finite interval starting at zero), that is increasing in the sense of inclusion, i.e.,

$$\mathcal{F}_s \subseteq \mathcal{F}_t \text{ for all } s \leq t \text{ in } \mathcal{I}.$$

Intuitively, the filtration at time $\mathcal{F}_t$ contains all the events that are decidable by time t. For example, assume that the event "The value of the asset i at time 2 is larger than £100" (which we would write as the set of outcomes for which $\{S_2^i > 100\}$), is something we **know** at time 2 and afterwards, so it belongs to $\mathcal{F}_t$ for all $t \geq 2$.

In the following, unless otherwise stated, we consider only filtrations defined over the set $0, 1, \dots, n$. This corresponds to the finite time discrete model case. Additionally, we make the following assumption

Assumption 1.1. The filtration $\{\mathcal{F}_t\}_{t=0,\dots,T}$ satisfies

$$\mathcal{F}_0 = \{\emptyset, \Omega\}; \quad \mathcal{F}_T = \mathcal{F}.$$

The above assumption simply says that $\mathcal{F}_0$-measurable functions are deterministic and in particular that the initial price of all assets is known. It also says that the whole set of outcomes is known by time T, which is a convenience assumption that signifies that the model ends at that time.

It is important to distinguish processes which are compatible with the information structure given by the filtration.

Definition 1.2. We say that a stochastic process is **adapted**, if X_t is $\mathcal{F}_t$-measurable for all t. We say that a stochastic process is **predictable**, if X_t is $\mathcal{F}_{t-1}$ measurable for all $t > 0$.

In our setup, the property of being *adapted* captures the intuition that processes are progressively discovered and they should not anticipate the future; likewise, the property of being *predictable* captures processes whose value at the end of a period is known at the beginning of it. Clearly, by the increasing nature of filtrations, any predictable process is adapted. It is then natural to make the following assumption

Assumption 1.3. The value processes S^i are adapted to the filtration $\{\mathcal{F}_t\}_{t=0,\dots,T}$.

This means that we assume that values at time t are revealed at that time, and those values are remembered from then on.

In some cases, it is easier to introduce a process that carries the information available on he market rather than the filtration. This motivates the following definition.

Definition 1.4 (Generated filtration)**.** A filtration generated by X (denoted $(\mathcal{F}_s^X)_{s\in\mathbb{N}}$) is the minimal filtration such that X_s is adapted.

Here the minimality means the following: assume that there is a filtration $(\mathcal{G}_t)_{t\in\mathbb{N}}$ such that X is adapted with respect to it. Then we have

$$\mathcal{F}_t^X \subseteq \mathcal{G}_t.$$

Intuitively, this filtration contains at each instant the sets required to make X adapted but nothing more.

Conditional expectat)ion

In probability, the expectation is an operator that gives us a punctual estimator on a random variable. Now, as information is evolving, the likeliness of events (measured by the probability) also evolves. Thus, what we deem the best punctual estimator now (at time 0, the start of the model) does not coincide with the best estimator at time t *given that we have more information about what actually happened (and what did not) at the times before* t. We capture this feature with the concept of conditional probability.

Definition 1.5. The **conditional expectation at time** t (denoted by $\mathbb{E}[.|\mathcal{F}_t]$, or $\mathbb{E}_t[.]$ if the filtration is clear from the context) is an operator such that for each $\mathcal{F}$ measurable random variable X with finite variance assigns an $\mathcal{F}_t$- measurable random variable given by,

$$\mathbb{E}_t[X] = \arg \inf_{Z\in L^2(\Omega,\mathcal{F}_t)} \mathbb{E}[(Z-X)^2]. \tag{1.1}$$

In the above, we call $L^2(\Omega, \mathcal{F}_t)$ to the set of all random variables X that are $\mathcal{F}_t$ measurable and such that $\mathbb{E}|X|^2 < \infty$. Hence, conditional expectation is the $\mathcal{F}_t$-measurable random variable that best approximates X in a mean quadratic sense. The operator can be extended to random variables with finite mean.

We give without proof a very useful characterisation of conditional expectations.

Proposition 1.6. *Given a random variable such that $\mathbb{E}|X| < \infty$, the conditional expectation $\mathbb{E}_t[X]$ is the only (up to measure zero) $\mathcal{F}_t$-measurable, integrable random variable such that*

$$\mathbb{E}[\hat{X}X] = \mathbb{E}[\hat{X}\mathbb{E}_t[X]] \textit{ for all } \mathcal{F}_t - \textit{measurable } \hat{X} \tag{1.2}$$

The above characterisation is handy whenever we want to verify a given candidate to be a conditional expectation.

Some properties of conditional expectation

We list now some important properties of conditional expectation. Their proof can be deduced directly from the characterisation (1.6) or from using similar properties of the expectation operator in Definition 1.1. Let X be $\mathcal{F}-$measurable.

1. (Invariance): If X is $\mathcal{F}_t$-measurable, $\mathbb{E}_t[X] = X$
2. (Homogeneity): If $\hat{X}$ is $\mathcal{F}_t$-measurable, $\mathbb{E}_t[X\hat{X}] = \hat{X}\mathbb{E}_t[X]$
3. (Orthogonality): If X is independent of $\mathcal{F}_t$, $\mathbb{E}_t[X] = \mathbb{E}[X]$
4. (Tower property): If $\ell \leq t$, $\mathbb{E}_\ell[X] = \mathbb{E}_\ell[\mathbb{E}_t[X]]$.

Moreover, as for the expectation operator we have

1. (Linearity): We have $\mathbb{E}_t[a(X + \hat{X})] = a(\mathbb{E}_t[X] + \mathbb{E}_t[\hat{X}])$ for all $a \in \mathbb{R}$
2. (Positivity): if $X \geq 0$, $\mathbb{E}_t[X] \geq 0$ (with linearity, we get monotonicity...)
3. (Independence): If $X \perp\!\!\!\perp \hat{X}$, $\mathbb{E}_t[X\hat{X}] = \mathbb{E}_t[X]\mathbb{E}_t[\hat{X}]$.
4. Monotone convergence: $0 \leq X_n \uparrow X$, then $E_t[X^n] \uparrow \mathbb{E}[X]$
5. Dominated convergence: $X^n \to X$ and $|X^n| \leq \hat{X}$ with $\hat{X}$ integrable then $\mathbb{E}_t[X^n] \to \mathbb{E}_t[X]$
6. Jensen's inequality: If $f \colon \mathbb{R} \to \mathbb{R}$ is convex, then $f(\mathbb{E}_t[X]) \leq \mathbb{E}_t[f(X)]$.

Example 1.7. *Assume that $(X_s)_{s\in\mathbb{N}}$ is i.i.d, with each entry being integrable, and assume we are working under the filtration generated by X. Assume also that $\hat{X}_s = \sum_{i=0}^{s} a_i X_i$. For all $t < s$,*

$$\mathbb{E}_t[\hat{X}_s] = \mathbb{E}_t\left[\sum_{i=0}^{s} a_i X_i\right] = \sum_{i=0}^{t} a_i \mathbb{E}_t[X_i] + \sum_{i=t+1}^{s} a_i \mathbb{E}_t[X_i] = \sum_{i=0}^{t} a_i X_i + \sum_{i=t+1}^{s} a_i \mathbb{E}[X_1].$$

Intuitively, conditional expectation $\mathbb{E}_t$ is "our best estimator" of a random variable given that "we know everything that happened up to t". For this reason, it will play an important role when considering prices of contingent claims, as will be seen in what follows.

Bank account

As announced before, the asset 0 will play a special role. A very convenient assumption is that it is a locally risk-free asset. It is customary to call such an asset *bank account* or *money market account.*

Definition 1.8. We call *money market*, *bank account* or *(locally) risk-free* asset to an asset whose price is strictly positive and predictable (see definition 1.2) with $S_0^0 = 1$. We call any asset whose price process is not predictable *risky.*

Hence, the bank account is an asset that can be overall random, but such that at each period, we have certainty on its value by the *end* of the period already at the start of the period. Observe, in particular, that if $T = 1$, the bank account is deterministic: in this context, it is also known as a risk-free asset. The assumption that $S_0^0 = 1$ is a normalisation and is useful to simplify some expressions.

In practice, some assets may act close to a *money market* account, like government bonds or overnight collateralised swaps. We will not systematically assume its existence, but it will be clear from the discussion that many developments are enhanced or simplified when one is available.

A money market account is a natural benchmark for all other assets. We will argue in the following chapters that investors usually require a larger average compensation in order to assume risk. This excess of mean return is denoted risk premium.

Definition 1.9. The *(conditional) risk premium* of asset i in the period $[t, t+1]$ is defined as

$$\mathbb{E}_t[R_{t+1}^i] - R_{t+1}^0. \tag{1.3}$$

Sometimes, the risk premium is also called **mean excess return**.

Discounted prices

It is often convenient to express prices relative to the money account, i.e.,

$$Y_t^i = \frac{S_t^i}{S_t^0}, \tag{1.4}$$

where we use the fact that this process is strictly positive. In financial terms, we say that Y^i is a *discounted price*. Note that this is equivalent to using S^0 as a *numéraire*.

Returns

Because we have assumed that an investor can buy any quantity of a given asset, it is possible and sometimes more convenient to focus on how prices change rather than what the actual prices are. This motivates the following definitions:

Definition 1.10. The *(gross) return* on asset i for the period $[t-1, t]$ for $t = 1, \ldots, T$ is defined as

$$R_t^i = \frac{S_t^i}{S_{t-1}^i}. \tag{1.5}$$

whenever the ratio is well-defined. Clearly, returns convey information in the case when total values are positive, so they are particularly useful when dealing with assets like stock prices.

Definition 1.11. The *rate of return* (or net return) on asset i is defined as

$$r_t^i = \frac{S_t^i - S_{t-1}^i}{S_{t-1}^i}. \tag{1.6}$$

For example, if an agent decides to invest an amount φ_0 in asset i at time 0, then we can get directly the amount they will have at the end of the first period by multiplying by the gross rate since

$$\varphi_0 R_1^i = \varphi_0 \frac{S_1^i}{S_0^i} = \theta_0^i S_1^i = \varphi_1$$

where θ_0^i represents the number of shares that can be bought with ϕ_0.

Likewise, investors can evaluate their net gain or loss (their actual earnings through the investment) using the rate of return, since

$$\varphi_0 r_1^i = \varphi_0 (R_1^i - 1) = \varphi_1 - \varphi_0.$$

Note that since it is possible to buy any fraction of a stock, for all modelling effects, two assets with the same gross return are indistinguishable.

We can also define a *discounted net gain* of an investment, which would be given by

$$\bar{Y}_t^i = \frac{S_t^i}{S_t^0} - \frac{S_{t-1}^i}{S_{t-1}^0}. \tag{1.7}$$

1.2 Some examples of market models

To illustrate our discussion up to now, we present here some well known market models in discrete time.

1.2.1 Independent and identically distributed returns

In this simple but important model, the vector of returns for all assets in different periods is assummed to be identically distributed and independent of each other. Moreover, the filtration is taken as the generated by the return process, so that for all $t \neq \ell$ and all $i = 0, \ldots, n$

$$\boldsymbol{R}_\ell \perp\!\!\!\perp \boldsymbol{R}_t; \text{ and } \boldsymbol{R}_\ell \sim \boldsymbol{R}_1,$$

and as a consequence if $\ell < t$,

$$\boldsymbol{R}_t \perp\!\!\!\perp \mathcal{F}_\ell. \tag{1.8}$$

Recalling 1.5, the value at time t is

$$S_t^i = S_0^i \prod_{\ell=1}^{t} R_\ell^i, \tag{1.9}$$

Now, take $\ell < t$. Using the tower and independence properties of conditional expectation, (1.8) and the i.i.d. property we get for each asset $i = 0, \ldots n$ that

$$\mathbb{E}_\ell[S_t^i] = \mathbb{E}_\ell[S_0^i \prod_{j=1}^{t} R_j^i] = S_\ell^i \mathbb{E}_\ell[\prod_{j=\ell+1}^{t} R_j^i] = S_\ell^i \prod_{j=\ell+1}^{t} \mathbb{E}[R_j^i] = S_\ell^i \mathbb{E}[R_1^i]^{t-\ell}.$$

Hence, assuming that $\mathbb{E}[R_1^i] \neq 0$, we find that for all $0 \leq \ell < t \leq T$

$$S_\ell^i = \frac{\mathbb{E}_\ell[S_t^i]}{\mathbb{E}[R_1^i]^{t-\ell}}. \tag{1.10}$$

Assuming i.i.d. returns is very convenient from an estimation point of view: observing

A log-normal model with deterministic money market account

A quite popular instance of the i.i.d. model postulates that the money market account is deterministic (i.e., $S^0_t = \psi(t)$) for some strictly positive function ψ (for example $\psi(t) = (1+r)^t$ for a fixed r), and that the returns of the other assets follow a log normal distribution, with

$$\boldsymbol{R}_t = \exp(\boldsymbol{Z}_t); \boldsymbol{Z}_t \sim \mathcal{N}(\mu, \Sigma),$$

where μ is a vector of means and Σ is the variance-covariance matrix. The advantage of the log-normal assumption is that we get from (1.9) that

$$S^i_t = S^i_0 \exp(\sum_{\ell=1}^{t} Z^i_\ell)$$

which, thanks to the fact that any linear function of a Gaussian vector is Gaussian, is also log normal, but with mean and **variance** rescaled by a factor of t: indeed,

$$\mathbb{E}[\sum_{\ell=1}^{t} Z^i_\ell] = \sum_{\ell=1}^{t} \mathbb{E}[Z^i_\ell] = t\mathbb{E}[Z^1_\ell],$$

$$\mathbb{V}[\sum_{\ell=1}^{t} Z^i_\ell] = \sum_{\ell=1}^{t} \mathbb{V}[Z^i_\ell] = t\mathbb{V}[Z^1_\ell].$$

The continuous version of this process is the well known Merton-Black-Scholes model.

The Cox, Ross and Rubinstein model (Binomial model)

Another, instance of the i.i.d. model that is usually applied in the case $n = 1$, keeps the deterministic money market account, and postulates that $R^1_t = (u-d)Z_t + d$, where $\{Z_t\}_{t=0,\ldots,T}$ are binomial random variables with probability α.

If $n = 1$, this model can be defined on a finite probability space $\Omega = (0,1)^T$, with $\mathbb{P}[\omega_t = 1] = \alpha$ so that $R^1_t = \begin{cases} u \text{ if } \omega_t = 1 \\ d \text{ if } \omega_t = 0 \end{cases}$. The value of the total stock value would correspond to the following binomial tree depicted in (1.1).

Note that case $n > 1$ might require a larger probability space to implement, depending on the joint structure between the different assets. If all assets are assumed to evolve independently, one can construct the model over the space $\Omega = \{0, \ldots, 2^n - 1\}^T$ with a well-defined probability structure.

Example 1.12. *To illustrate the meaning of Figure 1.1, suppose that $S^1_0 = 1$, $u = 1.2$ (i.e. a 20% net return per period), $d = 0.8$ (i.e. a 20% net less per period), and $\alpha = 0.6$. Then we have, for example, that*

$$S^1_2 = \begin{cases} 1.44 & \textit{with probability } 0.36 \\ 0.96 & \textit{with probability } 0.48 \\ 0.64 & \textit{with probability } 0.16. \end{cases}$$

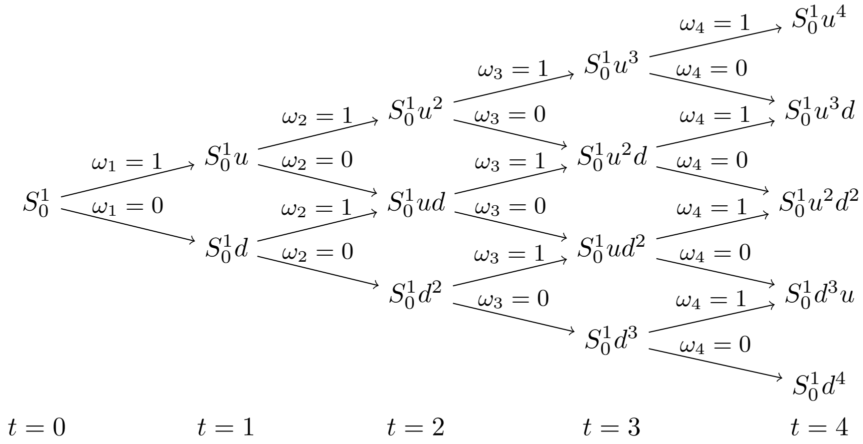

Figure 1.1: A binomial tree representing the total value process for an asset, implemented in a finite probability space, for $T = 4$. Each upper arrow has probability α and each lower arrow probability $1 - \alpha$.

1.2.2 Markovian models

The i.i.d. assumption is very strong: it means that returns do not evolve (in law) over time, and only . We can allow for changes in the returns but keeping one common trait: that prices are "memory-less" in the sense that the dynamics of the process after a time t only depend on its state at that time and not on the path the process took to arrive to it.

Let us formalise this intuition. Recall that for any two random variables $X, \hat{X}$ with finite expectation, we can define the conditional expectation $\mathbb{E}[X|\hat{X}]$ to be a random variable satisfying:

i. There exists a measurable h such that $\mathbb{E}[X|\hat{X}] = h(\hat{X})$ almost surely

ii. $\mathbb{E}[Xg(\hat{X})] = \mathbb{E}[E[X|\hat{X}]g(\hat{X})]$, for any measurable function g.

As expected, $\mathbb{E}[X|\hat{X}]$ satisfies analogous properties to the invariance and orthogonality properties of conditional expectation with respect to a filtration. Note that, in the particular case where $(X, \hat{X})$ has a joint density (say f), we have that $\mathbb{E}[X|\hat{X}] = K(\hat{X})$, where

$$K(y) = \mathbb{E}[X|\hat{X} = y] = \int xf(x, y)\mathrm{d}x.$$

We are ready to define Markov processes.

Definition 1.13. X is a Markov process if for every measurable function f and $\ell < t$,

$$\mathbb{E}[f(X_t)|\mathcal{F}_\ell] = \mathbb{E}[f(X_t)|X_\ell] = \hat{h}_{\ell,t}(X_\ell),$$

for some measurable function $\hat{h}_{\ell,t}$.

By definition, the i.i.d models before are all Markovian. Let us give a couple of examples of Markovian models for which returns are not i.i.d..

Auto regressive models of order 1 (AR(1)) with one factor

This model over the total value vector S assumes that the S_t^i are defined by

$$S_{t+1}^i = r^i S_t^i + b^i + \xi_{t+1},$$

with the ξ being i.i.d, and S_0^i known for each i. We assume the filtration is the one generated by ξ. Note that neither S^i nor R^i are i.i.d. in this case. However, we have

$$\mathbb{E}_t[S_{t+1}^i] = r^i S_t^i + b^i + \mathbb{E}_t[\xi_{t+1}] = r^i S_t^i + b^i + \mathbb{E}[\xi_1].$$

Hence, by induction, we get for all $0 \le \ell < t \le T$ that

$$\mathbb{E}_\ell[S_t^i] = (r^i)^{t-\ell} S_\ell^i + \frac{1 - r^{t-\ell}}{1-r}(b^i + \mathbb{E}[\xi_1]),$$

which clearly shows that this model is Markovian (the function $\hat{h}_{\ell,t}$ in the Definition (1.13) is simply linear).

A generalisation of this model to several factors $\xi^1, \xi^2, \dots, \xi^N$ is straightforward.

General 1-factor Markovian model

For this model over the total value vector S, we choose measurable functions ϕ^i, ψ^i and assume that the S_t^i are defined by $S_{t+1}^i = \psi^i(S_t^i) + \phi^i(S_t^i)\xi_{t+1}$, with the ξ being i.i.d, and S_0^i known for each i. As before, we assume the filtration is the one generated by ξ. In this case,

$$\mathbb{E}_t[S_{t+1}^i] = \mathbb{E}_t[\psi^i(S_t^i)] + \mathbb{E}_t[\phi^i(S_t^i)]\mathbb{E}[\xi_1];$$

however, using property i of conditional expectation, we get that

$$\mathbb{E}_t[S_{t+1}^i] = \hat{\psi}_t^i(S_t^i) + \hat{\phi}_t^i(S_t^i)\mathbb{E}[\xi_1],$$

for some measurable functions ψ_t^i, ϕ_t^i. An induction argument as in the auto-regressive case then shows that this model is Markovian. Once again, this model can be easily generalised to several factors.

1.2.3 Non-Markov models

Let us see some examples of possible **non**-Markovian models (a.k.a. path-dependent processes). Once more, we assume the filtration is the one generated by the process ξ, which is assumed to be i.i.d.

- Auto-regressive of second order (AR(2)): S^i defined by

  $$S_{t+1}^i = a_1^i S_t^i + a_2^i S_{t-1}^i + b^i + \xi_t.$$

  In this example, the state of variable X_{t+1} depends not only on the state at time t but also on the one at time $t-1$. Note in particular that

  $$\mathbb{E}[X_{t+1}|\mathcal{F}_t] = a_1^i S_t^i + a_2^i S_{t-1}^i + b^i \neq_1^i S_t^i + a_2^i \mathbb{E}[S_{t-1}^i|S_t^i] + b^i = \mathbb{E}[X_{t+1}|X_t]$$

  since $\mathbb{E}[S_{t-1}^i|S_t^i] \neq S_{t-1}^i$. Hence it is not Markovian. This model allows to include a 'trend' element in the series.

- Mild support and resistance process: X defined by

$$X_{t+1} = c_0^i(S_t^i - \min_{0\le s<t}(S_s^i))^+ - c_1^i(\max_{0\le s<t}(S_s^i) - S_t^i)^+ + a^i S_t^i + b^i + \xi_t;$$

  where $(x)^+ = \frac{|x|+x}{2}$ is the positive part function. This model adds a push 'up' (in the form of the coefficient c_0) when the price is reaching the historical minimal, and a push 'down' when (in the form of the coefficient c_1) when reaching the historical maximum. The presence of the running minimum and maximum makes the process non-Markovian.

In many cases, as the ones above, it is possible to render the model Markovian by extending the set of state variables, that is, we provide a model for the vector of total value prices **and** for some extra variables (for example, the process extended by the delayed prices).

Exercise. Choose a set of state variables to enlarge the set of state variables to render the above examples Markovian.

1.3 Arbitrage and completeness

We are ready to introduce the set of actions that a market participant can undertake on our market model and characterise some good properties of the market with respect to these actions. We start with a definition.

Definition 1.14 (Strategy)**.** A *strategy* is a set of actions that an investor decides to perform on the market (in order to obtain their goal). It is a predictable process $\boldsymbol{\theta}$ with values in $\mathbb{R}^{n+1}$ so that θ_t^j denotes the number of shares of asset j to be held during the period $[t-1, t]$.

The predictable property is imposed to reflect the fact that investors decide on their strategy with the information available to them at the start of each period.

Remark 1.1. In the one-period model, the only choices available to investors are at initial time. Hence, in the one-period model, choosing a strategy means simply choosing a portfolio composition, i.e., a vector $\boldsymbol{\theta} \in \mathbb{R}^{n+1}$.

From the definition, we see that the amount invested on the asset j at time $t-1$ is $S_{t-1}^j \theta_t^j$. At the end of the period, just before recomposition, the investor would have as a result $S_t^j \theta_t^j$.

We denote by $S^{\boldsymbol{\theta}}$ the value of a portfolio that follows the strategy $\boldsymbol{\theta}$. By definition,

$$S_{t-1}^{\boldsymbol{\theta}} = \sum_{j=0}^{n} S_{t-1}^j \theta_t^j = \boldsymbol{\theta}_t \boldsymbol{S}_{t-1},$$

and by time t, before any changes in the composition (that we denote $t-$), this would become

$$S_{t-}^{\boldsymbol{\theta}} = \boldsymbol{\theta}_t \boldsymbol{S}_t.$$

If investors only put forward some money at initial time, and later only redistribute the value of their portfolios, we say that the strategy is *self-financing*. In our model, this reads as follows

Definition 1.15 (Self-financing strategies)**.** A strategy is self-financing if for all $t = 1, \dots, T-1$,

$$S_{t-}^{\boldsymbol{\theta}} = \boldsymbol{\theta}_t \boldsymbol{S}_t = \boldsymbol{\theta}_{t+1} \boldsymbol{S}_t = S_t^{\boldsymbol{\theta}}.$$

Exercise. Show that a strategy is self-financing if and only if for all t,

$$S_t^{\boldsymbol{\theta}} = \boldsymbol{\theta}_1 \boldsymbol{S}_0 + \sum_{\ell=1}^{t} \boldsymbol{\theta}_\ell (\boldsymbol{S}_\ell - \boldsymbol{S}_{\ell-1}).$$

Remark 1.2. In many occasions, it might be more convenient to express strategies in alternative but equivalent units. For example, instead of expressing it in terms of the number of units of an asset to hold ($\boldsymbol{\theta}$), we can use the *amount* to hold on a given asset ($\boldsymbol{\varphi}$). We have

$$\varphi_{t-}^i = \theta_{t-1}^i S_{t-1}^i; \qquad \varphi_t^i = \theta_t^i S_{t-1}^i.$$

In terms of this convention, self-financing strategies satisfy the relation

$$\boldsymbol{\varphi}_{t-} \cdot \mathbf{1} = \sum_i \varphi_{t-1}^i = \sum_i \varphi_t^i = \boldsymbol{\varphi}_t \cdot \mathbf{1}.$$

Finding the strategy in terms of units from the strategy in terms of amounts is straight-forward.

Yet another useful way to express strategies consist in describing the *proportion* of invested wealth to hold on each asset ($\boldsymbol{\pi}$). We have,

$$\pi_t^i = \frac{\varphi_t^i}{\sum_j \varphi_t^j}.$$

This convention can be used only if explicit information on wealth is available in addition, so that

$$\varphi_t^i = W_t \pi_t^i.$$

If such information is not given, it is assumed that the strategy is self-financing: in this case, we can recover the strategy in terms of, say, amounts by keeping track of wealth through time.

We say that there is an arbitrage opportunity if we can construct a costless self-financing strategy that produces some returns but never losses by the end of the modelling time. Let us make this definition precise:

Definition 1.16. A market allows for an *arbitrage opportunity*, if there exists a self-financing strategy $\boldsymbol{\theta}$ such that

1. $S_0^{\boldsymbol{\theta}} \leq 0$
2. $\mathbb{P}[S_T^{\boldsymbol{\theta}} \geq 0] = 1$
3. $\mathbb{P}[S_T^{\boldsymbol{\theta}} > 0] > 0$.

A "good" market does **not** allow arbitrage opportunities to exist: indeed, an arbitrage opportunity would be rapidly detected and eliminated as a consequence of investors exploiting it. Thus, arbitrage opportunities are inefficiencies to be exploited, while a no-arbitrage condition intuitively implies that all investors have good access to the market and to information.

We call a market model where arbitrages are not possible *arbitrage-free*.

Another characteristic of a "good" market model is for it to be rich enough so that one can reproduce, at time T any desired wealth profile.

Definition 1.17. We say that a wealth profile W (at time T) can be *replicated* if there exists a self-financing strategy $\boldsymbol{\theta}$ such that

$$S_T^{\boldsymbol{\theta}} = W \text{ almost surely.}$$

The replication price is in this case simply the initial investment required to acquire such a self-financing strategy, i.e., $S_0^{\boldsymbol{\theta}}$. If such a strategy exists $\boldsymbol{\theta}$, we also say that $\boldsymbol{\theta}$ replicates W.

Definition 1.18. We say that a market is *complete* if every $\mathcal{F}_T$-measurable wealth W with $0 \leq W$ and $\mathbb{P}(W < \infty) = 1$ can be replicated.

Since strategies are so limited in one-period models, market completeness is a very strong assumption for this type of model: in essence, we need to have at least as many assets as different possible outcomes on the probability space. As more opportunities to rebalance a portfolio are given (i.e., when we move to multiperiod models), it becomes easier to replicate a given profile, or equivalently, we require fewer assets to guarantee completeness. In some continuous models (where constant rebalancing of a portfolio is allowed) complete models are possible even when only one risky and a risk-free asset are available.

A complete market is a well-developed market, as there is an incentive for financial institutions to introduce new assets to the market to account for uncovered profiles. However, in practice, completeness is limited by aspects like regulation, moral hazard, adverse selection and sophistication aversion.

Finding arbitrage-free prices is an important task in financial mathematics. In several cases, the problem can be solved by introducing some structural characteristics of an arbitrage-free market: either the existence of a special process, the *stochastic discount factor* or the introduction of a new measure *the risk neutral measure.* We study these in the following sections.

1.4 Stochastic discount factor (SDF)

In a "good" market model (in the sense of having the rules for buying assets and the no-arbitrage and completeness properties we have presented), there are some structural relationships between the asset values at the beginning and at the end of each period: for example, given that we know the random variable of the value of an asset at the end of the period, we know (by no arbitrage) that some initial prices would be ruled out. We will see in the following that the relationship is even stronger. Let us introduce some new concepts.

Definition 1.19. We say that an adapted process M in $\mathbb{R}$ is a stochastic discount factor (SDF) process, if we have that $M_0 = 1$, and for each $i = 0, \ldots, n$

$$\mathbb{E}[|M_T S_T^i|] < \infty; \tag{1.11}$$

$$M_t S_t^i = \mathbb{E}_t[M_{t+1} S_{t+1}^i]. \tag{1.12}$$

The intuition behind this definition is as follows: if we assume that the market is assumed to be risk neutral (i.e. pricing by taking only expectation), the current prices are explained if we suppose that the market discounts pay-offs depending on both the time of occurrence and the actual event that arises. This is encapsulated in the random process M. Indeed, it follows from the fact that $M_0 = 1$ and (1.12) that for any t,

$$S_0^i = \mathbb{E}[M_t S_t^i]. \tag{1.13}$$

The SDF can also be seen as connecting returns. Indeed, by dividing by S_t in (1.12) we get

$$M_t = \mathbb{E}_t[M_{t+1}R^i_{t+1}] \text{ for all } i = 0, \dots, n; \tag{1.14}$$

and by the properties of conditional expectation,

$$\mathbb{E}_t[M_{t+1}(R^i_{t+1} - R^j_{t+1})] = 0 \text{ for all } i, j = 0, \dots, n;$$

which implies that the SDF is orthogonal to relative excess returns.

Exercise. Using induction prove that for all $0 \le t \le T$,

$$M_\ell S^i_\ell = \mathbb{E}_\ell[M_t S^i_t] \text{ for all } \ell < t. \tag{1.15}$$

Example 1.20. *Let us look at an instance of the binomial model on returns presented in section* (1.2.1)*. We assume there is one deterministic asset and two risky assets in a multiperiod market model with* T *periods. More specifically, we assume that* $\Omega = \{0, 1, 2\}^T$ *with uniform probability and*

$$(R^1, R^2)_t = \begin{cases} (u_1, u_2) & \textit{if } \omega_t = 0 \\ (u_1, d_2) & \textit{if } \omega_t = 1\,; \\ (d_1, u_2) & \textit{if } \omega_t = 2 \end{cases} \qquad \textit{and} \qquad R^0_t = R^0 \textit{ is constant.}$$

Where we assume that $0 < d_1 < R^0 < u_1$*, and* $0 < d_2 < R^0 < u_2$*. Now, let us set*

$$Z_t = \begin{cases} \frac{3}{R^0}(a + b - 1) & \textit{if } \omega_t = 0 \\ \frac{3}{R^0}(1 - b) & \textit{if } \omega_t = 1 \\ \frac{3}{R^0}(1 - a) & \textit{if } \omega_t = 2 \end{cases}$$

with

$$a = \frac{R^0 - d_1}{u_1 - d_1}; b = \frac{R^0 - d_2}{u_2 - d_2}.$$

Note that $a, b \in (0, 1)$*, due to our constraints on the constants of the problem. We now show that*

$$M_t = \prod_{\ell=1}^{t} Z_\ell$$

is a strictly positive stochastic discount factor for this market model.

- *We first verify the initial condition. Indeed,* $M_0 = 1$.
- *We now verify that* $\mathbb{E}[|M_T S^i_T|] < \infty$*. However, this is trivial as long as* $T < \infty$ *since we are in a finite probability space.*
- *Finally, we check the property of conditional expectations. We get*

$$\begin{aligned} \mathbb{E}_t[M_{t+1}R^0_{t+1}] &= R^0\mathbb{E}_t[M_t Z_{t+1}] = R^0 M_t \mathbb{E}_t[Z_{t+1}] = M_t; \\ \mathbb{E}_t[M_{t+1}R^1_{t+1}] &= M_t\mathbb{E}_t[Z_{t+1}R^1_{t+1}] = \frac{M_t}{3}\left(\frac{3u_1}{R^0}(a + b - 1) + \frac{3u_1}{R^0}(1 - b) + \frac{3d_1}{R^0}(1 - a)\right) \\ &= M_t\left(\frac{u_1 a}{R^0} + \frac{d_1(1 - a)}{R^0}\right) = M_t; \\ \mathbb{E}_t[M_{t+1}R^2_{t+1}] &= M_t\mathbb{E}_t[Z_{t+1}R^2_{t+1}] = \frac{M_t}{3}\left(\frac{3u_2}{R^0}(a + b - 1) + \frac{3d_2}{R^0}(1 - b) + \frac{3u_2}{R^0}(1 - a)\right) \\ &= M_t\left(\frac{u_2 b}{R^0} + \frac{d_2(1 - b)}{R^0}\right) = M_t. \end{aligned}$$

Hence, M is indeed an SDF. Finally, note that if we add the condition

$$R^0 > \frac{u_1 u_2 - d_1 d_2}{(u_1 + u_2) - (d_1 + d_2)}$$

the SDF M defined above is strictly positive. □

Martingales

The structure of the processes $(M_t S_t^i)_{0 \le t \le T}$ is very special: the best estimation at time t for the future values of the process is precisely the value of the process at time t. Such processes are called martingales.

Definition 1.21. An **integrable** and adapted process X is a martingale if

$$\mathbb{E}_\ell[X_t] = X_\ell \text{ for all } \ell \le t,$$

i.e., if the best $\mathcal{F}_\ell$ -measurable approximation of the process at every time bigger that ℓ is its value at ℓ itself.

Example 1.22. *The definition of an SDF guarantees that MS^i (the total market value of each asset weighted by the stochastic discount factor) are martingales.*

The above example can be made stronger.

Proposition 1.23. *An adapted process M is a stochastic discount factor if and only if the process MS^θ is a martingale for any bounded self-financing strategy θ.*

Exercise. Prove Proposition 1.23.

Stochastic discount factors are not guaranteed to exist or to be unique. However, some remarkable results connect the existence and uniqueness of an SDF with the no-arbitrage and completeness market properties. These theorems are the so-called *fundamental theorems of asset pricing.* In our simple discrete-time finite world, they read as follows:

Theorem 1.24 (First fundamental theorem asset pricing)**.** *A market has no arbitrage opportunities if and only if there exists a (strictly) positive SDF.*

Theorem 1.25 (Second fundamental theorem asset pricing)**.** *An arbitrage-free market model is complete if and only if there is a unique SDF.*

We present the proof of these theorems in the particular case of finite probability space in the one-period model in section 1.7.3 below.

1.5 Risk neutral probability measure

As we explained before, assuming the market is risk-neutral, the introduction of an SDF accounted intuitively for our observed prices by introducing a discounting term. An alternative is to assume that the market has a different perception of the probabilities associated to each possible outcome. The risk neutral property will be expressed in the fact that under this alternative probability measure (that in the following we denote $\mathbb{Q}$), the conditional risk premium of any asset is zero.

Let us start by reminding ourselves the concept of equivalent measures.

Definition 1.26. We say that the probability measure Q_1 is absolutely continuous with respect to the probability measure Q_2 (denoted $Q_1 \ll Q_2$), if Q_1 assigns zero probability to any set with zero probability under Q_2 (i.e., $Q_2(A) = 0 \Rightarrow Q_1(A) = 0$). If both $Q_1 \ll Q_2$ and $Q_2 \ll Q_1$, we say the two measures are equivalent (denoted $Q_1 \sim Q_2$).

Thanks to the Radon-Nikodym theorem (see for example Durrett (2010)), we know that equivalent measures can be shown to be connected via a positive random variable with unit expectation that re-weights the observations and is called a *density*. In the following we show how to construct the risk neutral measure $\mathbb{Q}$ by defining appropriately its density.

Let us assume that the market is arbitrage-free. We know that in this case there is a strictly positive SDF (M). We can then define the random variable

$$Z := M_T S_T^0. \tag{1.16}$$

The expectation operator of the *risk neutral probability* $\mathbb{Q}$ for any random variable X is then given by

$$\mathbb{E}^{\mathbb{Q}}[X] := \mathbb{E}[ZX] = \mathbb{E}\left[M_T S_T^0 X\right]. \tag{1.17}$$

Equivalently, we can define $\mathbb{Q}$ for each event A by,

$$\mathbb{Q}[A] := \mathbb{E}^{\mathbb{Q}}[\mathbb{1}_A] = \mathbb{E}\left[Z\mathbb{1}_A\right] = \mathbb{E}\left[M_T S_T^0 \mathbb{1}_A\right]. \tag{1.18}$$

Let us verify that in fact Z is a density.

Proposition 1.27. *If M is a strictly positive SDF, the operator $\mathbb{Q}$ defined above is a true probability and $Z := \frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}}$.*

Proof. The claim follows from the fact that Z is strictly positive by the assumptions on M and S^0. Indeed, we can verify that $\mathbb{Q}$ so defined is a true probability. Indeed:

- $\mathbb{Q}[\Omega] = \mathbb{E}[Z\mathbb{1}_\Omega] = \mathbb{E}[Z] = \mathbb{E}[M_T S_T^0] = M_0 S_0^0 = 1$.
- Since Z is strictly positive, trivially $\mathbb{Q}[A] := \mathbb{E}^{\mathbb{Q}}[\mathbb{1}_A] \geq 0$
- Set $A_1, A_2, \ldots, \in \mathcal{F}$ pairwise disjoint, we get

$$\mathbb{Q}[\bigcup A_i] = \mathbb{E}[Z\mathbb{1}_{\bigcup A_i}] = \mathbb{E}[Z \sum \mathbb{1}_{\bigcup A_i}] = \sum \mathbb{E}[Z\mathbb{1}_{A_i}] = \sum \mathbb{Q}[A_i]$$

Finally, we also deduce that $\mathbb{Q}[A] = \mathbb{E}[Z\mathbb{1}_A] = 0 \Leftrightarrow E[\mathbb{1}_A] = \mathbb{P}[A] = 0$. Hence both probabilities are equivalent and Z is the Radon Nikodym density almost surely. $\square$

Lemma 1.28. *Let $0 \leq t \leq T$, X an arbitrary random variable and Z and $\mathbb{Q}$ defined as above. Then,*

$$\mathbb{E}_t^{\mathbb{Q}}[X] = \frac{\mathbb{E}_t[ZX]}{\mathbb{E}_t[Z]}$$

Proof. We use the characterisation presented in section 1.1.1:

i. $\frac{\mathbb{E}_t[ZX]}{\mathbb{E}_t[Z]}$ is clearly $\mathcal{F}_t$ measurable at is the composition of $\mathcal{F}_t$ measurable functions.

ii. Let $\hat{X}$ be an $\mathcal{F}_t$ measurable random variable. Then, using the properties of conditional expectation in 1.1.1 we have

$$\mathbb{E}^Q\left[\hat{X}\frac{\mathbb{E}_t[ZX]}{\mathbb{E}_t[Z]}\right] = \mathbb{E}\left[Z\{\underbrace{\hat{X}\frac{\mathbb{E}_t[ZX]}{\mathbb{E}_t[Z]}}_{\mathcal{F}_t-measurable}\}\right] = \mathbb{E}\left[\mathbb{E}_t[Z]\{\hat{X}\frac{\mathbb{E}_t[ZX]}{\mathbb{E}_t[Z]}\}\right]$$
$$= \mathbb{E}[\hat{X}\mathbb{E}_t[ZX]] = \mathbb{E}[\hat{X}ZX] = \mathbb{E}^{\mathbb{Q}}[\hat{X}X].$$

By our characterisation theorem, the claim follows. □

Definition 1.29 (Risk-neutral measure)**.** We say that a measure $\mathbb{Q}$ defined on $(\Omega, \mathcal{F})$ is a risk-neutral measure if it is equivalent to $\mathbb{P}$ and its Radon-Nikodym density $\frac{d\mathbb{Q}}{d\mathbb{P}}$ can be written as in (1.16) for some strictly positive stochastic discount factor M.

Remark 1.3. From the definition, we can deduce that **there is a one-to-one correspondence between risk-neutral probabilities and strictly positive stochastic discount factors**. As a consequence, arbitrage-free markets might have more than one risk neutral probability with uniqueness if and only if the market is complete.

The property of a process of being a martingale depends both on the filtration and the probability used to evaluate conditional expectations. Sometimes it will be convenient to evaluate the property with respect to a measure different from $\mathbb{P}$. The following theorem makes such a claim for a process when considering conditional expectations with respect to $\mathbb{Q}$. It is a very well known theorem in financial mathematics.

Theorem 1.30. *A measure $\mathbb{Q}$ is a risk-neutral probability if and only if $\mathbb{Q}$ is equivalent to $\mathbb{P}$ and discounted prices of market instruments (see* (1.4)*) are $\mathbb{Q}$-martingales.*

Proof. Assume first that $\mathbb{Q}$ defined by the density (1.16). We have already verified that this implies that $\mathbb{Q}$ is equivalent to $\mathbb{P}$. It remains to show that discounted prices are martingales. We check first that

$$\mathbb{E}^{\mathbb{Q}}|Y_T^i| = \mathbb{E}|M_T S_T^0 Y_T^i| = \mathbb{E}|M_T S_T^i| < \infty \tag{1.19}$$

by the definition of stochastic discount factor.

Now, recall from (1.15) that the process $(M_t S_t^0)_{t=0,\ldots,T}$ is a martingale. Set $0 \le \ell < t \le T$. Then, we have from Lemma 1.28 and the tower property of conditional expectation that

$$\mathbb{E}_\ell^{\mathbb{Q}}[Y_t^i] = \frac{\mathbb{E}_\ell[ZY_t^i]}{\mathbb{E}_\ell[Z]} = \frac{\mathbb{E}_\ell[\mathbb{E}_t[M_T S_T^0]Y_t^i]}{M_\ell S_\ell^0} = \frac{\mathbb{E}_\ell[M_t S_t^0 Y_t^i]}{M_\ell S_\ell^0} = \frac{\mathbb{E}_\ell[M_t S_t^i]}{M_\ell S_\ell^0} = Y_\ell^i. \tag{1.20}$$

Consider now the opposite direction, and assume that $\mathbb{Q}$ is an equivalent measure under which discounted prices are martingales. Let us define for all $0 \le t \le T$

$$M_t := \frac{1}{S_t^0}\mathbb{E}_t\left[\frac{d\mathbb{Q}}{d\mathbb{P}}\right].$$

Note that trivially, by setting $t = T$, $\frac{d\mathbb{Q}}{d\mathbb{P}} = M_T S_T^0$. We just need to check that such a process would be a strictly positive SDF. But this follows directly from (1.19) and (1.20). □

Due to Theorem 1.30, risk-neutral measures are also frequently called martingale measures.

Let us close this section by verifying our claim that under the risk-neutral measure there are no risk premia. We prove first the following small lemma:

Lemma 1.31. *Assume S^0 is a money market account. For all $t = 1, \ldots, T$ and any $\mathcal{F}_t$-measurable random variable X*

$$\mathbb{E}^{\mathbb{Q}}_{t-1}[X] = \frac{1}{\mathbb{E}_{t-1}[M_t]}\mathbb{E}_{t-1}[M_t X]$$

Proof. Using Lemma 1.28 we get,

$$\mathbb{E}^{\mathbb{Q}}_{t-1}[X] = \frac{\mathbb{E}_{t-1}[M_T S^0_T X]}{\mathbb{E}_{t-1}[M_T S^0_T]} = \frac{\mathbb{E}_{t-1}[M_T S^0_T X]}{\mathbb{E}_{t-1}[M_T S^0_T]} = \frac{\mathbb{E}_{t-1}[\mathbb{E}_t[M_T S^0_T] X]}{M_{t-1} S^0_{t-1}} = \frac{\mathbb{E}_{t-1}[M_t S^0_t X]}{M_{t-1} S^0_{t-1}}$$

where we used the martingale property of MS^0. Moreover, using the predictable property of the money market account, we get

$$\mathbb{E}^{\mathbb{Q}}_{t-1}[X] = \frac{\mathbb{E}_{t-1}[M_t X] S^0_t}{M_{t-1} S^0_{t-1}} = \frac{\mathbb{E}_{t-1}[M_t X] R^0_t}{M_{t-1}} = \frac{\mathbb{E}_{t-1}[M_t X]}{\mathbb{E}_{t-1}[M_t]},$$

where we used (1.14). □

We can now deduce the following result.

Proposition 1.32. *The conditional risk premium under the risk neutral probability $\mathbb{Q}$ of any market asset on any period is zero.*

Proof. Let us calculate the conditional risk premium for an arbitrary asset i on the period $[t-1, t]$. We get from Lemma 1.31 that

$$\mathbb{E}^{\mathbb{Q}}_{t-1}[R^i_t] = \frac{\mathbb{E}_{t-1}[M_t R^i_t]}{\mathbb{E}_{t-1}[M_t]} = \frac{M_{t-1}}{\mathbb{E}_{t-1}[M_t]} = R^0_t.$$

Replacing in the expression for the conditional risk premium (1.9) deduces the claim. □

1.6 Arbitrage-free pricing of financial products

Imagine now that you want to introduce a new asset to the market. To simplify the description, let us suppose that the asset is a *European claim*, that is an asset that produces a unique payoff at a time $0 < \tau \leq T$ (let us call the payoff H for the rest of the section). We assume that H is $\mathcal{F}_\tau$-measurable.

There are two important questions related to introducing a new assets: *what price should you charge for this asset?* and *is there any strategy that can be followed to eliminate any risk in offering the asset?*

In an arbitrage-free market, we can provide a satisfactory answer to these questions. Let us start with pricing: an important criterion to choose the price of a new asset is that it would not introduce an arbitrage in an arbitrage-free market. This motivates the following definition:

Definition 1.33. Consider an arbitrage-free market model. Assume that a contingent claim (i.e., an asset whose price depends on the value of other assets) with pay-off H at time τ wants to be introduced to the market. We say that p^{new} is an *arbitrage-free price* if the market extended with the new asset at the new price is still arbitrage-free.

There is a simple characterisation of arbitrage-free prices: they are precisely the prices that are obtained from using an SDF (or a risk-neutral measure). This is summarised in the following Proposition (you will prove it in Exercise 3.6.1)

Proposition 1.34. *A price p^{new} for a new European option with unique payoff H at time τ (i.e. $S_\tau^{new} = H$) is arbitrage-free if and only if $p^{new} = \mathbb{E}[M_\tau H]$ for some strictly positive SDF (M).*

Remark 1.4. Note that a similar proposition can be deduced in terms of the risk-neutral measures, but the condition would be $p^{new} = \mathbb{E}_T^{\mathbb{Q}}[H/S_\tau^0]$.

A direct consequence of Proposition 1.34 is that there are potentially as many arbitrage-free prices for an asset as distinct SDFs. Hence, there is only one possible arbitrage-free price in a complete market, but in general there might be more. Note that this does not mean that the asset can have more than one price without violating the no-arbitrage condition: it simply says that there is more than one *candidate* for a price, before introducing it, that would not produce an arbitrage. Naturally, once the asset is introduced, all other prices are ruled out; moreover, all SDFs that are incompatible with such price need to be dropped!

There are some interesting features of the *set* of arbitrage-free prices of European claims with payoff H at time τ (call it $A(H,\tau)$ for the rest of this section): for example, it is a convex set, bounded above and below (see Exercises 1.9 and 1.10). This implies that this set is, in fact, an interval.

Although keeping the market arbitrage-free is desirable, it might be an insufficient argument to convince a seller to choose a price in $A(H,\tau)$. Strikingly, though, the connection between existence of SDFs and no-arbitrage attains a dimension in pricing. We start with another definition.

Definition 1.35. Consider an asset with payoff H (paid at time τ). A (self-funding) strategy θ is called a super-hedging of H if

$$S_\tau^\theta \geq H.$$

If $S_\tau^\theta = H$ we say that θ is a *perfect hedging* or a *replicating strategy*

Super-hedging strategies are very useful: a seller of a European claim can in turn invest in the market using a super-hedge to completely cover the payoff (and in some cases it may even gain additional wealth. Now, here is the link between no-arbitrage prices and super-hedging strategies:

Proposition 1.36. *There exists a super-hedging strategy θ of a European claim with payoff H if and only if S_0^θ is an upper bound for $A(H,\tau)$.*

An apparently slightly stronger statement is stated in Exercise 1.11: it looks for super-hedges that are not replicating strategies with an additional constraint. To show Proposition 1.36 is to deduce a completely analogous claim: that we can only find a self-funding strategy that will be smaller than H at time τ if its initial value is smaller than all arbitrage-free prices.

A Corollary of the above is

Corollary 1.37. *In a complete market it is possible to find a replicating strategy of a European claim with payoff H at time τ if and only if the initial value of the strategy is the unique arbitrage-free price of the European claim.*

Remark 1.5. The case of claims with payoffs ranging throughout different times can be solved by considering independently each payoff and adding the prices and strategies for each one of them.

1.6.1 Models under $\mathbb{Q}$

We have seen that there is an equivalence between having a strictly positive SDF and finding a risk-neutral measure for pricing purposes.

However, from a *pricing* modelling point of view, there is a distinctive advantage of considering the risk measure $\mathbb{Q}$: we can define *market models directly under* $\mathbb{Q}$ instead of under $\mathbb{P}$; to do this, we estimate probability distributions under $\mathbb{Q}$ by using the market prices available in the market.

Frequently, the available prices are not enough to completely characterise these distributions. So, in practice, one can assume an overall model under $\mathbb{Q}$ (for example, an i.i.d. Gaussian model) and the parameters of the model are estimated from market values. This is a technique frequently used by quants in practice.

The advantage of such approach is that commonly for pricing purposes quants do not need to know the statistical measure $\mathbb{P}$. Note, though, that from a *risk measurement* point of view, this approach is flawed: the probabilities under $\mathbb{Q}$ are a mathematical construction to make sense of the structure in market prices, but do not reflect the possible losses that might appear.

1.7 A deeper look at one-period models in a finite probability space

In order to better illustrate the theory we have introduced up to now, let us consider in this section the case of market models in one period ($T = 1$) where the possible set of outcomes Ω is finite, i.e., where $\Omega = (\omega_1, \omega_2, ..., \omega_k)$ for some $k < \infty$. We take the complete sigma algebra (i.e., we can find the probability of every subset of Ω). Recall that in the finite case, the probability measure $\mathbb{P}$ is completely determined by the probability of each singleton $\mathbb{P}[\{\omega_1\}], \ldots, \mathbb{P}[\{\omega_k\}]$. We assume they are all different from zero (otherwise we reduce the number of states).

There are two main advantages of studying the finite probability setting: the first one is that due to its simplicity, we will be able to give a further economic interpretation to the SDF. Secondly, we will be able to rewrite many properties of the market in terms of matrix operations. Hence, we are able to use results in linear algebra to verify easily market properties like completeness or no-arbitrage, and show simply the F.T.A.Ps.

1.7.1 Understanding the SDF in terms of Arrow-Debreu securities

Let us assume that the market is arbitrage-free and complete. In this case, we can get a better intuition of what an SDF means. We introduce the concept of an Arrow-Debreu security, as a security that identifies outcomes on the probability space.

Definition 1.38. A security that pays one unit of the consumption good in a particular state ω_ℓ, and pays zero in all other states is called an *Arrow security* or an *Arrow-Debreu security.* In mathematical notation, the ℓ-th AD security payoff is written $\mathbb{1}_{\{\omega_\ell\}}$.

Let p_ℓ^{AD} be the price of the AD security corresponding to state ℓ. Informally, we can understand their price as the value in present time of guaranteeing a unit in a given scenario. Because of market completeness, we can replicate each AD security with market instruments and there is only one price for each market instrument. Hence,

$$p_\ell^{AD} = \mathbb{E}[M_1 \mathbb{1}_{\{\omega_\ell\}}] = \sum_{j=1}^{k} M_1(\omega_j)\mathbb{1}_{\{\omega_\ell\}}\mathbb{P}[\{\omega_j\}] = M_1(\omega_\ell)\mathbb{P}[\{\omega_\ell\}] \tag{1.21}$$

or equivalently,

$$M_1(\omega_\ell) = \frac{p_\ell^{AD}}{\mathbb{P}[\{\omega_\ell\}]} = \frac{\text{Price of scenario } \ell}{\text{Probability of the state } \omega_\ell}. \tag{1.22}$$

which can be understood as a re-weighting of the probability by the price of a unit on each state. For this reason, M is also called "*state-price density*" or "*pricing kernel.*"

Remark 1.6. Note that although we assumed market completeness to explain the notion of Arrow-Debreu securities, the AD prices can be directly found from (1.21), and thus they only require the existence of an SDF. Hence, there is a one-to-one correspondence between AD prices, SDFs and risk neutral probabilities in the finite probability space case.

Finally, note that AD prices can also be used to price an asset, since

$$S_0^i = \mathbb{E}[MS_1^i] = \sum_{j=1}^{k} S_1(\omega_j)M_1(\omega_j)\mathbb{P}[\{\omega_j\}] = \sum_{j=1}^{k} S_1(\omega_j)p_j^{AD} \tag{1.23}$$

1.7.2 Matrix interpretation

In the finite state model, we can also rewrite the market properties in terms of vectors and matrices. For example, we introduce the matrix $\mathcal{M}_{S_1} \in \mathbb{R}^{(n+1)\times k}$ to be given by

$$\mathcal{M}_{S_1} := \begin{bmatrix} S_1^0(\omega_1) & \dots & S_1^0(\omega_k) \\ \vdots & \ddots & \vdots \\ S_1^n(\omega_1) & \dots & S_1^n(\omega_k) \end{bmatrix},$$

which we write succinctly by $\mathcal{M}_{S_1} = (S_1^i(\omega_j))_{0\le i\le n, 1\le j\le k}$. Let us also set[2] $\boldsymbol{\nu} = (\mathbb{P}[\{\omega_1\}], \dots, \mathbb{P}[\{\omega_k\}])^\top$. Note that each one of the random variables S_1^i is included as a row vector in $\mathcal{M}_{S_1}$. Hence, the vector of expectations for each asset can be found by operating matrix multiplication with the vector $\boldsymbol{\nu}$ on the right, i.e.

$$\mathcal{M}_{S_1}\boldsymbol{\nu} = [\sum_{j=1}^{k} S_1^1(\omega_j)\mathbb{P}[\{\omega_j\}], \dots, \sum_{j=1}^{k} S_1^n(\omega_j)\mathbb{P}[\{\omega_j\}]]^\top,$$

while portfolios can be found by operating by the transpose of a vector $\boldsymbol{\theta} \in \mathbb{R}^{n+1}$ on the left, i.e.

$$\theta^\top \mathcal{M}_{S_1} = [\sum_{i=0}^{n} \theta_i S_1^i(\omega_1), \dots, \sum_{i=0}^{n} \theta_i S_1^i(\omega_k)].$$

[2]We use $\boldsymbol{\nu}$ as $\boldsymbol{P}$ has already been taken

This last property immediately implies the following characterisation

Proposition 1.39. *A one-period market model in a finite probability space is complete if and only if the range of* $\mathcal{M}_{S_1}^\top$ *is* $\mathbb{R}^k$.

Moreover, a vector of AD prices denoted $\boldsymbol{p}^{AD}$ would satisfy, thanks to (1.23)

$$\boldsymbol{S}_0 = \mathcal{M}_{S_1}\boldsymbol{p}^{AD} \tag{1.24}$$

Hence we conclude from the observation in Remark 1.6 and (1.21) and (accepting for now) the first theorem of asset pricing 1.24, that

Proposition 1.40. *A one-period market model in a finite probability space is arbitrage-free if and only if there is a* strictly positive *solution to* (1.24).

All in all, this means that we can check easily the main properties of the market and obtain its main descriptors using properties of the matrix $\mathcal{M}_{S_1}$.

Example 1.41. *Take a market with two assets and three equally likely scenarios. Assume that* $\boldsymbol{S}_0 = [1, 3]^\top$ *and*

$$\mathcal{M}_{S_1} := \begin{bmatrix} 1 & 2 & 1 \\ 3 & 2 & 1 \end{bmatrix}$$

Note that this market model does not contain any risk-free asset. We can check that this market model is incomplete and admits an arbitrage:

- *Since the range of* $\mathcal{M}_{S_1}^\top$ *is at most* $2 < 3$ *(the number of outcomes), the market is incomplete. For example* $X = (0, 1, 0)$ *cannot be replicated.*
- *The market admits an arbitrage. To check it, we look at equation* (1.24)*. To solve the linear system we extend the matrix and perform matrix reduction operations to get*

$$\left(\begin{array}{ccc|c} 1 & 2 & 1 & 1 \\ 3 & 2 & 1 & 3 \end{array}\right) \to \left(\begin{array}{ccc|c} 1 & 2 & 1 & 1 \\ 2 & 0 & 0 & 2 \end{array}\right) \to \left(\begin{array}{ccc|c} 1 & 0 & 0 & 1 \\ 0 & 2 & 1 & 0 \end{array}\right)$$

  i.e. $p_1^{AD} = 1$, *and* $2p_2^{AD} + p_3^{AD} = 0$. *Since the last equation cannot have a strictly positive solution, we conclude the market admits an arbitrage.*

  Alternatively, one could explicitly propose an arbitrage strategy. For example, take $\boldsymbol{\theta} = [3, -1]^\top$. *Clearly it has not cost at the start time, produces zero outcome in case that* ω_1 *occurs, and positive outcome if either* ω_2 *or* ω_3 *occur.*

1.7.3 Proof of the F.T.A.P.

In this section, we prove the two fundamental theorems of asset pricing in the case when the probability space is finite. In Example (1.41) we used an extended matrix to study the no-arbitrage condition. Inspired by this, we define

$$\mathcal{M}_{ext} := [-\boldsymbol{S}_0 \ \mathcal{M}_{S_1}] = \begin{bmatrix} -S_0^0 & S_1^0(\omega_1) & \dots & S_1^0(\omega_k) \\ \vdots & \vdots & \ddots & \vdots \\ -S_0^n & S_1^n(\omega_1) & \dots & S_1^n(\omega_k) \end{bmatrix}$$

Note that we have changed the signs of the initial values of S_0. This will be useful to take profit of separation theorems in the following.

An interesting result that will help us prove the F.T.A.P. is the following alternative theorem. It essentially says that we can either find a point in the intersection of a linear system L and a hypercube K, or instead finding a vector that is perpendicular to all L and that points "within" K.

Theorem 1.42 (Rockafellar (1970))**.** *Let L be a subspace of $\mathbb{R}^N$, and let K be a possibly degenerate but non-empty hypercube i.e., $K = I_1 \otimes I_2 \otimes .. \otimes I_N$ where each I_i is an interval not necessarily bounded and possibly only a point. Then one and only one of the following alternatives holds:*

- *There exists a vector $\boldsymbol{z} \in L \cap K$*
- *There exists a vector $\boldsymbol{z}^* \in \mathbb{R}^N$ such that $\boldsymbol{z}^* \cdot \boldsymbol{a} = 0$ for all $\boldsymbol{a} \in L$ and $\boldsymbol{z}^* \cdot \boldsymbol{b} > 0$ for all $\boldsymbol{b} \in K$.*

We will not reproduce the proof, but rather give an intuitive explanation: if there is no point in the intersection of L and K, then we can divide the space $\mathbb{R}^N$ in two around L and K will be contained in one of the two parts. In particular L cannot be the whole space, and K cannot be unbounded in all directions. If we take any element of K, this point can be expressed as the sum of one element of L and one perpendicular to L. One can then prove that the perpendicular component will have the desired property of pointing "toward" K.

We can deduce two easy corollaries that will be useful in the following and illustrate the uses of the Theorem.

Corollary 1.43. *Let $A \in \mathbb{R}^{m \times n}$. If one can show that there is no solution $\boldsymbol{x} \in \mathbb{R}^n$ to $A\boldsymbol{x} > \boldsymbol{0}$ then there exists $\boldsymbol{y} \in \mathbb{R}^m$, $\boldsymbol{y} > \boldsymbol{0}$ such that $\boldsymbol{y}^\top A\boldsymbol{x} = 0$, for all $\boldsymbol{x} \in \mathbb{R}^n$.*

Proof. In Theorem 1.42, set $N = m$, L be the image of the matrix A, and $K = \mathbb{R}^m_{++} = \bigotimes_{i=1}^m (0, \infty)$.

We have that there is a vector $\boldsymbol{y} \in \mathbb{R}^m$ such that $\boldsymbol{y}^\top A\boldsymbol{x} = 0$ for all $\boldsymbol{x} \in \mathbb{R}^N$ and $\boldsymbol{y} \cdot \boldsymbol{b} > 0$ for all $\boldsymbol{b} > \boldsymbol{0}$. But this later inequality implies that $\boldsymbol{y} > \boldsymbol{0}$. Indeed, if a component y_i would be such that $y_i \leq 0$, one could choose $\boldsymbol{b} = \boldsymbol{e}_i$ (only one in the direction 'i') and obtain a contradiction.

□

Corollary 1.44 (A version of Farkas' lemma)**.** *Let $A \in \mathbb{R}^{m \times n}$. If one can show that there is no solution $\boldsymbol{x} \in \mathbb{R}^n$ to $A\boldsymbol{x} = \boldsymbol{b}$ then there exists $\boldsymbol{y} \in \mathbb{R}^m$, such that $\boldsymbol{y}^\top A\boldsymbol{x} = 0$, for all $\boldsymbol{x} \in \mathbb{R}^n$, and $\boldsymbol{y} \cdot \boldsymbol{b} > 0$.*

Proof. In Theorem 1.42, set $N = m$, L be the image of the matrix A, and $K = \{\boldsymbol{b}\}$ as proceed as in Corollary 1.43. □

First F.T.A.P.

Proof of Theorem 1.24 in finite probability space case. Claim: No arbitrage $\Leftrightarrow$ there exists a strictly positive SDF.

"" $\Leftarrow$": We are going to show that if a strictly positive SDF exists, any strategy that generates no losses and some gains must have a strictly positive initial cost.

Let M_1 be a strictly positive S.D.F. Let $\boldsymbol{\theta} \in \mathbb{R}^{n+1}$ be a portfolio such that $\boldsymbol{\theta}^\top \mathcal{M}_{S_1} \geq \mathbf{0}$ and, for some ℓ, $\boldsymbol{\theta}^\top \boldsymbol{S}_1(\omega_\ell) > 0$. Then, by definition of M_1, we have that, for each $i = 0, ..., n$.

$$S_0^i = \mathbb{E}[M_1 S_1^i] = \sum_{j=1}^k M_1(\omega_j) S_1^i(\omega_j) \mathbb{P}\{\omega_j\} \tag{1.25}$$

By linearity, we can then deduce that

$$\begin{aligned} \boldsymbol{\theta} \cdot \boldsymbol{S}_0 &= \sum_{i=0}^n \theta^i \sum_{j=1}^k M_1(\omega_j) S_1^i(\omega_j) \mathbb{P}\{\omega_j\} \\ &= \sum_{j=1}^k \underbrace{M_1(\omega_j)\mathbb{P}\{\omega_j\}}_{>0} \underbrace{\sum_{i=0}^n \theta^i S_1^i(\omega_j)}_{\geq 0 \text{ and } >0 \text{ if } j=\ell} > 0 \end{aligned}$$

Hence, the initial cost of $\boldsymbol{\theta}$ must be positive.

"⇒": Let us remark first that no-arbitrage implies that there is no solution to $\boldsymbol{\theta}^\top \mathcal{M}_{ext} > 0$. Then, by Corollary (1.43) (with $A = \mathcal{M}_{ext}^\top$), we can find a strictly positive vector $\boldsymbol{\eta} \in \mathbb{R}_+^{k+1}$ such that $\boldsymbol{\theta}^\top \mathcal{M}_{ext} \boldsymbol{\eta} = \mathbf{0}$ for all $\theta \in \mathbb{R}^{n+1}$. Hence,

$$\begin{aligned} (-\sum_{j=1}^k \theta^i S_0^i)\eta_0 + \sum_{i=0}^n \sum_{j=1}^k \theta^i S_1^i(\omega_j)\eta_j &= 0 \\ \Rightarrow \sum_{i=0}^n \sum_{j=1}^k \theta^i S_1^i(\omega_j) \frac{\eta_j}{\eta_0} &= \sum_{j=1}^k \theta^i p^i \end{aligned}$$

where we can divide using the strict positivity of η_0. To finish, take $M_1(\omega_j) := \frac{\eta_j}{\eta_0 \mathbb{P}[\omega_j]}$. □

Second F.T.A.P.

Proof of Theorem 1.25 in finite probability space case. Claim: Arbitrage-free market is complete ⇔ unique SDF.

""⇒":Consider the Arrow-Debreu securities. Since the market is complete, for each $j = 1, \ldots, k$ there exists $\boldsymbol{\theta}_j$ such that $\boldsymbol{\theta}_j \cdot \boldsymbol{S}_1 = \mathbb{1}_{\omega_j}$. But since the market is arbitrage-free, p_j^{AD} the price of $\mathbb{1}_{\omega_j}$ satisfies $p_j^{AD} = \boldsymbol{\theta}_j \cdot \boldsymbol{S}_0$.

Now, assume there are two SDFs M_1, $\tilde{M}_1$. Then, they satisfy $M_1(\omega_j) = \frac{p_j^{AD}}{\mathbb{P}\{\omega_j\}} = \tilde{M}_1(\omega_j)$ for all $j = 1, \ldots, k$, so they are equal.

"⇐": Assume the market model is not complete. Then, there exists $\boldsymbol{b} \in \mathbb{R}^k$ such that $\boldsymbol{b} \notin \text{Range}(M_{S_1}^\top)$. By Corollary 1.44 we can find a non-trivial $\boldsymbol{\eta} \in \mathbb{R}^k$ such that for any $\boldsymbol{\theta} \in \mathbb{R}^{n+1}$, $\theta^\top M_{S_1} \boldsymbol{\eta} = 0$. Without loss of generality, we pick η such that for some j^*, $\eta_{j^*} > 0$.

Now, let $\boldsymbol{p}^{AD}$ be a vector of AD prices. (we know it exists thanks to the first F.T.A.P.) and (1.6). Define

$$\alpha = \frac{1}{2} \max_{j:\eta_j > 0} \left\{ \frac{p_j^{AD}}{\eta_j} \right\}.$$

Note that $\alpha > 0$. Set $\tilde{\boldsymbol{p}}^{AD} := \boldsymbol{p}^{AD} - \alpha\boldsymbol{\eta}$. We have $\tilde{\boldsymbol{p}}^{AD} \neq \boldsymbol{p}^{AD}$, $\tilde{\boldsymbol{p}}^{AD} > 0$, and for all $\boldsymbol{\theta} \in \mathbb{R}^{n+1}$,

$$\begin{aligned} \boldsymbol{\theta}^\top \mathcal{M}_{S_1} \boldsymbol{p}^{\tilde{A}D} &= \boldsymbol{\theta}^\top \mathcal{M}_{S_1}(\boldsymbol{p}^{AD} - \alpha\boldsymbol{\eta}) && (1.26) \\ &= \boldsymbol{\theta}^\top \mathcal{M}_{S_1} \boldsymbol{p}^{AD} - \alpha\boldsymbol{\theta}^\top M_{S_1}\boldsymbol{\eta} && (1.27) \\ &= \boldsymbol{\theta}^\top \mathcal{M}_{S_1} \boldsymbol{p}^{AD} = \boldsymbol{\theta}^\top \boldsymbol{S}_0 && (1.28) \end{aligned}$$

Hence, we have two different sets of possible Arrow-Debreu prices. By remark (1.6), the claim follows.

□

1.8 Exercises

Exercise 1.1. Let X be an i.i.d. discrete time stochastic process and consider the filtration associated to it. For each of the following examples, establish if they are: Predictable, Adapted, Markovian or Martingales (see definitions 1.2, 1.13 and 1.21). Justify your answer by proving or giving a counterexample.

1. $\hat{X}_0 = 0$ and $\hat{X}_t = \frac{t-1}{t}\hat{X}_{t-1} + \frac{1}{t}X_t$ for $t > 0$
2. $\hat{X}_0 = 1$ and $\hat{X}_t = \exp(\frac{1}{t}\sum_{i=1}^t X_i)$ for $t > 0$
3. X is in addition a Gaussian process (that is, every $(X_{t_1}, \dots, X_{t_k})$ is jointly Gaussian), $\hat{X}_0 = 1$ and $\hat{X}_t = \exp(\sum_{i=1}^t X_i - t\mathbb{E}[X_1] - \frac{1}{2}t\mathrm{var}[\mathrm{X}_1])$ for $t > 0$
4. $\xi_0 = (0,1)^\top$ and for $\alpha_0, \alpha_1, \beta_1 \geq 0$ and $t > 0$,
$$\xi_t = \begin{pmatrix} \hat{X}_t \\ \sigma_t^2 \end{pmatrix} = \begin{pmatrix} \sigma_t X_t \\ \alpha_0 + \alpha_1 \hat{X}_{t-1}^2 + \beta_1 \sigma_{t-1}^2 \end{pmatrix}.$$
5. $\hat{X}_t = \max_{0 \leq s \leq t} X_s$
6. $\hat{X}_t = \min_{0 \leq s \leq t}(\{s : X_s > 1\} \cup \{t\})$
7. $\hat{X}_t = \min_{t \leq s}\{s : X_s > 1\}$.

Exercise 1.2. Consider a market model with $n = 1$ and $T = 2$, given by a generalised binomial market model, where $R_t^0 = 1$ and $R_t^1 \in \{d, u\}$ for $d < 1 < u$, but where the returns at different times are not necessarily i.i.d.

Let $0 < \delta < \frac{1}{2}$ and define $\boldsymbol{\pi}, \boldsymbol{\pi}^+$ and $\boldsymbol{\pi}^-$ by

$$\begin{aligned} \pi_t &= \left[\frac{1}{2}, \frac{1}{2}\right]^\top ; \text{ for } t = 1, 2 \\ \pi_1^+ &= \pi_1^- = \pi_1 \\ \pi_2^+ &= \left[\frac{1}{2} - \delta(\mathbb{1}_{\{R_1^1 = u\}} - \mathbb{1}_{\{R_1^1 = d\}}), \frac{1}{2} + \delta(\mathbb{1}_{\{R_1^1 = u\}} - \mathbb{1}_{\{R_1^1 = d\}})\right]^\top ; \\ \pi_2^- &= \left[\frac{1}{2} + \delta(\mathbb{1}_{\{R_1^1 = u\}} - \mathbb{1}_{\{R_1^1 = d\}}), \frac{1}{2} - \delta(\mathbb{1}_{\{R_1^1 = u\}} - \mathbb{1}_{\{R_1^1 = d\}})\right]^\top ; \end{aligned}$$

1. Can $\boldsymbol{\pi}, \boldsymbol{\pi}^+$ and $\boldsymbol{\pi}^-$ be interpreted as valid strategies in terms of proportions (see remark 1.2 ? Justify your answer.

2. Construct an explicit instance of a generalised binomial model where $S^{\pi^+} > S^{\pi^-}$, where this is taken to mean $S^{\pi^+}(\omega) > S^{\pi^-}(\omega)$ for (almost) all $\omega \in \Omega$.

3. Give an explicit instance of a generalised binomial model where $S^{\pi^+} < S^{\pi^-}$.

4. Can you construct an instance of the model where $S^{\pi}(\omega) > S^{\pi^-}(\omega)$ and $S^{\pi}(\omega) > S^{\pi^+}(\omega)$?

Exercise 1.3. Consider a multiperiod market model where S^0 is a money market account and M is a stochastic discount factor. Show that for all $t = 0, \ldots, T-1$; $i = 0, \ldots, n$:

$$\mathbb{E}_t[R^i_{t+1}] - R^0_{t+1} = -R^0_{t+1}\mathrm{Cov}_t\left(\frac{M_{t+1}}{M_t}, R^i_{t+1}\right). \tag{1.29}$$

Note: The previous exercise shows that when stochastic discount factor exists, the conditional risk premium is determined by the covariance of the asset and the geometrical increment of the SDF. Note also, that the sign structure is very interesting. In general, as we expect risk premia and risk-free returns to be positive, we expect to see a negative correlation between the SDF and the return of market instruments.

Exercise 1.4. Consider a one-period model, and set $\Omega = \{\omega_1, \omega_2, \omega_3\}$, and assume that $\mathbb{P}$ is such that $\mathbb{P}[\omega_i] > 0$ for $i = 1, 2, 3$. Assume we have three assets with no pay-off and such that

$$\boldsymbol{S}_0 = (1, 2, 7)^\top.$$

and

$$\boldsymbol{S}_1(\omega_1) = \begin{bmatrix} 1 \\ 3 \\ 9 \end{bmatrix}, \boldsymbol{S}_1(\omega_2) = \begin{bmatrix} 1 \\ 1 \\ 5 \end{bmatrix}, \boldsymbol{S}_1(\omega_3) = \begin{bmatrix} 1 \\ 5 \\ 10 \end{bmatrix}.$$

- Is there any risk-less asset in this market? If so, what is the risk-free rate?
- Calculate the return and risk premia of the risky assets (Assume the probability law is uniform).
- Is the market complete? Arbitrage free? Justify your answers.

Exercise 1.5. Still on a one-period model, set $\Omega = \{\omega_1, \omega_2\}$, and assume that there are two assets, a risk-free asset with return R^0 and a risky asset. Let us call $R^u := R^1_1(\omega_1)$ and $R^d := R^1_1(\omega_2)$, and assume that $R^u > R^d$.

1. Find conditions on R^0, R^u, R^d equivalent to the absence of arbitrage.

2. Compute the risk neutral probabilities of each state

3. Assuming these conditions hold, compute the unique vector of Arrow-Debreu prices

4. Suppose that a *call option* is introduced in the market: it pays $(S^1_1 - K)^+$, for some K. Assuming that the price of the risky asset is p, compute the price of the call option for $K = pR^0$ if the assumptions found before hold.

Exercise 1.6. We say that a market has *the law of one price* if for any two portfolios represented by $\boldsymbol{\theta}$ and $\boldsymbol{\xi}$ such that $\boldsymbol{\theta} \cdot \boldsymbol{S}_1 = \boldsymbol{\xi} \cdot \boldsymbol{S}_1$, we have $\boldsymbol{\theta} \cdot \boldsymbol{S}_0 = \boldsymbol{\xi} \cdot \boldsymbol{S}_0$.

- Show that an arbitrage-free market has the law of one price.
- Give an example of a market that admits arbitrage but such that the law of one price is satisfied.

Exercise 1.7. Consider a one-period market model. Let $M_1, \tilde{M}_1$ be two strictly positive SDFs. Show that any convex combination of the two is also a strictly positive SDF, i.e. show that

$$M_1^{\alpha} = \alpha M_1 + (1-\alpha)\tilde{M}_1, \text{ for } \alpha \in (0,1),$$

is also a strictly positive SDF.
Note: We are showing that the set of SDFs is a *convex set*.

Exercise 1.8. Show that a price p^{new} is arbitrage-free for an asset with final value S_T^{new} (see Definition 1.33) if and only if $p^{new} = \mathbb{E}[M_T S_T^{new}]$ for some strictly positive SDF (M).

Exercise 1.9. Show that if $\tilde{p}$, $\hat{p}$ are two arbitrage-free prices for an asset, then any convex combination of the two prices is also an arbitrage free price.

Exercise 1.10. Let $A(H)$ be the set of all arbitrage free prices for a European claim with with payoff H payed at time $\bar{T}$ with $0 < \bar{T} \leq T$. Show that $A(H)$ is bounded below and above.
Hint: Use Exercise 3.6.1.

Exercise 1.10 together with Exercise 1.9 imply that $A(H)$ is a bounded interval.

Exercise 1.11. Let

$$p_{sup}(H) = \sup\{A(H)\},$$

i.e. the upper limit of the interval $A(H)$, where $A(H)$ is defined in Exercise 1.10. Show that there exists a strategy such that $S_0^{\boldsymbol{\theta}} = \gamma$, $\mathbb{P}[S_1^{\boldsymbol{\theta}} \geq D] = 1$ and $\mathbb{P}[S_1^{\boldsymbol{\theta}} > D] > 0$ if and only if $\gamma \geq p_{sup}(D)$ and $\gamma \notin A(D)$.

Hint: The "if" implication is harder. Remember that this means that if we would introduce the asset at this price, an arbitrage opportunity would appear in the extended market.

1.9 Summary

- Market models are a mathematical representation of the time-series of prices. They are useful to solve problems in mathematical finance like pricing, risk measurement and portfolio choice.
- Probability theory and stochastic processes provide a framework and properties to construct meaningful models.
- Simple market models are based on the i.i.d returns framework (like the log-normal and binomial models). More complex models (like Markovian or non-Markovian models) are sometimes more appropriate, though.

- Two important properties of markets are absence of arbitrage and completeness. These are idealizations towards which real markets can aspire to.
- Under quite strong assumptions (no transaction costs, full information, possibility to buy or sell any fraction) we presented a theory that links the above properties with the existence and uniqueness of a special process called a Stochastic Discount Factor. This relation is the subject of the Fundamental Theorems of Asset pricing.
- The SDF plays an essential role in asset pricing: it can be used to obtain arbitrage-free prices, which in turn impose a limitation on possible super-hedging strategies.
- Risk-neutral measures are also important objects: they are defined in such a way that all discounted prices are martingales, and so that there is no mean excess return. We can find a one-to-one correspondence between risk neutral measures and strictly positive SDFs, so they have similar market properties. Many practitioners prefer to use risk-neutral measures since they can be introduced directly on the modelling stage.
- We can give a fuller understanding to the above notions in the simpler case where the set of possible outcomes Ω is assumed to be finite. In this case, all probability statements can be translated into simple linear algebra statements.

Chapter 2

Utility functions

> Money has no utility to me beyond a certain point.
>
> *Bill Gates*

2.1 Utility functions to model behaviour

Now that we have presented an abstract theory on how to model the market, we can proceed to model investor's decisions and their connection to risk.

A very important idea that underlies such an effort is the assumption that the choices that investors make are not completely arbitrary or random, but rather that they can be modelled as consequences of the information at hand. This is an old idea, but got a quantitative taste in the 18th century, where the concept of *utility* as a measurement or degree of "well-being" or "pleasure" was proposed and develop by, among others, Hume and Bentham[1]. This concept, although highly controversial from a philosophical point of view, has been very successful and useful in economics, in particular supporting the development of a whole area now known as microeconomics.

An additional important element in the utility theory is its connection to uncertainty. Probably the first quantitative connection between the two concepts is found in D. Bernoulli's solution to the famous "St. Petersburg Paradox" [2], proposed in 1738, in a paper called *Exposition of a new theory on the measurement of risk*. In this paper, the concept of expected utility first appeared. Later, these ideas became more solidly-founded when derived from theory based on a rationality assumption and a simple set of rules thanks to the axiomatic development of Von Neumann and Morgenstern in 1947 (we include a short description in Section 2.6 below).

The utility maximisation theory, the centre of the present chapter, allows for a very convenient mathematical treatment of problems as pricing, portfolio choice and risk measurement.

[1]Yes, the Bentham whose auto-icon can be found on the South Cloisters here at UCL

[2]The original paper can be found translated from Latin in Bernoulli (1954) . For a quick review, check `https://en.wikipedia.org/wiki/St._Petersburg_paradox`

However, empirical studies including clinical behavioural data show several limitations on the capacity of utility theory to act as a *descriptive* theory : individuals were not consistent in their choices, and there were non-rational biases related to, for example, framing and relative risk aversion, that were observed in empirical studies. This lead to a revision of the paradigm initiated by the Prospect Theory of D. Kahneman and A. Tversky (see for example Tversky and Kahneman (1986) and Tversky and Kahneman (1992)). The works of other Nobel laureates like Schiller also questioned rationality on the markets. A branch of Behavioural Economics has developed ever since. For further reading including a detailed summary, other criticisms, and extensions: see Gilboa (2009).

Even with its flaws, utility theory and the rational assumptions remain relevant. On the one hand, they provide a natural reference or benchmark: for example, one characterises exuberance by comparison with a rational market. Hence, utility theory can be used to study a model where everything is *as it should be*, an use it to characterise deviations from it seen in practice. This means that the theory has a *normative* value. A second observation is that even if the simple utility theory is flawed, the essential idea of being able to capture an investor's (and in general people's) preferences quantitatively is more alive than ever: the changes brought by Behavioural Economics entailed only modifications of the model, not a fundamental reconsideration of the idea that decisions can be modelled. This notion, that people's decisions can be characterised, lies behind the extraordinary revolution in 'client knowledge' and 'prediction analysis' driven by machine learning, supported by the large amount of data gathered in our ever connected world and used by companies and governments alike to push for their interests.

All the above makes it relevant to understand and study the simple utility function paradigm in one period. The idea of utility function means essentially that investors establish their preferences by using a function that measures their satisfaction and drives their choices.

In the classical economical framework this well-being is written as a function of *consumption.* For simplicity, in the one-period case we can express it in terms of wealth. Thus, in a one-period framework, we define a utility function as follows:

Definition 2.1. A utility function (in one-period) is an increasing continuous mapping $u : E \to \mathbb{R}$, with E an interval in $\mathbb{R}$.

Utility functions encode our satisfaction or well-being. This is why it is understood as an increasing function: well-being should not decrease when wealth increases (in other words, we expect that *more should feel better than less*). Continuity, on the other hand, follows from the intuition that well-being should not have sudden changes at arbitrary levels.

Example 2.2. *Here are some examples of utility functions:* $u(x) := -\exp(-x)$ *(exponential, defined over* $\mathbb{R}$*) ,* $u(x) := \log(x)$ *(logarithmic, defined over* $\mathbb{R}_+$*).*

What happens when a market participant faces an uncertain bet? The key proposition of Bernoulli is that investors are sensitive to their expected well-being Hence, among two possible investments, *an investor will choose the one with the biggest expected utility.* That is, when given a choice between two different random investments W_1 and W_2,

$$\text{the investor strictly prefers } W_1 \text{ over } W_2 \text{ if and only if } \mathbb{E}[u(W_1)] > \mathbb{E}[u(W_2)]$$

If both investments share the same expected utility they are said to be equivalent.

2.2 Risk attitude and certainty equivalence

One of the consequences of modelling the preferences of an investor as being ruled by maximising expected utility is that we can understand how they react to risk: for example, we can understand if they fear or look for risk.

Definition 2.3. An investor is said to be (weakly) risk-averse if

$$u(\mathbb{E}(W)) \geq \mathbb{E}[u(W)]. \tag{2.1}$$

It is called (weakly) risk-seeking if

$$u(\mathbb{E}(W)) \leq \mathbb{E}[u(W)]. \tag{2.2}$$

In other words, a risk-averse investor prefers to avoid a fair bet. This is precisely Bernoulli's explanation to solve St. Petersburg Paradox: people are, in general, risk-averse, and hence they are willing to offer a quantity much smaller than the (infinity) expectation value.

Mathematically, there are some properties of u implying risk aversion. We start by including some definitions.

Definition 2.4. A set $E \subset R^n$ is convex if for all $x, y \in E$ and $\alpha \in [0, 1]$, $\alpha x + (1 - \alpha)y \in E$.

Example 2.5. *Consider E the closed unit ball in $\mathbb{R}^d$, $E = \{x \in \mathbb{R}^d : |x| \leq 1\}$. Note that for $x, y \in E$, $\alpha \in (0, 1)$*

$$|\alpha x + (1 - \alpha)y| \leq \alpha|x| + (1 - \alpha)|y| \leq 1 \Rightarrow \alpha x + (1 - \alpha)y \in E.$$

Definition 2.6. Let $E \subset \mathbb{R}^n$ convex and $u : E \to \mathbb{R}$. A function u is concave if for all x in its domain,

$$u(\alpha x + (1 - \alpha)y) \geq \alpha u(x) + (1 - \alpha)u(y); \text{for all } x, y \in \mathbb{R}; \alpha \in [0, 1].$$

In the case where $n = 1$, if u is twice differentiable it is concave if and only if $u''(\cdot) \leq 0$.

Example 2.7. *The functions $f_1(x) = \log(x)$ and $f_2(x) = a - bx^2$ for $b \geq 0$ are concave.*

Theorem 2.8 (Jensen's inequality)**.** *Let W be a random variable and $u : \mathbb{R}^n \to \mathbb{R}$ be concave and such that $\mathbb{E}[|u(W)|] < \infty$. Then*

$$\mathbb{E}[u(W)] \leq u(\mathbb{E}[W])$$

Thus, if the utility function u is concave, the investor is (weakly) risk-averse. The converse is also true: risk-averse investors must have a concave utility function.

Proposition 2.9. *Assume that $\mathbb{P}$ is continuous* [3] *and that for every random variable W,*

$$\mathbb{E}[u(W)] \leq u(\mathbb{E}[W]).$$

Then u is concave.

As a consequence of this proposition, at least in the continuous case, there is an equivalence between a risk-averse and a concave utility function. Due to this equivalence, concavity is sometimes directly used as the key property to define risk-aversion.

Exercise. Show Proposition 2.9

[3] i.e., it has continuous density with respect to the Lebesgue measure

Certainty equivalence

An important tool for investors that decide using their expected utility as a criterion, is to understand what would be a deterministic equivalent of a given random wealth.

Definition 2.10. Let W be a wealth gamble. We denote x_W to be *certainty equivalent* of W:

$$x_W = u^{-1}(\mathbb{E}[u(W)]).$$

The name comes form the fact that x_W is a deterministic value that produces the same utility on average than W. In view of our assumptions, this value is unique.

Note that an investor is (weakly) risk-averse if $x_W \leq \mathbb{E}[W]$ for all random wealth W. The difference between these two quantities can be understood as a premium that the investor would be willing to pay in order to avoid the uncertainty.

Definition 2.11. Let W be a wealth gamble. The *certainty premium*[4] of W is the (unique) number $\eta = \eta(W)$ such that

$$u(\mathbb{E}[W] - \eta) = \mathbb{E}[u(W)]. \tag{2.3}$$

2.3 (Arrow-Pratt) coefficients of risk aversion

Let us fix a given random wealth W. Intuitively, the more risk-averse investors are, the bigger the certainty premium they are willing to pay. Thus, the certainty premium can be understood as a measure of risk aversion.

Finding the certainty premium might require numerical techniques [5]. We look in this section for a local approximation to a value that depends only on the derivatives of u when they exist.

To do this, let us fix a constant $w \in \mathbb{R}$, and consider a small random perturbation ξ, such that

$$\mathbb{E}[\xi] = 0; \quad \mathrm{var}(\xi) = 1; \quad \mathbb{E}[\xi^3] < \infty.$$

Let us take a small value ϵ. We are going to find the certainty premium required on the random endowment $W^\epsilon = w + \epsilon\xi$ as $\epsilon \to 0$. Clearly, $\mathbb{E}[W^\epsilon] = \mathbb{E}[w + \epsilon\xi] = w$. For simplicity, we write in this section $\eta(\epsilon) := \eta(W^\epsilon)$.

Let us suppose that u and η are two times continuously differentiable on a neighbourhood of w, and that $u'(w) > 0$ (i.e. the utility grows at that point). Applying the Taylor theorem as functions of ϵ around 0, we have on the one hand that

$$\mathbb{E}[u(W^\epsilon)] = \mathbb{E}[u(w+\epsilon\xi)] = \mathbb{E}\left[u(w) + u'(w)\xi\epsilon + \frac{1}{2}u''(w)\xi^2\epsilon^2 + R_3(\epsilon)\right] = u(w)+\frac{1}{2}u''(w)\epsilon^2+o(\epsilon^2);$$

where the notation $o(\epsilon^2)$ denotes a value that goes to zero faster than ϵ^2. On the other hand,

$$u(w - \eta(\epsilon)) = u(w) - u'(w)\eta'(0)\epsilon + \frac{1}{2}[(\eta'(0))^2 u''(w) - u'(w)\eta''(0)]\epsilon^2 + o(\epsilon^2).$$

[4]Also known as risk premium in books like Back (2010)
[5]See for example the Python Notebook 4 on Expected utility

By definition, we have equality between the previous two terms. By a scale analysis, we conclude that $\eta'(0) = 0$. Hence, the leading term is quadratic and

$$\eta''(0) = -\frac{u''(w)}{u'(w)}.$$

Integrating (or replacing on the Taylor expansion), we have that for $|\epsilon| \ll 1$,

$$\eta(\epsilon) \approx -\frac{u''(w)}{u'(w)}\frac{\epsilon^2}{2}. \tag{2.4}$$

Since ϵ^2 represents the variance of W^ϵ, we see that for small perturbations, the certainty premium is linear with respect to the variance, with a coefficient depending only on derivatives of u.

2.3.1 Absolute coefficient of risk aversion

Inspired by (2.4), we introduce the following quantity.

Definition 2.12. The coefficient of "*absolute*" risk aversion at the wealth level w is defined by

$$\alpha(w) = -\frac{u''(w)}{u'(w)}. \tag{2.5}$$

Therefore, the risk-premium depends (at least as an approximation) on the utility function through the risk aversion, and on the gamble through its variance.

Remark 2.1. Some observations:

1. The absolute risk aversion is **not** affected by a monotone affine transform of the utility function.
2. A risk averse investor has an absolute risk aversion $\alpha(\cdot) \geq 0$.

Exercise. Justify the observations in Remark 2.1.

2.3.2 Relative risk aversion coefficient

If instead of studying the case $W = w + \epsilon\xi$, we consider $W = w + w\hat{\epsilon}\xi$, i.e. if we perturb by a quantity proportional to the actual wealth, and we express the certainty premium in relative terms, i.e. $\eta = w\hat{\eta}$ we would have found that

$$\hat{\eta}''(0) = -\frac{wu''(w)}{u'(w)} = w\alpha(w). \tag{2.6}$$

This motivates the following definition.

Definition 2.13. The coefficient of "*relative*" risk aversion at the wealth level w is defined by

$$\rho(w) = w\alpha(w) = -w\frac{u''(w)}{u'(w)}. \tag{2.7}$$

Remark 2.2. Relative risk aversion coefficients convey information in the case when the original wealth is strictly positive.

2.4 Notable examples of utility functions

In this section we examine some utility functions commonly examined in the literature.

2.4.1 Constant absolute risk aversion (CARA)

If absolute risk aversion is the same at **every** level of wealth, that is, $\alpha(w) \equiv \alpha \in \mathbb{R}$, then an investor has CARA utility.

We can identify two cases. If $\alpha = 0$, then

$$u(x) = ax + b$$

where $a \geq 0$ for u to be non-decreasing. This is case is known as *linear utility* or *risk neutral* given that $\alpha \equiv 0$. On the other hand, if $\alpha \neq 0$, we get

$$u(x) = -ae^{-\alpha x} + b \tag{2.8}$$

where α is the absolute risk aversion parameter, and a must satisfy $a\alpha > 0$ and $b \in \mathbb{R}$. In the case $\alpha > 0$, then $a > 0$ and this is a negative exponential utility.

CARA utility implies that risk aversion is insensitive to the total wealth . We will verify that this implies that **portfolio choices** are independent of the initial wealth.

We can calculate the certainty premium η for the CARA utility as follows:

$$\begin{aligned} u(\mathbb{E}[W] - \eta) &= \mathbb{E}[u(W)] \\ \Longrightarrow -ae^{-\alpha(\mathbb{E}[W]-\eta)} + b &= -a\mathbb{E}[e^{-\alpha W}] + b \\ \Longrightarrow \eta &= \frac{1}{\alpha}\log\left(\mathbb{E}\left[e^{-\alpha \hat{W}}\right]\right), \end{aligned}$$

for $\hat{W} = W - \mathbb{E}[W]$. Observe that in this case the certainty premium η does not depend on $\mathbb{E}[W]$, a nor b but only on $W - \mathbb{E}[W]$.

Example 2.14. *If we assume that $W_1 \sim \mathcal{N}(0, \sigma^2)$ then*

$$\eta_1 = \frac{1}{\alpha}\log\left(\mathbb{E}[e^{-\alpha W_1}]\right) = \frac{1}{\alpha}\log\left(e^{\frac{1}{2}\alpha^2\sigma^2}\right) = \frac{1}{2}\alpha\sigma^2. \tag{2.9}$$

That is, in the case of CARA utility and a normally distributed wealth gamble, the approximation of (2.4) *is actually exact for every σ.*

On the other hand, if $W_2 \sim U([-a, a])$, then we would have

$$\eta_2 = \frac{1}{\alpha}\log\left(\mathbb{E}[e^{-\alpha W_2}]\right) = \frac{1}{\alpha}\log\left(\frac{\sinh(a\alpha)}{a\alpha}\right). \tag{2.10}$$

For this example, let $\psi(x) = \sinh(x)/x$. Since log *is strictly increasing, the only way to assure that both premia are the same is by choosing $\sigma^2 = 2\alpha^{-2}\psi(\alpha a)$. However, this does not imply that we would have the same variance since the one on the uniform case is simply $a^2/3$.*

Example 2.15. *Imagine two investors P(oor) and R(ich), with wealths $w_P = 10$ and $w_R = 100000$ and the same utility function $u(x) = -e^{-\alpha x}$, face the following situation. A coin is tossed, the investors have to pay* 1 *if tail shows up, otherwise they obtain* 1*. Thus, $W = w + 1$ and $W = w - 1$ each with probability $1/2$ and $\sigma^2 = \mathrm{var}(W) = 1$. Since investor R has far more money than investor P and the game is fair, one would expect that R would pay only very little to avoid the game (and much less than P). However, this is not true in the case of CARA utility.*

2.4.2 Constant relative risk aversion (CRRA)

Recall the definition of ρ, the coefficient of relative risk aversion, in (2.7). If relative risk aversion is the same at **every** level of wealth, that is, $\rho(w) \equiv \rho \in \mathbb{R} \setminus \{0\}$, then an investor has CRRA utility. Since relative risk aversion is only sensible with positive wealths, we only cosnider the case $E = (0, \infty)$ here . Any CRRA utility function with **positive** relative risk aversion has **decreasing** absolute risk aversion

$$\alpha(w) = \frac{\rho}{w}. \tag{2.11}$$

Any monotone CRRA utility function is of the form:

(i) $u(x) = a \log(x) + b$;

(ii) $u(x) = ax^\gamma + b; \quad 0 < \gamma < 1$;

(iii) $u(x) = -ax^\gamma + b; \quad \gamma < 0$.

for $a > 0$, $b \in \mathbb{R}$. Cases (ii) & (iii) can be summarised (and rewritten) by

$$u(x) = \frac{a}{1-\rho} x^{1-\rho} + b; \quad \rho \in (0,1) \cup (1, \infty). \tag{2.12}$$

Thus, CRRA implies either *logarithmic* or *power utility.*

Remark 2.3. We make a few observations concerning the logarithmic utility:

1. The logarithmic utility has constant relative risk aversion $\rho = 1$.
2. The logarithmic utility is the limiting case of the power utility for $\rho \to 1$, in the following sense:

$$\frac{1}{1-\rho} x^{1-\rho} - \frac{1}{1-\rho} \longrightarrow \log(x) \quad (\rho \to 1) \tag{2.13}$$

(Recall that monotone affine transformations do not change the preference structure.)

Exercise. Prove (2.13). (Hint: use L'Hôpital's rule)

The fraction of wealth that a CRRA investor would pay to avoid a gamble that is **proportional** to the initial wealth, is **independent** of the investor's wealth. To wit, if there is a wealth gamble $W = w(1 + \xi)$ for some w and some ξ with $\mathbb{E}[\xi] = 0$ then the risk premium is $w\hat{\eta}$ for some $\hat{\eta} \geq 0$, which does not depend on w.

Exercise. In the case of a CRRA investor, consider a wealth gamble $W = w(1 + \xi)$, where ξ is a random variable with $\mathbb{E}[\xi] = 0$ and $w > 0$ is a constant. Prove that the certainty premium for W is of the form $\eta = \eta' w$, where η' does not depend on w. (Hint: Recall the functional form of a CRRA utility function u.)

2.4.3 Hyperbolic absolute risk-averse utility functions (HARA)

We say that a utility function belongs to the HARA family, if it has the following form

$$u(x) = a \left(\frac{1-\gamma}{\gamma} \right) \left(\frac{x}{1-\gamma} - \hat{x} \right)^\gamma + b$$

where $\gamma \neq 0$, $\frac{x}{1-\gamma} - \hat{x} > 0$ and $a > 0$. Note that the domain E where these are defined has to be carefully defined.

HARA utilities encompass, among other, the CRRA case and the CARA case (the latter as a limit case), and the linear utility case.

2.4.4 Quadratic utility

Whenever $\gamma = 2$ and $\hat{x} > 0$, we get the quadratic utility function. For example, when $a = 1, b = 0$ it reads

$$u_Q(x) = -\frac{1}{2}(x - \hat{x})^2. \tag{2.14}$$

Note that the utility function is concave and therefore shows risk aversion. However, this kind of utility function is not very realistic (an investor would feel very unhappy to have a very large wealth!). Indeed, to make sense of this utility function we need to add the requirement $E = (-\infty, \hat{x})$.

Computing expected utility we note that

$$\mathbb{E}[u_Q(W)] = \mathbb{E}\left[-\frac{1}{2}(W - \hat{x})^2\right] = -\frac{1}{2}\{(\mathbb{E}[W] - \hat{x})^2 + \mathrm{var}(W)\} \tag{2.15}$$

Preferences over wealth gambles therefore depend **only** on their mean **and** variance when an investor has quadratic utility.

Exercise. Show (2.15).

2.5 Utility functions in a multi-period setting

Up to now, we focused on how to understand utility functions in one-period. We need to consider a generalisation of utility functions that make multiple periods intervene.

As stated in the introduction, the classical economic theory postulates that agents would like to maximise their perceived expected happiness, via the use of a utility function of their current and future consumption.

Generalising the case of one dimension we say that a multi-period utility function is a continuous and entry-wise increasing function $u : E \to \mathbb{R}$, with $E \subset \mathbb{R}^{T+1}$ a convex set. It determines an investor's choices.

We can also generalise the concept of risk-aversion: for any adapted stochastic process C in [0,T], and any $0 \leq s < T$, we have that

$$\mathbb{E}_s[u(C)] \leq u(C_0, \ldots, C_s, \mathbb{E}_s[C_{s+1}], \mathbb{E}_s[C_{s+2}], \ldots \mathbb{E}_s[C_T]); \quad \text{almost surely.}$$

Remark 2.4. In one period, we could interpret well-being as being wealth or consumption almost indistinctly. By contrast, in a multi-period setting, there is a difference since one quantity refers to a *stock* (cumulative) and the other to a *flow* (changes). In the above, we interpret the entries of a multi-period utility function as either consumption of additional wealth, i.e., we keep track of flows.

Jensen's inequality (this time in a more complicated set) guarantees that concavity (in $\mathbb{R}^T$) is a sufficient condition for risk aversion.

Given that we have also a time component, it makes sense to understand the behaviour of an investor with respect to time. Most investors show a preference for obtaining their consumption or income sooner rather than later. Let $\{\boldsymbol{e}_1, \ldots, \boldsymbol{e}_T\}$ denote the canonical vectors on $\mathbb{R}^T$. We say that a utility function is (weakly) time-discounting if there exists $\epsilon > 0$ such that for all $0 \leq \ell \leq t \leq T$ and all $r \in (0, \epsilon)$

$$u(\boldsymbol{x} + r\boldsymbol{e}_t) \leq u(\boldsymbol{x} + r\boldsymbol{e}_\ell).$$

2.5.1 Time additive utility

A particularly useful family of utility functions in this setting is known as *time-additive utility functions*. A utility function is time-additive with constant discount factor $0 < \delta < 1$, if

$$u(\boldsymbol{x}) = \sum_{t=0}^{T} \delta^t \hat{u}(x_t)$$

where $\hat{u}$ is a one-period utility function.

This represents the situation whereby investors are equally satisfied among periods up to a factor representing the preference of having their consumption now. Note that in this case, the increasing and concavity conditions on u become the usual ones on $\hat{u}$.

This type of utility immediately extends the studied families to several periods.

2.6 Axiomatisation of utility maximisation(*)

As it seems like a very ad-hoc assumption, several works have been devoted to deduce the utility function maximisation in agents preferences, from simpler assumptions related to rational preference. Here we focus on the one-period case. In **?**, an axiomatisation was presented that proved an equivalence between the expected utility approach and rational choice.

To describe this set-up, the key element is the set $\mathcal{A}$ of all *probability distributions on the measurable space* $(\mathbb{R}, \mathcal{B})$ *of real numbers with Borel's sigma algebra.* More general spaces can be used (and indeed need to be used when considering dynamic problems instead of our one-period case). Thus, any wealth profile is identified with its probability law. Consequently, two wealth profiles with the same distribution are treated as identical. Recall that

$$X \sim \nu \in \mathcal{A} \text{ iff } \mathbb{E}[f(X)] = \int f(x)\nu(dx) \text{ for all measurable } f.$$

In economic literature, these probability distributions are sometimes called "lotteries".

We consider a binary relation $\precsim$ on the space $\mathcal{A}$ such that $\nu_2 \precsim \nu_1$ is interpreted as "the investment in ν_1 is (weakly-)preferable to the investment on ν_2": it means that an investor would not be unhappy if they had to accept the lottery ν_1 over the lottery ν_2. We denote the strong preferences by $\nu_2 \prec \nu_1$ which means that $\nu_1 \not\precsim \nu_2$.

Von-Neumann-Morgenstern propose the following rules that $\precsim$ should obey:

Property 2.16 (Von Neumann-Morgenstern)**.**

1. *Completeness: For all* $\nu_1, \nu_2 \in \mathcal{A}$*, either* $\nu_1 \precsim \nu_2$ *or* $\nu_2 \precsim \nu_1$.
2. *Transitivity: If* $\nu_1 \precsim \nu_2$ *and* $\nu_2 \precsim \nu_3$ *then* $\nu_1 \precsim \nu_3$.
3. *Continuity (or Archimedian): For all* $\nu_1, \nu_2, \nu_3 \in \mathcal{A}$*, there exist* $\alpha \in (0,1)$ *such that if* $\nu_1 \prec \nu_2 \prec \nu_3$*, then*[6]
$$\alpha\nu_1 + (1-\alpha)\nu_3 \prec \nu_2 \prec (1-\alpha)\nu_1 + \alpha\nu_3$$
4. *Independence: For all* $\nu_1, \nu_2 \in \mathcal{A}$*,*
$$\nu_1 \precsim \nu_2 \quad \textit{implies} \quad \alpha\nu_1 + (1-\alpha)\nu_3 \precsim \alpha\nu_2 + (1-\alpha)\nu_3 \textit{ for every } \alpha \in (0,1) \textit{ and } \nu_3 \in \mathcal{A}$$

The stated assumptions can be interpreted in the following way: *completeness* means that an investor can always decide between any two investments; *transitivity* means that ordering of preferences are consistent; *continuity* means that we can replace one investment with a mix of this investment and a third one, provided that the event determining the mix is small enough (it is a rather technical assumption, whose main purpose is to ease the mathematics rather than to answer to a fundamental property). Finally, the *independence* assumption can be understood as follows: an individual prefers an investment over another if and only if they will still prefer it conditionally on any given event. It is a very strong assumption, and hence it has been debated and modified in the literature.

The main result in Von-Neumann-Morgenstern's theory is that there exists such a binary relation if and only if an agent acts *as if* they would like to maximise some *expected utility*, that is
$$\nu_1 \precsim \nu_2 \quad \text{if and only if} \quad \mathbb{E}[u(X_1)] \leq \mathbb{E}[u(X_2)], \tag{2.16}$$
for some (and therefore all) $X_1 \sim \nu_1, X_2 \sim \nu_2$.

Moreover, we can prove that the utility function is unique up to a *monotone affine transform.* This means that if instead of u, an investor decides using a map of the form $\widetilde{u}(\cdot) = a + bu(\cdot)$, where $a \in \mathbb{R}$ and $b > 0$ is considered, then the investor's preferences remain unchanged.

Further contributions refined these ideas. It is worth mentioning here the work of Savage (1954) who proposed a different set of axioms acting directly over random variables, and that is characterised not just by a concave function but also by a probability function (the subjective view on probability). This approach solves many of the inconsistencies of using the basic utility approach as a descriptive model. Many others, however were still unexplained (see for example Tversky and Kahneman (1992)) where a different approach where utility is defined with respect to a 'reference' or 'frame' is introduced. A yet alternative approach was based on introducing uncertainty aversion, that is, on admitting that the probability measure $\mathbb{P}$ is not fixed but rather that the investors have some doubts and consider a family of distributions around $\mathbb{P}$. The alternative probabilities are penalised according to how implausible they seem to the investor. For further reading see Gilboa (2009) and Föllmer and Schied (2011).

[6] This is well defined, as the convex combination of probability distributions is also a probability distribution. For general measurable spaces a strengthening of the Archimedian property to true continuity of $\precsim$ with respect to the weak topology is necessary. See Föllmer and Schied (2011)

2.7 Exercises

Exercise 2.1. Solve the following two problems.

1. An investor with initial wealth $w_0 = 200$ has utility function $u_A(x) = x^\gamma/\gamma$ for some $\gamma \in (0,1)$. She faces two options, and decides according to her expected utility. Option 1: doing nothing . Option 2: Betting in a lottery where she might lose 100 or win 100 with equal probability ($X_2 = 100$ or $X_2 = -100$, each with probability 0.5). How does she decide? What about another investor who has utility function $u_B(x) = -\exp(-x)$?
2. An investor with initial wealth $w_0 = 1$ has utility function $u_C(x) = \log(x)$ and decides between two investments. Investment 1 pays 10 with probability 1/3 and pays 30 with probability 2/3. Investment 2 pays 20 with probability 0.9 and nothing with probability 0.1. Which one does she chose?

Exercise 2.2. Check that the two functions in Example 2.7 are concave.

Exercise 2.3. Write a Python program that evaluate the expected log utility function if the wealth W follows a

1. Pareto distribution with $\alpha = 2$ and scale factor $x_m = 1/2$.
2. Exponential $\lambda = 1$
3. Log-normal, i.e. $W = e^X$ with $X \sim \mathcal{N}(0,1)$.

Exercise 2.4. A monotone affine function is an affine map that preserves ordering.

- Show that if $M : \mathbb{R} \to \mathbb{R}$ is a monotone affine map, then $M(x) = ax + b$ for some $a > 0$ and $b \in \mathbb{R}$.
- Show that every twice differentiable CARA utility function is a monotone affine transform of the form (2.8).

Exercise 2.5. Show that if $W \sim \mathcal{N}(0, \sigma^2)$, then the certainty premium for the CARA utility function is given by (2.9).

Exercise 2.6. What is the absolute risk aversion of a quadratic utility? of a CRRA utility? Are these coefficients increasing or decreasing in their domain?

Exercise 2.7. Prove that the utility function of (2.12) has constant relative risk aversion ρ

Exercise 2.8. Consider an investor with relative risk aversion ρ.

i. Verify that the fraction of wealth that he/she is willing to pay to avoid a gamble that is **proportional to wealth** is independent of the initial wealth.

ii. Consider a gamble ξ. Assume that the gross return of bet is $R = \exp(Z)$ with $Z \sim \mathcal{N}(-\sigma^2/2, \sigma^2)$. What is the average of R? Find the certainty premium that a CRRA investor with initial wealth w_0 would pay to avoid betting on R.

Exercise 2.9. This exercise, sourced from Back (2010), is based on a simple bid-ask model by Stoll (1978). Consider an individual with constant risk aversion coefficient α. Assume that, if nothing is done, the investor is set to have a wealth at the end of the period of $W \sim \mathcal{N}(w, \sigma_w^2)$. Consider a pay-off D, which is jointly Gaussian with W. Assume that D has mean m, variance σ^2 and $\mathrm{cor}(D, W) = \rho$.

i. Compute the maximum amount the individual would pay (at the end of the period) to obtain the pay-off D; that is, compute p_{BID} satisfying

$$E[u(W)] = E[u(W + D - p_{BID})|]$$

ii. Compute the minimum amount the individual would require (at the end of the period) to accept the pay-off $-D$; that is, compute p_{ASK} satisfying

$$E[u(W)] = E[u(W - D + p_{ASK})].$$

(*) Exercise 2.10. Show that the operator $\precapprox$ defined over the space of probability functions in $(\mathbb{R}, \mathcal{B})$ such that

$$\nu_1 \precapprox \nu_2 \Leftrightarrow \int u(x)\nu_1(\mathrm{d}x) \leq \int u(x)\nu_2(\mathrm{d}x)$$

for some increasing and concave u, satisfies the Von-Neumann-Morgenstern properties listed in Property (2.16)

Exercise 2.11. Consider an investor with log utility.

i. Construct a gamble W such that $\mathbb{E}[W] = \infty$

ii. Construct a gamble W such that $W > 0$ in every state of the world, but $\mathbb{E}[u(W)] = -\infty$
Hint: think about the St. Petersburg paradox.

2.8 Summary

- Utility functions are a way to model the decision-making process of agents and individuals. Mathematically we assume them as increasing and continuous functions of consumption. In the one-period case, it is equivalent to work with wealth instead of consumption.
- Under the expected utility maximisation paradigm, agents facing uncertainty aim to maximise their (perceived) expected utility.
- Risk-aversion is a tendency of certain agents to obtain a lower utility from a random (wealth/consumption) distribution than from its mean. We have several forms of measuring the self-aversion of a given agent including insurance premia and coefficients of risk-aversion.
- In a multi-period setting, it is convenient to consider also how consumption at different periods impacts on utility. If investors prefer to consume sooner rather than later, they are said to have time-discounting utility functions.

Chapter 3

Portfolio choice

When it comes to investing, there is no such thing as a one-size-fits-all portfolio.

Barry Ritholtz

In this chapter, we use the utility function theory to establish how an investor would decide to divide their wealth into a set of investments. This is known as *portfolio choice theory.*

We focus first in the one-period case (the investor evaluates the results only for one period): in this case, we just need to use a one-period market model, and most results can be deduced from a mild extension of traditional convex optimisation theory. In the second part of the chapter, we examine the multi-period case: here new tools like the Dynamic Programming Principle can be used.

Before, we continue, let us first give a quick review of the optimisation background needed for the rest of the Chapter.

3.1 Optimisation problems

In an optimisation problem, we look for one or more points within a given set of admissible values that would optimise (i.e. minimise or maximise) a given criterion function. More precisely, given $K \subset \mathbb{R}^d$, we define the problem

$$\max_{\boldsymbol{x} \in K} F(\boldsymbol{x}) \tag{3.1}$$

where K is the set of admissible values, F is the *criterion* or *objective function.* If the set $K = \mathbb{R}^d$, we say that the problem is *unconstrained.* Note that the minimisation problem is covered by considering $-F(x)$ instead. The set K can be explicitly or implicitly given. For example, a typical form is

$$K = \{\boldsymbol{x} \in \mathbb{R}^n : G(\boldsymbol{x}) \leq 0\} \cap \{\boldsymbol{x} \in \mathbb{R}^n : H(\boldsymbol{x}) = 0\}$$

for some functions G, H satisfying some assumptions as we shall see below.

We say that $\boldsymbol{x}^*$ is a (global) optimal to the optimisation problem (3.1) if $\boldsymbol{x}^* \in K$ and $F(\boldsymbol{x}^*) \geq F(\boldsymbol{y})$ for all $\boldsymbol{y} \in K$.

We say that $\boldsymbol{x}^*$ is a (local) optimal to the optimisation problem (3.1) if $\boldsymbol{x}^* \in K$ and there exists a neighbourhood of $\boldsymbol{x}^*$, $V(\boldsymbol{x}^*)$, such that $F(\boldsymbol{x}^*) \geq F(\boldsymbol{y})$ for all $\boldsymbol{y} \in K \cap V(\boldsymbol{x}^*)$.

There is a large mathematical theory on the conditions under which global and local solutions exist, understanding assumptions that could guarantee that they are unique, and characterising the minima with verifiable criteria. This area has created some beautiful and useful results, largely outside the scope of this module (the interested reader can refer to Boyd and Vandenberghe (2004) for an application driven approach and to Rockafellar (1970) or Clarke (2013) for a nice mathematical treatment). Here, we focus on stating some of the useful theoretical results and signalling how to solve numerically the problem (3.1).

Let us start by reminding the most important *existence* theorem for the solution of problem (3.1). Recall that $F : K \to \mathbb{R}$ is upper semi-continuous if and only if $\{\boldsymbol{x} \in K : F(\boldsymbol{x}) < a\}$ is an open set for every $a \in \mathbb{R}$ (in particular, continuous functions are upper semi-continuous).

Theorem 3.1. *If K is compact and F is upper semi-continuous, then problem* (3.1) *has at least one global solution.*

The proof of Theorem (3.1) can be found in most topology textbooks (see for example, Thm. 27.4 in Munkres (2014) where the claim asks for continuity but the proof only uses upper semi-continuity.).

The question of *how many global optima* and *how* to find them becomes simpler when F is concave.

3.1.1 The concave case

Optimisation theory has developed a good amount of important results in the case where we aim to maximise a concave function (Definition 2.6) over a convex set (Definition 2.4). Other results require some even stronger properties on the function F.

Definition 3.2 (Strict and strong concavity)**.** Let $K \subseteq \mathbb{R}^d$ be a *convex* set, and let $F : K \to \mathbb{R}$.

- The function is called *strictly concave* if

$$F(\alpha \boldsymbol{x} + (1-\alpha)\boldsymbol{y}) > \alpha F(\boldsymbol{x}) + (1-\alpha)F(\boldsymbol{y}).$$

- The function F is said to be *strongly concave* if there exists $\theta > 0$ such that for all $x, y \in K$ and $\alpha \in (0,1)$,

$$F(\alpha \boldsymbol{x} + (1-\alpha)\boldsymbol{y}) > \alpha F(\boldsymbol{x}) + (1-\alpha)F(\boldsymbol{y}) + \frac{\theta}{2}(\alpha)(1-\alpha)|\boldsymbol{x}-\boldsymbol{y}|^2. \tag{3.2}$$

  where $|.|$ is the canonical norm on $\mathbb{R}^n$.

Clearly, strongly concave functions are strictly concave, and these in turn are concave.

Proposition 3.3. *Let $K \subseteq \mathbb{R}^d$ be a convex set. Let $F : K \to \mathbb{R}$ be differentiable in all K, then*

i. F is concave if and only if $F(\boldsymbol{y}) \leq F(\boldsymbol{x}) + \nabla F(\boldsymbol{x}) \cdot (\boldsymbol{y} - \boldsymbol{x})$; for all $\boldsymbol{x}, \boldsymbol{y} \in K$.

ii. F is strictly concave if and only if $F(\boldsymbol{y}) < F(\boldsymbol{x}) + \nabla F(\boldsymbol{x}) \cdot (\boldsymbol{y} - \boldsymbol{x})$; for all $\boldsymbol{x}, \boldsymbol{y} \in K$.

iii. *F is strongly concave with parameter θ if and only if*

$$F(\boldsymbol{y}) \leq F(\boldsymbol{x}) + \nabla F(\boldsymbol{x}) \cdot (\boldsymbol{y} - \boldsymbol{x}) - \frac{\theta}{2}|\boldsymbol{y} - \boldsymbol{x}|^2; \qquad \text{for all } \boldsymbol{x}, \boldsymbol{y} \in K.$$

Proof. We show only the first claim (the others are similar). We first deduce that concavity plus differentiability deduces the inequality in the statement. Let $\lambda \in (0, 1)$. By concavity,

$$F(\boldsymbol{x} + \lambda(\boldsymbol{y} - \boldsymbol{x})) = F(\lambda \boldsymbol{y} + (1 - \lambda)\boldsymbol{x}) \geq \lambda F(\boldsymbol{y}) + (1 - \lambda)F(\boldsymbol{x}) = F(\boldsymbol{x}) + \lambda(F(\boldsymbol{y}) - F(\boldsymbol{x})).$$

Equivalently,

$$\frac{F(\boldsymbol{x} + \lambda(\boldsymbol{y} - \boldsymbol{x})) - F(\boldsymbol{x})}{\lambda} \geq F(\boldsymbol{y}) - F(\boldsymbol{x}).$$

Taking the limit $\lambda \to 0$, recognising the form of a derivative and using the chain rule, we conclude

$$\nabla F(\boldsymbol{x}) \cdot (\boldsymbol{y} - \boldsymbol{x}) \geq F(\boldsymbol{y}) - F(\boldsymbol{x}),$$

which gives us the statement after reordering.

Conversely, assume that for a differentiable F the inequality holds. Take $\boldsymbol{x}, \boldsymbol{y} \in K$ arbitrary, and let $\boldsymbol{z} = \boldsymbol{x} + \lambda(\boldsymbol{y} - \boldsymbol{x})$. Then we have

$$f(\boldsymbol{x}) \leq f(\boldsymbol{z}) + \nabla f(\boldsymbol{z}) \cdot (\boldsymbol{x} - \boldsymbol{z}) = f(\boldsymbol{z}) - \lambda \nabla f(\boldsymbol{z}) \cdot (\boldsymbol{y} - \boldsymbol{x}) \tag{3.3}$$

$$f(\boldsymbol{y}) \leq f(\boldsymbol{z}) + \nabla f(\boldsymbol{z}) \cdot (\boldsymbol{y} - \boldsymbol{z}) = f(\boldsymbol{z}) + (1 - \lambda)\nabla f'(\boldsymbol{z}) \cdot (\boldsymbol{y} - \boldsymbol{x}). \tag{3.4}$$

Multiplying the equations by $(1 - \lambda)$ and λ respectively and then adding them up, we get

$$(1 - \lambda)f(\boldsymbol{x}) + \lambda f(\boldsymbol{y}) \leq f(\boldsymbol{z}) = f((1 - \lambda)\boldsymbol{x} + \lambda \boldsymbol{y}),$$

so f is concave. □

Proposition 3.4. *Let $K \subseteq \mathbb{R}^d$ be a convex set. Let $F : K \to \mathbb{R}^d$ be differentiable in all K, then*

i. *F is concave if and only if $(\nabla F[\boldsymbol{y}] - \nabla F[\boldsymbol{x}]) \cdot (\boldsymbol{y} - \boldsymbol{x}) \leq 0$ for all $\boldsymbol{x}, y \in K$.*

ii. *F is strictly concave if and only if $(\nabla F[\boldsymbol{y}] - \nabla F[\boldsymbol{x}]) \cdot (\boldsymbol{y} - \boldsymbol{x}) < 0$ for all $\boldsymbol{x}, \boldsymbol{y} \in K$.*

iii. *F is strongly concave with parameter θ if and only if*

$$(\nabla F[\boldsymbol{y}] - \nabla F[\boldsymbol{x}]) \cdot (\boldsymbol{y} - x) \leq -\theta|\boldsymbol{x} - \boldsymbol{y}|^2 \quad \text{for all } \boldsymbol{x}, \boldsymbol{y} \in K.$$

Moreover, if F is twice differentiable in K, then

iv. *F is concave if and only if $(\boldsymbol{y}-\boldsymbol{x})^\top \nabla^2 F(\boldsymbol{x})(\boldsymbol{y}-\boldsymbol{x}) \leq 0$ for all $\boldsymbol{x}, \boldsymbol{y} \in K$ (equivalently $-\nabla^2 F(\boldsymbol{x})$ is positive definite).*

v. *F is strongly concave with parameter θ if and only if $(\boldsymbol{y} - \boldsymbol{x})^\top \nabla^2 F(x)(\boldsymbol{y} - \boldsymbol{x}) \leq -\theta|\boldsymbol{y} - \boldsymbol{x}|^2$ for all $\boldsymbol{x}, \boldsymbol{y} \in K$.*

Proof. Claim i follows from applying Proposition 3.3- i twice to get

$$F(\boldsymbol{y}) \leq F(\boldsymbol{x}) + \nabla F(\boldsymbol{x}) \cdot (\boldsymbol{y} - \boldsymbol{x}) \qquad F(\boldsymbol{x}) \leq F(\boldsymbol{y}) + \nabla F(\boldsymbol{y}) \cdot (\boldsymbol{x} - \boldsymbol{y}).$$

Summing both equations and reorganising deduces the claim. Claims ii and iii are deduced similarly.

To prove iv, we can use i to write, for any $\lambda \in (0, 1)$,

$$\{\nabla F(\boldsymbol{x} + \lambda(\boldsymbol{y} - \boldsymbol{x})) - \nabla F(\boldsymbol{x})\} \cdot (\lambda(\boldsymbol{y} - \boldsymbol{x})) \leq 0$$

Dividing by λ^2, passing to the limit when $\lambda \to 0$ and using the chain rule deduces the claim. □

Example 3.5. *As we mentioned before strongly concave implies strictly concave which in turn implies concavity. We show example to show the opposite is not true. Let us $K = \mathbb{R}$. Note that*

- *$F(x) := x$ is concave but not strictly concave (and* a fortiori *not strongly concave). (Use the definition).*
- *$F(x) := -x^{3/2}$ is strictly concave but not strongly concave (use Proposition 3.4 - ii and iii).*

We now connect some of these versions of concavity with properties optimisation problem in a convex setting.

Proposition 3.6. *Concerning the optimisation problem* (3.1)*, assume that K is a convex set and that the function F is concave. Then, any local solution of the optimisation problem is a global solution.*

Proof. Clearly global maxima are local maxima. Hence, if there is a unique local maximum, then it is the global maximum, and the claim follows.

Suppose then that we have more than one local maxima. We need to show that all of them have the same function value, so that they must all be global maxima.

Indeed, suppose that $\boldsymbol{x}, \boldsymbol{y}$ are both local maxima for F. But since x is a local maximum, for α sufficiently close to 1, we have

$$F(\boldsymbol{x}) \geq F(\alpha\boldsymbol{x} + (1-\alpha)\boldsymbol{y}) \geq \alpha F(\boldsymbol{x}) + (1-\alpha)F(\boldsymbol{y})$$

and by reordering

$$F(\boldsymbol{x}) \geq F(\boldsymbol{y}).$$

Similarly, for α sufficiently close to 0,

$$F(\boldsymbol{y}) \geq F(\alpha\boldsymbol{x} + (1-\alpha)\boldsymbol{y}) \geq \alpha F(\boldsymbol{x}) + (1-\alpha)F(\boldsymbol{y})$$

so that

$$F(\boldsymbol{y}) \geq F(\boldsymbol{x}),$$

and thus $F(\boldsymbol{x}) = F(\boldsymbol{y})$. □

The above proof can be easily adapted to deduce the following result.

Corollary 3.7. *Concerning the optimisation problem* (3.1)*, assume that K is a convex set and that the function F is strictly concave. Then, there is at most a unique global solution.*

Hence, for concave functions there is a strong connection between local and global properties. We can therefore solve the global optimisation problem by looking at local properties of maximisation candidates. Here is an example in the case when F is differentiable.

Proposition 3.8. *Assume that K is convex. Let $\boldsymbol{x}^*$ be a local optimal solution to problem* (3.1) *and assume that F is differentiable at x^*. Then,*

$$\nabla F(\boldsymbol{x}^*) \cdot (\boldsymbol{y} - \boldsymbol{x}^*) \leq 0 \text{ for all } \boldsymbol{y} \in K.$$

Moreover, if in addition $\boldsymbol{x}^* \in \mathrm{Int}(K)$ (that is, if there is a non-degenerate ball around $\boldsymbol{x}^*$ contained in K), then $\nabla F(\boldsymbol{x}^*) = \mathbf{0}$.

Corollary 3.9. *If K is convex, and F is differentiable and concave, then $\boldsymbol{x}^*$ is a global optimisation problem if and only if*

$$\nabla F(\boldsymbol{x}^*) \cdot (\boldsymbol{y} - \boldsymbol{x}^*) \leq 0 \text{ for all } \boldsymbol{y} \in K.$$

Moreover, assume further that K is open (for example $K = \mathbb{R}^d$). Then, the statement of the above inequality can be changed for

$$\nabla F(\boldsymbol{x}^*) = 0.$$

Proof. Consider the first claim. We just need to add the *only if* direction to the result in Proposition 3.8. Assume the inequality condition holds. Using concavity and Proposition 3.3 - i we deduce that

$$F(\boldsymbol{y}) \leq F(\boldsymbol{x}^*) + \nabla F(\boldsymbol{x}^*) \cdot (\boldsymbol{y} - \boldsymbol{x}) \leq F(\boldsymbol{x}^*); \text{for all } \boldsymbol{y} \in K,$$

so that $\boldsymbol{x}^*$ is a local maximum.

□

In many cases, we can explicitly find all of those solutions. Otherwise, gradient descent methods and root-finding techniques (like the Newton algorithm) can be used to solve the optimisation problem. We will use some of these numerical techniques in the Python exercises.

Let us end this section by returning to the existence problem in the unconstrained case. Here is a result that extends Theorem 3.1

Theorem 3.10. *Assume that F is u.s.c. (automatic when F concave), $F(\boldsymbol{x}) > -\infty$ for some $\boldsymbol{x} \in \mathbb{R}^d$ and*

$$\lim_{a \uparrow \infty} F(a\boldsymbol{x}) = -\infty, \forall \boldsymbol{x} \neq \mathbf{0}.$$

Then, there is a (global) solution to the maximisation problem.

Moreover, if F is strictly concave the above condition is also necessary.

Proof. Let us prove that the condition is sufficient. By assumption, there exists $\boldsymbol{x}_0$ such that $F(\boldsymbol{x}_0) > -\infty$. Now, because the function F tends to $-\infty$ in every direction, we can find a real $r > 0$ large enough so that

$$F(\boldsymbol{x}) < F(\boldsymbol{x}_0); \quad \forall |\boldsymbol{x}| > r. \tag{3.5}$$

Now, by Theorem 3.1 there exists a maximum $\boldsymbol{x}^*$ of F within the closed ball of radius r around zero. So that $F(\boldsymbol{x}^*) \geq F(\boldsymbol{x})$ for all $|\boldsymbol{x}| \leq r$. But by (3.5), $F(\boldsymbol{x}^*) \geq F(\boldsymbol{x}_0) > F(\boldsymbol{x})$ whenever $|\boldsymbol{x}| > r$. Hence, $F(\boldsymbol{x}^*) \geq F(\boldsymbol{x})$ for all $\boldsymbol{x} \in \mathbb{R}^n$ and it is thus optimal.

To prove that the condition is necessary if F is strictly concave, suppose by contradiction that there is a direction

□

3.2 Lagrangian optimisation method for constrained optimisation

As we saw, it is convenient to deal with an unconstrained concave optimisation problem since it essentially is equivalent to finding roots of the gradient equation. Consider now a problem of the type

$$\max_{\boldsymbol{x}\in\mathbb{R}^d} \quad F(\boldsymbol{x}) \tag{3.6}$$

$$\text{s.t.} \quad G(\boldsymbol{x}) \leq 0 \tag{3.7}$$

$$H(\boldsymbol{x}) = 0 \tag{3.8}$$

where $F : \mathbb{R}^d \to \mathbb{R}$ is concave, $G : \mathbb{R}^d \to \mathbb{R}^k$ is convex and $H : \mathbb{R}^d \to \mathbb{R}^\ell$ is linear. In this case $K = \{\boldsymbol{x} \in \mathbb{R}^d : G(\boldsymbol{x}) \leq 0, H(\boldsymbol{x}) = 0\}$, which can easily be shown to be a convex set.

This type of problem can be sometimes made simpler by turning it into an unconstrained optimisation problem via the multipliers method. We start by defining a function that extends the value function with the constraints.

Definition 3.11. The *Lagrangian* of the maximisation problem (3.6) with constraints (3.7) and (3.8) is a mapping

$$\mathcal{L} : \mathbb{R}^d \times \mathbb{R}^k_+ \times \mathbb{R}^\ell \to \mathbb{R}$$

given by

$$\mathcal{L}(\boldsymbol{x}, \boldsymbol{\gamma}, \boldsymbol{\lambda}) = F(\boldsymbol{x}) - \boldsymbol{\gamma} \cdot G(\boldsymbol{x}) + \boldsymbol{\lambda} \cdot H(\boldsymbol{x}).$$

In this context, we call $\boldsymbol{\gamma}, \boldsymbol{\lambda}$ (Lagrange) *multipliers.*

Be mindful about the sign of each entry in $\boldsymbol{\gamma}$. Before giving the main result connecting the constrained optimisation problem with the Lagrangian, we need an important concept that generalises the idea of an inflection point.

Definition 3.12. A point $(\boldsymbol{x}^*, \boldsymbol{\gamma}^*, \boldsymbol{\lambda}^*) \in \mathbb{R}^d \times \mathbb{R}^k_+ \times \mathbb{R}^\ell$ is called a saddle point of the function $\mathcal{L}$, if

$$\inf_{(\boldsymbol{\gamma},\boldsymbol{\lambda})\in\mathbb{R}^k_+\times\mathbb{R}^\ell} \mathcal{L}(\boldsymbol{x}^*, \boldsymbol{\gamma}, \boldsymbol{\lambda}) = \mathcal{L}(\boldsymbol{x}^*, \boldsymbol{\gamma}^*, \boldsymbol{\lambda}^*) = \sup_{\boldsymbol{x}\in\mathbb{R}^d} \mathcal{L}(\boldsymbol{x}, \boldsymbol{\gamma}^*, \boldsymbol{\lambda}^*).$$

The saddle point is a point of unstable equilibrium, i.e., it is minimal in one component but maximal in the remaining. We are ready to state the main result of this section.

Theorem 3.13 (Kuhn-Tucker)**.** *Assume the optimisation problem* (3.6) *with constraints* (3.7) *and* (3.8) *has a solution. Suppose in addition that there is at least one* $\boldsymbol{x} \in K$ *such that the constraints* (3.7) *are satisfied with strict inequality. Then,* $\boldsymbol{x}^*$ *is a solution of the optimisation problem if and only if there exist* $(\boldsymbol{\gamma}^*, \boldsymbol{\lambda}^*) \in \mathbb{R}^k_+ \times \mathbb{R}^\ell$ *such that* $(\boldsymbol{x}^*, \boldsymbol{\gamma}^*, \boldsymbol{\lambda}^*)$ *is a saddle point for the Lagrangian.*

Proof.

- Assume first that $(\boldsymbol{x}^*, \boldsymbol{\gamma}^*, \boldsymbol{\lambda}^*)$ is a saddle point for $\mathcal{L}$. Since

$$\begin{aligned}\mathcal{L}(\boldsymbol{x}^*, \boldsymbol{\gamma}^*, \boldsymbol{\lambda}^*) &= \inf_{(\boldsymbol{\gamma}, \boldsymbol{\lambda}) \in \mathbb{R}^k_+ \times \mathbb{R}^\ell} \mathcal{L}(\boldsymbol{x}^*, \boldsymbol{\gamma}, \boldsymbol{\lambda}) = F(\boldsymbol{x}^*) + \inf_{(\boldsymbol{\gamma}, \boldsymbol{\lambda}) \in \mathbb{R}^k_+ \times \mathbb{R}^\ell} \{-\boldsymbol{\gamma} \cdot G(\boldsymbol{x}^*) + \boldsymbol{\lambda} \cdot H(\boldsymbol{x}^*)\},\\ &= F(\boldsymbol{x}^*) - \sup_{(\boldsymbol{\gamma}, \boldsymbol{\lambda}) \in \mathbb{R}^k_+ \times \mathbb{R}^\ell} \{\boldsymbol{\gamma} \cdot G(\boldsymbol{x}^*) - \boldsymbol{\lambda} \cdot H(\boldsymbol{x}^*)\},\end{aligned}$$

it follows that $G(\boldsymbol{x}^*) \le 0$ and $H(\boldsymbol{x}^*) = 0$ as otherwise the supremum would be infinity. Hence $\boldsymbol{x}^* \in K$. Moreover, since $\gamma \cdot G(\boldsymbol{x}^*) \le 0$ for all $\gamma \in \mathbb{R}^k_+$, it must follow that $\boldsymbol{\gamma}^* \cdot G(\boldsymbol{x}^*) = 0$. We then get that $\mathcal{L}(\boldsymbol{x}^*, \boldsymbol{\gamma}^*, \boldsymbol{\lambda}^*) = F(\boldsymbol{x}^*)$.

Now, take any $\boldsymbol{y} \in K$. We get from the fact that $\boldsymbol{\gamma}^* \cdot G(\boldsymbol{y}) \le 0$ that

$$F(\boldsymbol{x}^*) = \mathcal{L}(\boldsymbol{x}^*, \boldsymbol{\gamma}^*, \boldsymbol{\lambda}^*) = \sup_{\boldsymbol{x} \in \mathbb{R}^d} \mathcal{L}(\boldsymbol{x}, \boldsymbol{\gamma}^*, \boldsymbol{\lambda}^*) \ge \mathcal{L}(\boldsymbol{y}, \boldsymbol{\gamma}^*, \boldsymbol{\lambda}^*) = F(\boldsymbol{y}) - \boldsymbol{\gamma}^* \cdot G(\boldsymbol{y}) \ge F(\boldsymbol{y}).$$

Thus $\boldsymbol{x}^*$ is maximal in K.

- Conversely, let us now assume that $\boldsymbol{x}^*$ solves the optimisation problem.

Step 1. Consider the set

$$C = \{(-F(\boldsymbol{x}) + \delta, G(\boldsymbol{x}) + \Delta, H(\boldsymbol{x})) : \delta \ge 0, \Delta \ge 0, \boldsymbol{x} \in \mathbb{R}^d\}.$$

This set can be shown to be convex, thanks to the convexity of $-F, G, H$. Clearly $(-F(\boldsymbol{x}^*), 0, 0) \in C$ (taking $\Delta = -G(\boldsymbol{x}^*)$) and moreover we can deduce that it is actually in the boundary of C: otherwise, C would contain a point $\bar{\boldsymbol{x}}$ of the form

$$(-F(\bar{\boldsymbol{x}}) + \delta, G(\bar{\boldsymbol{x}}) + \Delta, H(\bar{\boldsymbol{x}})) = (-F(\boldsymbol{x}^*) - \epsilon, 0, 0) \text{for some } \epsilon > 0,$$

but such $\bar{\boldsymbol{x}}$ would be admissible and thus this would contradict the maximality of $\boldsymbol{x}^*$ on K.

Step 2. The convexity of C implies there is a triplet $(\eta, \boldsymbol{\gamma}, \boldsymbol{\lambda}) \neq 0$ normal to C at the boundary point $(-F(\boldsymbol{x}^*), 0, 0)$, that is, such that for all $\boldsymbol{x} \in \mathbb{R}^d, \delta \in \mathbb{R}_+, \Delta \in \mathbb{R}^k_+$,

$$\eta(-F(\boldsymbol{x}) + \delta + F(\boldsymbol{x}^*)) + \boldsymbol{\gamma} \cdot (G(\boldsymbol{x}) + \Delta) + \boldsymbol{\lambda} \cdot H(\boldsymbol{x}) \ge 0 \tag{3.9}$$

We can deduce several properties of $(\eta, \boldsymbol{\gamma}, \boldsymbol{\lambda})$ by taking different choices of $\boldsymbol{x}, \delta$ and Δ in (3.9):

- ∗ Choosing $\boldsymbol{x} = \boldsymbol{x}^*, \delta = 1, \Delta = -G(\boldsymbol{x}^*)$ we deduce that $\eta \ge 0$.
- ∗ Choosing $\boldsymbol{x} = \boldsymbol{x}^*, \delta = 0, \Delta = -G(\boldsymbol{x}^*) + \boldsymbol{\alpha}$ for any $\boldsymbol{\alpha} > \mathbf{0}$ we deduce that $\boldsymbol{\gamma} \ge \mathbf{0}$.
- ∗ Choosing $\boldsymbol{x} = \boldsymbol{x}^*, \delta = 0, \Delta = 0$ we deduce from the fact that $\boldsymbol{\gamma} \ge \mathbf{0}$ and $G(\boldsymbol{x}^*) \le 0$ that $\boldsymbol{\gamma} \cdot G(\boldsymbol{x}^*) = 0$.
- ∗ From the assumption, we can choose $\bar{\boldsymbol{x}} \in K$ such that $G(\bar{\boldsymbol{x}}) < 0$. Choosing $\boldsymbol{x} = \bar{\boldsymbol{x}}, \delta = 0, \Delta = 0$, it follows that $F(\bar{\boldsymbol{x}}) < F(\boldsymbol{x}^*)$ and $\eta^* > 0$. Note that this already tells us that the optimal of F does not satisfy strictly all inequality assumptions.

We can then define $\boldsymbol{\gamma}^* = \boldsymbol{\gamma}/\eta$, $\boldsymbol{\lambda}^* = \boldsymbol{\lambda}/\eta$. It then follows from (3.9) (with $\delta = 0, \Delta = 0$) and the above properties that

$$\mathcal{L}(\boldsymbol{x}^*, \boldsymbol{\gamma}^*, \boldsymbol{\lambda}^*) = F(\boldsymbol{x}^*) = \sup_{x \in \mathbb{R}} \mathcal{L}(x, \boldsymbol{\gamma}^*, \boldsymbol{\lambda}^*).$$

Step 3. Finally, since for any $\boldsymbol{\gamma} \in \mathbb{R}_+^k$, we have that $\gamma \cdot G(\boldsymbol{x}^*) \leq 0$, it follows that

$$\mathcal{L}(\boldsymbol{x}^*, \boldsymbol{\gamma}, \boldsymbol{\lambda}) = F(\boldsymbol{x}^*) - \gamma \cdot G(\boldsymbol{x}^*) \geq F(\boldsymbol{x}^*) = \mathcal{L}(\boldsymbol{x}^*, \boldsymbol{\gamma}^*, \boldsymbol{\lambda}^*) \text{ for all } (\gamma, \lambda) \in \mathbb{R}_+^k \times \mathbb{R}^\ell,$$

so that

$$\mathcal{L}(\boldsymbol{x}^*, \boldsymbol{\gamma}^*, \boldsymbol{\lambda}^*) = \inf_{(\gamma,\lambda) \in \mathbb{R}_+^k \times \mathbb{R}^\ell} \mathcal{L}(\boldsymbol{x}^*, \gamma, \lambda)$$

□

Corollary 3.14. *Suppose that the assumptions of Theorem 3.13 hold. Assume in addition that F, G are differentiable. Then, $\boldsymbol{x}^*$ solves the maximisation problem* (3.6) *with constraints* (3.7) *and* (3.8) *if and only if there exists $\boldsymbol{\gamma}^*, \boldsymbol{\lambda}^* \in \mathbb{R}_+^k \times \mathbb{R}^\ell$ such that*

$$\begin{aligned} \nabla_x \mathcal{L}(\boldsymbol{x}^*, \boldsymbol{\gamma}^*, \boldsymbol{\lambda}^*) &= 0; \\ \nabla_\lambda \mathcal{L}(\boldsymbol{x}^*, \boldsymbol{\gamma}^*, \boldsymbol{\lambda}^*) &= 0; \\ \nabla_\gamma \mathcal{L}(\boldsymbol{x}^*, \boldsymbol{\gamma}^*, \boldsymbol{\lambda}^*) &\leq 0; \\ \boldsymbol{\gamma}^* \cdot G(\boldsymbol{x}^*) &= 0. \end{aligned}$$

Proof. The claim is deduced from Theorem (3.13) and Proposition 3.8. □

Geometric idea

To complement the discussion above, we illustrate geometrically the case with $d = 2, k = 0, \ell = 1$ (i.e. only a unique equality constraint in dimension 2), when F is concave and differentiable. In this case, we can read the condition on the gradient of the Lagrangian to be

$$\nabla \mathcal{L}(\boldsymbol{x}^*, \boldsymbol{\lambda}^*) = \nabla F(\boldsymbol{x}^*) + \boldsymbol{\lambda}^* \nabla \cdot H(\boldsymbol{x}^*) = 0$$

or equivalently,

$$\nabla F(\boldsymbol{x}^*) = -\boldsymbol{\lambda}^* \nabla H(\boldsymbol{x}^*),$$

i.e. both gradients are parallel. Hence, we deduce that F and H are tangent at the point $\boldsymbol{x}^*$. This is illustrated in Figure 3.1.

3.3 The investor's choice problem - one-period

For this section, we assume that the market model only has one period $T = 1$. An investor has initially an amount w_0 completely destined to be invested in the market. At the end of the period, the investor receives in addition to the value of her investment an *endowment process I*. This endowment models sources of income or loses not related to the market that can be both certain or uncertain (e.g. due to labour income).

This investor would like to optimise her wealth. The question that we want to address is *how should the investor compose her portfolio to achieve her goal?* This is known as the *optimal investment problem.*

Mathematically, this is formalised as follows:

Problem:

$$\max_{\boldsymbol{\theta} \in \mathbb{R}^{n+1}} \mathbb{E}[u(W_1)] \tag{3.10}$$

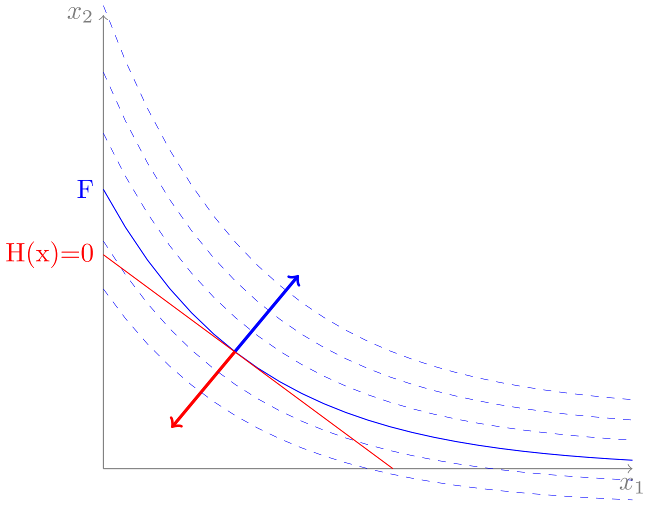

Figure 3.1: Graphical illustration: The blue dotted lines represent the level sets for a concave function F, while the red line represents the feasible set ($H(x) = 0$). The arrows point to the direction of increase for F and H respectively. The maximum of F is found at the point of tangency of the feasible set and the level sets of F.

subject to the constraints

$$\begin{aligned} t = 0; &\quad w_0 = \boldsymbol{\theta}\boldsymbol{S}_0; \\ t = 1; &\quad W_1 = I_1 + \boldsymbol{\theta}\boldsymbol{S}_1. \end{aligned} \tag{3.11}$$

We call (3.10) the *objective* or *goal* (in this case choosing the best strategy to maximise expected utility), the variable $\boldsymbol{\theta}$ is known as the control (as it is our decision variable), and (3.11) are known as the "*budget constraints*". As before, θ^i represents the number of shares invested in asset i.

Different reformulations are possible in terms of the gross returns and the rate of returns. For example, instead of (3.10) and (3.11), we could define the problem :

$$\max_{\boldsymbol{\pi} \in \mathbb{R}^{n+1}} \mathbb{E}[u(W_1)]$$

Subject to

$$\begin{aligned} 1 &= \sum_{i=0}^{n} \pi^i = \boldsymbol{\pi} \cdot \mathbf{1}; \\ W_1 &= I_1 + w_0(\boldsymbol{\pi}\boldsymbol{R}_1), \end{aligned} \tag{3.12}$$

That is, we choose different control variables (in this case π) and modify the budget constraints accordingly.

Note that we can show the equivalence of (3.11) and (3.12) through the relation

$$\pi^i = \frac{\theta^i S_0^i}{\boldsymbol{\theta}\boldsymbol{S}_0}.$$

Exercise. Check that (3.11) and (3.12) are equivalent.

Exercise. State and interpret how (3.11) would be written in terms of the rate of returns of the assets

In addition to the budget constraints which have a financial interpretation, we might need to add a technical constraint: that $W_1 \in E$, where E is the domain of u. We only consider two cases: either $E = \mathbb{R}$ (and in this case there is no technical constraint, for example for CARA utility) or $E = [a, \infty)$ for some $a \in \mathbb{R}$ (ex: $E = (0, \infty)$ for the CRRA utilities). This is translated as the additional constraint

$$W_1 \geq a.$$

We say that a portfolio $\boldsymbol{\theta}$ *is feasible* if it satisfies almost surely, all the constraints.

Note that the budget constraints and the technical constraints are all affine as functions of the control variables. This is a nice feature that is useful to understand if we can actually solve the problem.

3.3.1 Existence and uniqueness of the optimal investment problem

Let us consider the optimisation problem defined by (3.10) and (3.11). Before trying to solve the problem, we might wonder, *is there an optimal strategy?* Suppose for example that the market admits arbitrage. Without loss of generality, we can show that this arbitrage can be constructed with initial price zero. Then, this implies that there exists a strategy that produces money without any cost. This strategy could be scaled to produce unbounded gains, so that for every strategy that one considers, one can obtain a better strategy by arbitraging a bigger quantity! Therefore, we might only expect to find a relevant solution if the market admits no-arbitrage.

Now, let us suppose for a minute that the problem has a solution, in the sense that it is possible to find a strategy giving optimal mean utility. Is this solution attained in a unique way, or there exist different equivalent strategies (portfolios) giving the same mean utility?

If we can show that the function

$$f : \mathbb{R}^{n+1} \to \mathbb{R} : f(\boldsymbol{\theta}) = \mathbb{E}\left[(u(I_1 + \boldsymbol{\theta} \cdot \boldsymbol{S}_1)\right]$$

is concave, and recalling that the constraints are all affine, we can make use of the optimisation theory we presented in the previous section to find suitable assumptions.

Exercise. Show that if u is concave (resp. convex) then f defined above is also concave (resp. convex).

A consequence of concavity is that there is a unique maximal value for the function, and that the set of maximal arguments is convex. To render this set a singleton (i.e., to show that there is only one maximum), a sufficient condition is to require the function f to be strictly concave, i.e.

$$f(\alpha\boldsymbol{\theta} + (1-\alpha)\boldsymbol{\xi}) > \alpha f(\boldsymbol{\theta}) + (1-\alpha) f(\boldsymbol{\xi}).$$

In particular, this is achieved if the function u is strictly increasing, strictly concave and if the market does not have *redundancies*, that is, no instrument in the market can be replicated by the remaining instruments.

The following result, proved for example in Föllmer and Schied (2011), formalises the previous discussion:

Theorem 3.15. *Assume u is a strictly concave, and either that*

- $E = \mathbb{R}$ *and u is bounded from above; or*
- $E = (a, \infty]$, *and*
$$\mathbb{E}|u(I_1 + \boldsymbol{S}_1^{\boldsymbol{\theta}})| < \infty$$
for all feasible $\boldsymbol{\theta}$.

Then, there exists a solution to the optimal investment problem if and only if the market is arbitrage-free. Moreover, this solution is unique if and only if the market has no redundancies.

Idea of the proof. Although we omit the full proof of the result, let us summarise here the main elements in deducing Theorem 3.15:

- If there are redundancies in the market, the solution cannot be in general unique: it suffices to use any redundancies to change the composition of the portfolio and get the same expected utility as the maximal.
- If there are no redundancies, uniqueness is deduced from strict concavity of the function to optimise.
- If there is an arbitrage on the market, there is no maximum. Indeed, in this case we can create unbounded outcomes with no extra violation of initial wealth. Since F is increasing, this means the optimal would diverge.

To complete the existence argument, the two different domain cases are analysed to make use of the existence theorems we showed in the first part:

- In the case $E = [a, \infty)$
  - We show that the space of feasible portfolios is compact and the function $F(\boldsymbol{\theta}) := \mathbb{E}[u(I_1 + \boldsymbol{S}_1^{\boldsymbol{\theta}})]$ is upper semicontinuous.
- In the case $E = \mathbb{R}$,
  - We show that F is upper semicontinuous using Fatou's lemma, and that F tends to $-\infty$ at infinity using the no-arbitrage assumption.

□

In the following, we assume all the hypothesis in Theorem 3.15 and thus we guarantee the existence of a unique solution. Moreover, we assume that the function u is differentiable in E

3.3.2 First order conditions

The case $E = \mathbb{R}$

Let us first treat the case $E = \mathbb{R}$. Knowing that our portfolio choice problem has a unique solution, we use the *Lagrange multipliers* technique to transform the optimisation problem with constraints (3.11) into an unconstrained optimisation problem.

Define the Lagrangian

$$\mathcal{L}(\boldsymbol{\theta}, \lambda) := \mathbb{E}\bigg[u\Big(I_1 + \boldsymbol{\theta} \cdot \boldsymbol{S}_1\Big)\bigg] - \lambda\left(\boldsymbol{\theta} \cdot \boldsymbol{S}_0 - w_0\right) \tag{3.13}$$

where λ is the Lagrange multiplier. It is well known that any maximal (and under our assumptions the maximum) must satisfy

$$\partial_{\theta^i}\mathcal{L}(\boldsymbol{\theta}^*, \boldsymbol{\lambda}^*) = 0 \text{ for all } i = 0, \dots n; \quad \text{ and } \quad \partial_\lambda \mathcal{L}(\boldsymbol{\theta}^*, \boldsymbol{\lambda}^*) = 0. \tag{3.14}$$

Assume that $\mathbb{E}|u'(I_1 + \boldsymbol{\theta} \cdot \boldsymbol{S}_1)S_1^i| < \infty$ almost surely for all i uniformly in $\boldsymbol{\theta}$. We can then exchange the partial derivatives and expectation[1], that is to use that

$$\frac{\partial}{\partial \theta^i}\mathbb{E}[u(\cdot)] = \mathbb{E}\left[\frac{\partial}{\partial \theta^i}u(\cdot)\right]. \tag{3.15}$$

Then, (3.14) becomes

$$\mathbb{E}[u'(W(\boldsymbol{\theta}^*))S_1^i] - \boldsymbol{\lambda}^* S_0^i = 0, \text{ for all } i = 0, \dots, n \quad \text{ and } \quad \boldsymbol{\theta}^* \cdot \boldsymbol{S}_0 = w_0 \tag{3.16}$$

where $W(\boldsymbol{\theta}^*) = I_1 + \boldsymbol{\theta}^* \cdot \boldsymbol{S}_1$ is the end of period wealth.

In this case, if $\lambda \neq 0$ we can rewrite (3.16) to get

$$\mathbb{E}\left[\frac{u'(W(\boldsymbol{\theta}^*))}{\boldsymbol{\lambda}^*}S_1^i\right] = S_0^i \text{ for all } i = 0, \dots, n. \tag{3.17}$$

In this case $\frac{u'(W(\boldsymbol{\theta}^*))}{\boldsymbol{\lambda}^*}$ is an SDF. From a theoretical point of view, this result is useful to study the existence of SDFs in terms of solving an optimal investment problem.

To give a more economic interpretation, assume further that $\boldsymbol{\lambda}^* > 0$, $u' > 0$ and suppose we are in the finite probability space. Since from (1.22) any SDF is essentially a (weighted) the price of some possible outcome in the probability space, (3.18) reads

$$u'(W(\boldsymbol{\theta}^*))(\omega_j) = \boldsymbol{\lambda}^* \frac{p_j^{AD}}{\mathbb{P}\{\omega_j\}} \tag{3.18}$$

Thus, in this case, the marginal utility of an optimally investing agent at a given state is proportional to the ratio of the "price" of a state and its likeliness

If $S_0^i \neq 0$, (3.17) can also be written

$$\mathbb{E}\left[\frac{u'(W(\boldsymbol{\theta}^*))}{\boldsymbol{\lambda}^*}R_1^i\right] = 1, \text{ for all } i = 0, \dots, n;$$

which in turn implies

$$\mathbb{E}[u'(W(\boldsymbol{\theta}^*))(R_1^i - R_1^j)] = 0 \quad \forall i, j.$$

The pay-off $R_1^i - R_1^j$ is called the "*excess return of the asset i over the asset j*". Thus, the last expression means that marginal utility evaluated at the **optimal wealth** is orthogonal to each excess return. In particular, adding or subtracting a little of a zero-cost portfolio does not improve utility.

[1] The assumptions on differentiability and the bound of the expectation of the derivative allows us to apply Leibniz's rule

The case $E = [a, \infty)$

The technical constraint cannot in general be treated using Lagrange multipliers (because u is not defined in those cases). Let us start by defining

Definition 3.16. Let $f : \mathbb{R}^N \to \mathbb{R}$. The directional derivative of f in the direction $\boldsymbol{v} \in \mathbb{R}^N$ evaluated at the point $\boldsymbol{x} \in \mathbb{R}^N$ is defined by

$$Df(x; \boldsymbol{v}) := \lim_{\epsilon \downarrow 0} \frac{f(\boldsymbol{x} + \epsilon \boldsymbol{v}) - f(\boldsymbol{x})}{\epsilon}.$$

We can state a more general first order optimality condition in terms of directional derivatives.

$$D\mathcal{L}[(\boldsymbol{\theta}^*, \boldsymbol{\lambda}^*)^\top, (\boldsymbol{\theta} - \boldsymbol{\theta}^*)] \leq 0 \text{ for all } (\boldsymbol{\theta} - \boldsymbol{\theta}^*) \in E. \tag{3.19}$$

The condition (3.19) simply states that the investment cannot be improved by changing the composition of the portfolio to any other feasible portfolio. Note in particular that if the maximum is in the interior of the feasible set, conditions (3.19) becomes (3.14). For this reason, in several practical applications, it is useful to consider separately the interior and border of the feasible set.

3.3.3 Examples: single risky asset

Assumptions:

(i) risk-free asset with return R^0;

(ii) one risky asset with return R;

(iii) no end-of-period endowment, i.e., $I_1 = 0$.

Agent invests ϕ in risky asset and $w_0 - \phi$ in risk-free asset:

$$W = \phi R + (w_0 - \phi)R^0 = w_0 R^0 + \phi(R - R^0) \tag{3.20}$$

First-order condition for the optimal portfolio:

$$\mathbb{E}[u'(W)(R - R^0)] = 0 \tag{3.21}$$

Exercise. How does ϕ relate to θ in Subsection 3.3?

Investment in risky asset is positive if risk premium is positive

Let us now assume that $\mathbb{E}[R] > R^0$ and that R has a finite third moment. Then, from (3.20),

$$W = w_0 R^0 + \phi(R - R^0) = w_0 R^0 + \phi(\mathbb{E}[R] - R^0) + \phi(R - \mathbb{E}[R]) = \mathbb{E}[W] + \xi,$$

where $\xi = \phi(R - \mathbb{E}[R])$ is a random variable that satisfies $\mathbb{E}[\xi] = 0$.

We recall from (2.3) the definition of certainty premium η of W as

$$u(\mathbb{E}[W] - \eta(W)) = \mathbb{E}[u(W)].$$

Let us suppose that ϕ is very small. We can then use the approximation (2.4), which writes here as

$$\eta(W) \approx \frac{1}{2}\sigma^2\phi^2\alpha(\mathbb{E}[W]), \tag{3.22}$$

(compare with (2.4), and recall that in that case $\sigma = 1$). Note that if the utility function is concave, η will be positive as all the terms are positive.

We want to determine which of the terms $\phi(\mathbb{E}[R] - R^0)$ and η is larger. Assuming that $\mathbb{E}[R] - R^0 > 0$, we have that

$$\lim_{\phi\to 0} \frac{\phi(\mathbb{E}[R] - R^0)}{\eta} = \lim_{\phi\to 0} \left(\frac{\phi^2}{\eta}\frac{\mathbb{E}[R] - R^0}{\phi}\right) = \infty.$$

Since both terms are positive, we conclude that $\phi(\mathbb{E}[R] - R^0) > \eta$ for sufficiently small $\phi > 0$. This implies

$$\mathbb{E}[W] - \eta = w_0R^0 + \phi(\mathbb{E}[R] - R^0) - \eta > w_0R^0$$

and thus, using the increasing property of the utility function, $u(w_0R^0) < u(\mathbb{E}[W] - \eta) = \mathbb{E}[u(W)]$. Thus, some investment in the risky asset is better than none whenever the certainty premium is positive.

The case of CRRA in Binomial model

Our investor has a current wealth of w_0 and may choose to invest any part of it in the risky asset. Let

$$\phi = \text{ the amount invested in the risky asset}$$

Then $w_0 - \phi$ is the amount which the investor puts into the risk-less asset.

Let us assume, that R only takes the two values[2] $d \in (0, R^0)$ and $u \in (R^0, \infty)$ (the Binomial model), each with probability $1/2$. Thus, W either takes value $(w_0 - \phi)R^0 + \phi u$ or $(w_0 - \phi)R^0 + \phi d$, each with probability $1/2$.

As in (2.12), since the investor has CRRA ρ, we have

$$\begin{aligned} u(x) &= \frac{1}{1-\rho}x^{1-\rho}, \qquad \rho \in (0,1) \cup (1,\infty); \\ u(x) &= \log(x), \qquad \rho = 1. \end{aligned}$$

Remark 3.1. Note that all our assumptions are satisfied, so that a unique solution exists.

Let us assume in the following that $\rho \neq 1$. Then

$$\mathbb{E}[u(W(\phi))] = \frac{1}{2(1-\rho)}\left((w_0R^0 - \phi R^0 + \phi u)^{1-\rho} + (w_0R^0 - \phi R^0 + \phi d)^{1-\rho}\right) \tag{3.23}$$

To find the optimal ϕ^*, let us consider the first order condition

$$0 = \frac{\partial}{\partial\phi}\mathbb{E}[u(W(\phi^*))] = \frac{1}{2}\left((w_0R^0 - \phi^*R^0 + \phi^*u)^{-\rho}(u - R^0) + (w_0R^0 - \phi^*R^0 + \phi^*d)^{-\rho}(d - R^0)\right)$$

[2]The choice of the domain of definition for d and u is related to a no-arbitrage condition, as shown in the exercises of the chapter on asset pricing.

Rearranging we obtain

$$(w_0R^0 - \phi^*R^0 + \phi^*d)^{-\rho}(R^0 - d) = (w_0R^0 - \phi^*R^0 + \phi^*u)^{-\rho}(u - R^0)$$

and then

$$\left(\frac{w_0R^0 - \phi^*R^0 + \phi^*u}{w_0R^0 - \phi^*R^0 + \phi^*d}\right)^{\rho} = \frac{u - R^0}{R^0 - d}.$$

To give simplify the final expression, let us fix

$$K = \left(\frac{u - R^0}{R^0 - d}\right)^{\frac{1}{\rho}}$$

Then, we get

$$w_0R^0 - \phi^*R^0 + \phi^*u = K\left(w_0R^0 - \phi^*R^0 + \phi^*d\right)$$

and by rearranging,

$$\phi^* = \frac{w_0R^0(K-1)}{u - R^0 + K(R^0 - d)} \tag{3.24}$$

The optimal utility is found by replacing in (3.23) $\phi = \phi^*$.

Exercise. How does ϕ^* change as the relative risk aversion ρ increases?

3.3.4 Example: CARA with multivariate Normal returns

Assumptions:

(i) CARA investor

(ii) risk-free asset with return R^0_1;

(iii) n risky assets, jointly normal, with the following returns and notation:

$$\hat{\boldsymbol{R}} = \begin{pmatrix} R^1_1 \\ R^2_1 \\ \vdots \\ R^n_1 \end{pmatrix}; \quad \mathbb{E}[\hat{\boldsymbol{R}}_1] = \boldsymbol{\mu} = \begin{pmatrix} \mathbb{E}[R^1_1] \\ \mathbb{E}[R^2_1] \\ \vdots \\ \mathbb{E}[R^n_1] \end{pmatrix} = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_n \end{pmatrix}; \tag{3.25}$$

$$\mathrm{Cov}(R^i_1, R^j_1) = \Sigma_{ij}; \qquad i, j = 1, \ldots, n \tag{3.26}$$

with non-singular[3] covariance matrix

$$\mathrm{Cov}(\hat{\boldsymbol{R}}) = \bar{\bar{\Sigma}} = \begin{pmatrix} \Sigma_{11} & \Sigma_{21} & \ldots & \ldots & \Sigma_{1n} \\ \Sigma_{21} & \Sigma_{22} & & & \Sigma_{2n} \\ \vdots & & \ddots & & \vdots \\ \vdots & & & \Sigma_{n-1n-1} & \Sigma_{n-1n} \\ \Sigma_{n1} & \ldots & \ldots & \Sigma_{nn-1} & \Sigma_{nn} \end{pmatrix} \tag{3.27}$$

[3]In this framework, the non-singular condition is related to the non-replicability condition of the market

(iv) $\bar{\bar{\Sigma}}$ is invertible.

(v) no end-of-period endowment, i.e., $I_1 = 0$.

Amount invested in risk-free asset: ϕ^0.
Amount invested in risky assets: ϕ^i, and $\hat{\boldsymbol{\phi}} := (\phi^1, \ldots, \phi^n)^\top$.

Budget constraint:

$$w_0 = \sum_{i=0}^{n} \phi^i; \tag{3.28}$$

equivalently,

$$\phi^0 = w_0 - \mathbf{1} \cdot \hat{\boldsymbol{\phi}}. \tag{3.29}$$

End-of-period wealth, at $t = 1$:

$$W = \boldsymbol{\phi} \cdot \boldsymbol{R}_1 = \phi^0 R_1^0 + \hat{\boldsymbol{\phi}} \cdot \hat{\boldsymbol{R}}_1 = w_0 R_1^0 + \hat{\boldsymbol{\phi}} \cdot (\hat{\boldsymbol{R}}_1 - R_1^0 \mathbf{1}), \tag{3.30}$$

where the third equality comes from substituting (3.29) for ϕ^0.

Remark 3.2. Note that all our assumptions are satisfied, so that a unique solution exists.

We obtain

$$\mathbb{E}[W] = \mathbb{E}[w_0 R_1^0 + \hat{\boldsymbol{\phi}} \cdot (\hat{\boldsymbol{R}}_1 - R_1^0 \mathbf{1})] = w_0 R_1^0 + \hat{\boldsymbol{\phi}} \cdot (\boldsymbol{\mu} - R_1^0 \mathbf{1})$$

and

$$\operatorname{var}[W] = \operatorname{var}[\phi^0 R_1^0 + \hat{\boldsymbol{\phi}} \cdot \hat{\boldsymbol{R}}_1] = \operatorname{var}[\hat{\boldsymbol{\phi}} \cdot \hat{\boldsymbol{R}}_1] = \operatorname{var}\left[\sum_{i=1}^{n} \phi^i R_1^i\right]$$

Exercise. Show that $\operatorname{var}[W] = \hat{\boldsymbol{\phi}}^\top \bar{\bar{\Sigma}} \hat{\boldsymbol{\phi}}$.

Note that W is also normally distributed. By using the properties of the log-normal, we can then choose $\hat{\boldsymbol{\phi}}$ to maximise

$$\mathbb{E}[-e^{-\alpha W}] = -e^{-\alpha[w_0 R_1^0 + \hat{\boldsymbol{\phi}}^\top (\boldsymbol{\mu} - R_1^0 \mathbf{1}) - \frac{1}{2}\alpha \hat{\boldsymbol{\phi}}^\top \bar{\bar{\Sigma}} \hat{\boldsymbol{\phi}}]}. \tag{3.31}$$

Since $-\exp(-x)$ is an increasing function, we get that $\hat{\boldsymbol{\phi}}^*$ maximises (3.31) if and only if it maximises

$$\alpha \hat{\boldsymbol{\phi}}^\top (\boldsymbol{\mu} - R_1^0 \mathbf{1}) - \frac{1}{2}\alpha^2 \hat{\boldsymbol{\phi}}^\top \bar{\bar{\Sigma}} \hat{\boldsymbol{\phi}}, \tag{3.32}$$

which is a simple quadratic function on our control variable. We can apply first order conditions and set the gradient equal to zero to obtain

$$\boldsymbol{\mu} - R_1^0 \mathbf{1} = \alpha \bar{\bar{\Sigma}} \hat{\boldsymbol{\phi}}^*$$

The optimal portfolio is

$$\hat{\boldsymbol{\phi}}^* = \frac{1}{\alpha} \bar{\bar{\Sigma}}^{-1} (\boldsymbol{\mu} - R_1^0 \mathbf{1}). \tag{3.33}$$

Importantly, the optimal investment is independent of the initial wealth! As before, this answers to the fact that the utility function is CARA. We remark that this is not always the case (see for example Exercise 5.3).

If the returns were pairwise independent, then for $i = 1, \dots, n$ we would obtain

$$\phi^{\star,i} = \frac{\mu_i - R_1^0}{\alpha\sigma_i^2},$$

so that the optimal investment is to choose amounts proportional to the risk premium divided by variance.

Remark 3.3. Observe that solving the investor's maximisation problem under the assumptions of this subsection corresponds to maximising $\mathbb{E}[W] - \frac{\alpha}{2}\mathrm{var}[W]$, see (3.32).

3.4 Exercises - One period case

Exercise 3.1. Let $\alpha, \beta, p_1, p_2 \geq 0$. Consider the problem

$$\begin{aligned} &\max_{(x_1,x_2)\in\mathbb{R}^2} \{\alpha \ln(x_1) + \beta \ln(x_2)\} \\ &\text{s.t.} \quad p_1 x_1 + p_2 x_2 = c. \end{aligned}$$

Show the solution

$$p_1 x_1^* = \frac{c\alpha}{\alpha+\beta}, \qquad p_2 x_2^* = \frac{c\beta}{\alpha+\beta}.$$

Exercise 3.2. Let $\alpha, \beta, p_1, p_2 \geq 0$. Find the solution to

$$\begin{aligned} &\max_{(x_1,x_2)\in\mathbb{R}^2} \quad \{x_1^\alpha x_2^\beta\} \\ &\text{s.t.} \quad p_1 x_1 + p_2 x_2 = c. \end{aligned}$$

Exercise 3.3. Let $\alpha, \beta \geq 0$. Solve

$$\begin{aligned} &\max_{(x_1,x_2)\in\mathbb{R}^2} \quad \{\alpha x_1 + \beta x_2\} \\ &\text{s.t.} \quad x_1^2 + x_2^2 \leq c \end{aligned}$$

using Lagrange multipliers. Also construct a graphical representation.

Exercise 3.4. Show that the expression for the optimal investment portfolio given in (3.24) is also valid when $\rho = 1$, i.e. when $u(x) = \log(x)$.

Exercise 3.5. Consider the CRRA case of the previous exercise.

Assume that short positions are banned from the market (that is, that no investment can be negative). Find conditions in terms $u - R^0, d - R^0, R^0$ of such that the optimal investment in the risky asset is i) Larger, ii) Equal and iii) Smaller than in the previous exercise.

Exercise 3.6. Show that the optimal investment portfolio $\boldsymbol{\phi}^*$ for the example in Section 3.3.4 (CARA with multivariate Normal returns) when there is no risk-free asset is

$$\boldsymbol{\phi}^* = \frac{1}{\alpha}\bar{\bar{\Sigma}}^{-1}\boldsymbol{\mu} + \left(\frac{\alpha w_0 - \mathbf{1}^\top \bar{\bar{\Sigma}}^{-1}\boldsymbol{\mu}}{\alpha\, \mathbf{1}^\top \bar{\bar{\Sigma}}^{-1}\mathbf{1}}\right)\bar{\bar{\Sigma}}^{-1}\mathbf{1}$$

Exercise 3.7. Consider the optimal investment-consumption problem in one period, for an agent with initial wealth w_0, and who has time additive utility function $v(c_0, c_1) = u(c_0) + u(c_1)$, where u has constant absolute risk aversion α. Assume that in the market there are n risky assets jointly normal and one risk-free asset as in Section 3.3.4.

- Find the initial consumption c_0^* and the optimal portfolio in terms of the amount invested in each asset $(\phi^{\star\,0}, \ldots, \phi^{\star\,n})$. Compare with (3.33) and comment.
- Express the optimal portfolio in terms of the number of shares on each asset $(\theta^{\star\,0}, \ldots, \theta^{\star\,n})$

Exercise 3.8. Solve the optimal investment-consumption problem in the Binomial model (as in Section 3.3.3), for an agent with initial wealth w_0, and who has time additive utility function $v(c_0, c_1) = u(c_0) + \delta u(c_1)$, where $\delta < 1$ and u is CRRA. *Hint*: Express the initial consumption as a percentage of the initial wealth $c_0 = \kappa_0 w_0$.

Exercise 3.9. Consider the consumption-investment problem with only one riskless asset. Assume that an investor with time additive utility function $v(c_0, c_1) = u(c_0) + \delta u(c_1)$, has a random endowment I_1 at the end of the period.

- Write a maximisation problem without constraints equivalent to this problem, with the initial consumption as control variable.
- Suppose that $u'' < 0, u''' > 0$ and that $\mathbb{E}[I_1] = 0$. Show that the optimal c_0 is smaller than if we had $I_1 = 0$.

3.5 The multi-period case

Let us consider now a case when investors evaluate their strategies after several periods.

It becomes particularly interesting to analyse the multiperiod model in two cases: in the pure investor problem with additional costs (like *inventory* or *transaction* costs); and in the case where utility depends on time and on *consumption*. We consider first the latter, and we close the section with an example of application of the former.

In what follows, we set $(I_t)_{1\le t\le T}$ to be an adapted process denoting random endowment (income, salary, ...); $(C_t)_{0\le t\le T}$ an adapted process denoting consumption; $(\boldsymbol{\theta}_t)_{1\le t\le T}$ denotes a strategy; and $(W_t)_{0\le t\le T}$ denotes the adapted process of wealth.

Generalising the one-period section, we can also set

$$\boldsymbol{\pi}_t^i := \frac{\theta_t^i S_{t-1}^i}{\boldsymbol{\theta}_t \cdot \boldsymbol{S}_{t-1}}$$

to be the proportion of investment on the period $[t-1,t]$ devoted to asset i.

3.5.1 The portfolio-consumption choice problem

The portfolio choice problem is then the following: given W_0 and a time horizon T, find $(C, \boldsymbol{\theta})$ that solves

$$\max_{C;\boldsymbol{\theta}} \mathbb{E}[u(C_0, \dots, C_T)]$$

subject to:

$$W_{t+1} = I_{t+1} + \boldsymbol{\theta}_{t+1}\boldsymbol{S}_{t+1} \text{ for } 0 \le t < T; \tag{3.34}$$
$$\boldsymbol{\theta}_{t+1}\boldsymbol{S}_t = (W_t - C_t) \text{ for } 0 \le t < T; \tag{3.35}$$
$$W_T = C_T. \tag{3.36}$$

The constraint (3.34) states that wealth only comes from the endowment and the returns on investment; (3.35) states that the excess consumption on a given period is the only source for investment; and (3.36) is a closing constraint consistent with our assumption that the investment horizon is T.

As in the one-period case, this problem can be restated in terms of the proportions invested on each asset. In this case, it reads

$$\max_{C;\boldsymbol{\pi}} \mathbb{E}[u(C_0, \dots, C_T)]$$

subject to:

$$W_{t+1} = I_{t+1} + (W_t - C_t)\boldsymbol{\pi}_{t+1}\boldsymbol{R}_{t+1} \text{ for } 0 \le t < T; \tag{3.37}$$
$$\boldsymbol{\pi}_{t+1} \cdot \mathbf{1} = 1 \text{ for } 0 \le t < T; \tag{3.38}$$
$$W_T = C_T. \tag{3.39}$$

Note that:

- We can recover the one-period case by setting $T = 1$, and taking W_0 to represent the total wealth after consumption on time zero, i.e. fixing $C_0 = 0$.
- In essence, we are modelling a small investor: their actions do not have an impact on the market.

The budget constraints (3.37), (3.38) and (3.39) (or their equivalent (3.34), (3.35) and (3.36)) imply an important "conservation" lemma.

Lemma 3.17. *Let M be an SDF process, and suppose that $(C, \boldsymbol{\theta})$ denote a consumption planning and an investment strategy such that* (3.37), (3.38) *and* (3.39) *are satisfied. Then*

$$\sum_{s=0}^{T} \mathbb{E}\left[M_s C_s\right] = \sum_{s=1}^{T} \mathbb{E}\left[M_s I_s\right] + W_0. \tag{3.40}$$

In other words, the sum of all consumption is equivalent to the sum of all investment plus the initial wealth, when discounted by M.

Proof of Lemma 3.17 . From (3.34), multiplying by M_{t+1} on both sides and taking conditional expectation at time t, we get for all $0 \leq t < T$

$$\begin{aligned} \mathbb{E}_t[M_{t+1}W_{t+1}] &= \mathbb{E}_t[M_{t+1}I_{t+1}] + \mathbb{E}_t[M_{t+1}\boldsymbol{\theta}_{t+1}\boldsymbol{S}_{t+1}] \\ &= \mathbb{E}_t[M_{t+1}I_{t+1}] + M_t(\boldsymbol{\theta}_{t+1}\boldsymbol{S}_t) \\ &= \mathbb{E}_t[M_{t+1}I_{t+1}] + (W_t - C_t)M_t \end{aligned} \tag{3.41}$$

where the first equality follows since M is an SDF and the second uses (3.35). From (3.36)

$$\mathbb{E}_{T-1}[M_T C_T] = \mathbb{E}_{T-1}[M_T W_T],$$

and thus, replacing (3.41) with $t = T - 1$, we get

$$\mathbb{E}_{T-1}[M_T C_T] = \mathbb{E}_{T-1}[M_T I_T] + (W_{T-1} - C_{T-1})M_{T-1}$$

or, taking conditional expectation at time $T - 2$ and reordering

$$\mathbb{E}_{T-2}[M_T C_T] + \mathbb{E}_{T-2}[M_{T-1}C_{T-1}] = \mathbb{E}_{T-2}[M_T I_T] + \mathbb{E}_{T-2}[W_{T-1}M_{T-1}].$$

By iterating the argument (replacing (3.41), taking conditional expectation and reordering) $T-2$ times we finally get (3.40) as claimed. □

Remark 3.4. Let us emphasise that a similar lemma can be deduced using instead of an SDF risk-neutral measures discounted by the return of the money market account. In this case, we have

$$\sum_{s=0}^{T} \mathbb{E}^{\mathbb{Q}}\left[\frac{C_s}{S_s^0}\right] = \sum_{s=1}^{T} \mathbb{E}^{\mathbb{Q}}\left[\frac{I_s}{S_s^0}\right] + W_0. \tag{3.42}$$

3.5.2 Euler condition for maximality (First Order Conditions)

In this section we study optimality conditions for the multi-period consumption-investment problem under some regularity conditions. We start by looking at necessary conditions.

Theorem 3.18 (First order conditions - necessity)**.** *Assume that $(C^*, \boldsymbol{\theta}^*)$ is an optimal solution for the consumer-investor problem with $W_0 = w_0$ and T periods. Assume that $u \in C^1(\mathbb{R}^{T+1}, \mathbb{R})$ and such that ∇u is bounded. Then,*

$$\mathbb{E}_t[\partial_{c_t} u(C^*)]S_t^i = \mathbb{E}_t[\partial_{c_{t+1}} u(C^*) S_{t+1}^i] \text{ for all } 0 \le t < T; i = 1, \dots n \tag{3.43}$$

Proof. We are going to show the necessity of our claim by perturbing the optimal strategy solution and using optimality. To achieve this, let us fix $t \in \{0, \dots, T\}$. Let $\boldsymbol{Z}$ be an arbitrary random variable in $L^2(\mathcal{F}_t)$ with values in $\mathbb{R}^{n+1}$ (taken as a row vector).

We consider the perturbed investment strategy: for any $\epsilon > 0$, let $\theta_s^{t,\epsilon} = \theta_s^* + \epsilon \boldsymbol{Z} \mathbb{1}_{s=t+1}$. Note this strategy is a slightly perturbed version of the optimal, the perturbation appearing only for the period $[t, t+1]$ (in a random amount given by $\epsilon \boldsymbol{Z}$). In order to be able to do so and still satisfy the budget constraints, we need to compensate for the additional investment at time t by changing the consumption at the beginning and at the end of the period $[t, t+1]$. Indeed, from (3.35) we get that the consumption we must keep at time t is

$$\boldsymbol{\theta}_{t+1}^{t,\epsilon} \boldsymbol{S}_t = (W_t^* - C_t^{t,\epsilon}) \to C_t^{t,\epsilon} = C_t^* - \epsilon \boldsymbol{Z} \boldsymbol{S}_t$$

while from (3.34) we get

$$W_{t+1}^{t,\epsilon} = I_{t+1} + \boldsymbol{\theta}_{t+1}^{t,\epsilon} \boldsymbol{S}_{t+1} = W_{t+1}^* + \epsilon \boldsymbol{Z} \boldsymbol{S}_{t+1}.$$

Using once more (3.34) but for the period $[t+1, t+2]$, and remembering that the strategy keeps all other data constant, we get

$$\boldsymbol{\theta}_{t+2}^* \boldsymbol{S}_{t+1} = (W_t^{t,\epsilon} - C_{t+1}^{t,\epsilon}) \to C_{t+1}^{t,\epsilon} = C_{t+1}^* + \epsilon \boldsymbol{Z} \boldsymbol{S}_{t+1}$$

Now, by optimality

$$\mathbb{E}[u(C^{t,\epsilon})] \le \mathbb{E}[u(C^*)],$$

and we can re-arrange terms, divide by ϵ and take limits, to find

$$\lim_{\epsilon \downarrow 0} \frac{1}{\epsilon} \mathbb{E}[u(C^{t,\epsilon}) - u(C^*)] \le 0.$$

Using the assumption on boundedness of the derivative of u and its continuity, we can use the dominated convergence theorem to exchange the order of derivatives and limits and write

$$\mathbb{E}[\lim_{\epsilon \downarrow 0} \frac{1}{\epsilon} \{u(C^{t,\epsilon}) - u(C^*)\}] = \mathbb{E}\left[\frac{\mathrm{d}}{\mathrm{d}\epsilon} u(C^{t,\epsilon}) \bigg|_{\epsilon=0} \right] \le 0.$$

By using the chain rule and the expressions we have for $C^{t,\epsilon}$ we get

$$\mathbb{E}[\partial_{C_t} u(C^*) \boldsymbol{Z} \boldsymbol{S}_t] \le \mathbb{E}[\partial_{C_{t+1}} u(C^*) \boldsymbol{Z} \boldsymbol{S}_{t+1}].$$

Now, repeating the above analysis using this time $\tilde{\theta}^{t,\epsilon}_s := \theta^*_s - \epsilon Z \mathbb{1}_{s=t+1}$ we obtain,

$$\mathbb{E}[\partial_{C_t} u(C^*) \boldsymbol{Z}\boldsymbol{S}_t] \geq \mathbb{E}[\partial_{C_{t+1}} u(C^*) \boldsymbol{Z}\boldsymbol{S}_{t+1}].$$

Thus,

$$\mathbb{E}[\partial_{C_t} u(C^*) \boldsymbol{Z}\boldsymbol{S}_t] = \mathbb{E}[\partial_{C_{t+1}} u(C^*) \boldsymbol{Z}\boldsymbol{S}_{t+1}].$$

Since $\boldsymbol{Z}$ is arbitrary, we can choose it to be a vector with non-zero entries only in the $i - th$ position. We then conclude that for any $i = 0, \dots, n$

$$\mathbb{E}[\partial_{C_t} u(C^*) \tilde{Z} S^i_t] = \mathbb{E}[\partial_{C_{t+1}} u(C^*) \tilde{Z} S^i_{t+1}].$$

Recalling the characterisation of conditional expectation in Proposition (1.6), we deduce the claim. ☐

Remark 3.5. The assumption on boundedness of the derivatives can be relaxed. Note that it is used to exchange derivative and integral: for example, one can restrict it to hold only in a neighbourhood of the optimal. This later observation also serves to understand that we can extend the result to cases where the utility function is not defined on all $\mathbb{R}^{T+1}$: those cases are more delicate as the perturbation must keep the control within the admissible region; but, being able to limit oneself to a neighbourhood can help in these cases.

Remark 3.6. A simpler expression is found when the utility is time-additive. Here we have

$$u'(C_t) S^i_t = \mathbb{E}_t[\delta u'(C_{t+1}) S^i_{t+1}] \text{ for all } 0 \leq t < T; i = 1, \dots n \tag{3.44}$$

Note that condition proposes a natural candidate to serve as an SDF of the market.

Lemma 3.19. *Assume that* $\mathbb{E}[|\partial_{C_T} u(C^*) \boldsymbol{S}_T|] < \infty$ *and that*

$$\mathbb{E}_t[\partial_{c_t} u(C^*)] S^i_t = \mathbb{E}_t[\partial_{c_{t+1}} u(C^*) S^i_{t+1}] \textit{ for all } 0 \leq t < T; i = 1, \dots n$$

then we can construct an SDF M by

$$M_t = \frac{\mathbb{E}_t[\partial_{C_t} u(C^*)]}{\mathbb{E}[\partial_{C_0} u(C^*)]} \tag{3.45}$$

Corollary 3.20. *Under the assumptions of Theorem we can define an SDF*

We can use Lemma 3.19 together with the conservation result in Lemma (3.17), to deduce the sufficiency of the first order conditions when we add to the assumptions of Theorem a concavity assumption on the utility function.

Theorem 3.21 (Euler condition - sufficiency)**.** *Consider the consumer-investor problem with* $W_0 = w_0$ *and T periods. Assume that* $u \in C^1(\mathbb{R}^{T+1}, \mathbb{R})$ *and* ∇u *is bounded. Assume further that u is concave.*

Let $(C^*, \boldsymbol{\theta}^*)$ *be such that it satisfies all the budget constraints and*

$$\mathbb{E}_t[\partial_{c_t} u(C^*)] S^i_t = \mathbb{E}_t[\partial_{c_{t+1}} u(C^*) S^i_{t+1}] \textit{ for all } 0 \leq t < T; i = 1, \dots n$$

Then, $(C^*, \boldsymbol{\theta}^*)$ *is an optimal solution for the problem.*

Proof. From the assumptions on u and Lemma 3.19, we conclude that we can define an SDF M as in (3.45).

Now, let $(C, \boldsymbol{\pi})$ be some other consumption process that satisfies the budget constraints (3.34), (3.35) and (3.36). From concavity (see for example Proposition 3.4, we get that

$$\mathbb{E}[u(C)] - \mathbb{E}[u(C^*)] = \mathbb{E}[u(C) - u(C^*)] \leq \mathbb{E}[\nabla u(C^*) \cdot (C - C^*)] = \mathbb{E}\left[\sum_{t=0}^{T} \partial_{c_t} u(C^*)(C_t - C_t^*)\right]$$

Recalling the definition of M in (3.45), we can write

$$\frac{\mathbb{E}[u(C)] - \mathbb{E}[u(C^*)]}{\mathbb{E}[\partial_{C_0} u(C^*)]} \leq \sum_{t=0}^{T} \mathbb{E}\left[M_t(C_t - C_t^*)\right] = \sum_{t=0}^{T} \mathbb{E}[M_t C_t] - \sum_{t=0}^{T} \mathbb{E}[M_t C_t^*] = 0,$$

where we used in the last step the fact that both strategies $(C, \boldsymbol{\theta})$ and $(C^*, \boldsymbol{\theta}^*)$ and the conservation property of Lemma (3.17). Recalling that the denominator is always positive, we conclude that the expected utility of $(C^*, \boldsymbol{\theta}^*)$ is larger or equal than that of any other control satisfying the budget constraints. □

3.5.3 Solving the multi-period case

We have discussed some properties of the optimal consumption-investment problem in several periods. Let us now discuss how to solve such problems in practice.

The first method mimics our approach on the one-period problem: we can make use of the First Order Conditions in (3.5.2) to find possible candidates to a given multi-period problem. Then, if it happens that only one candidate is available, this must be necessarily the optimal. Otherwise, we can then maximise over the smaller space of candidates satisfying the FOC.

Another alternative comes from optimal control theory and is known as Dynamic Programming Principle. It is a very useful approach, particularly as a template for a computational solution of the problem.

Dynamic programming principle

The key idea behind the Dynamic Programming Principle is that instead of looking at an overall optimisation problem, we consider a family of optimisation problems with different conditions in such a way that solving each element in the family is simpler, and the solution of the whole family provides a way of retrieving the solution of the original problem.

We explain the concept of Dynamic Programming by introducing it to the consumption-investment problem. We assume that R, I, W are Markovian (see definition (1.13)). And consider the portfolio choice problem

$$J^* := \max_{C, \boldsymbol{\pi}} \; \mathbb{E}\left[\sum_{t=0}^{T} \delta^t u(C_t)\right]$$
$$\text{s.t. } W_{t+1} = I_{t+1} + (W_t - C_t)\pi_t^\top R_{t+1}; \qquad W_T = C_T; \qquad \pi_t^\top \bar{1} = 1.$$

We can then introduce the **value function** $J : \{0,\dots,T\} \times \mathbb{R} \times \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}$ defined by

$$J(T,w,r,y) := u(w); \text{ and}$$
$$J(t,w,r,y) := \max_{c\in\mathbb{R},\boldsymbol{\lambda}^\top \mathbf{1}=1} \left[u(c) + \mathbb{E}[\delta J(t+1, I_{t+1} + (w-c)(\boldsymbol{\lambda}\boldsymbol{R}_{t+1}), R_{t+1}, I_{t+1}) | R_t = r, I_t = y]\right].$$

We can interpret the function $J(t,w,r,y)$ as the *best cumulated expected utility that can be obtained given that at time t the available wealth is w, the gross returns where r and the income was y.* Indeed, $J(t,w,r,y)$ is obtained as an optimisation on the consumption at time t (denoted by the dummy variable c) and the best discounted expected cumulated utility for the next period, given we maximise our investment strategy.

By the definition above, we can find the value function successively backwards in time (i.e. going from T to 0). If at each time we can find optimal controls, they can be seen as depending on time and initial values. We can then write them as $c^*(t,w,r,y), \boldsymbol{\lambda}^*(t,w,r,y)$. These controls are said to be *in feedback form*, because we can determine them as a function of known values at time t.

In the case we consider as an example here, it will follow that

$$J^* = J(0, W_0, \mathbf{1}, 0)$$

i.e., the solution of the original problem coincides with finding the DDP solution when the actual initial conditions are given: we take the values $\boldsymbol{R}_0 = \mathbf{1}$ (initial closing condition), $I_0 = 0$ (no endowment at zero). In general, we say that a multiperiod stochastic optimisation problem satisfies a Dynamic Programming Principle if the equality given above follows. Hence, the DPP holds in our simple investment problem.

We present in the following an example that can be explicitly worked out and illustrates the way in which the Dynamic Programming method works.

Example: CRRA for i.i.d returns and finite horizon

Assumptions:

- No endowments $I_t = 0$
- R_t are i.i.d and positive
- u is CRRA with $\rho > 1$

Problem: Find optimal consumption and portfolio investment.

Solution: We use a DPP approach.

In this case

$$J(t,w) := \max_{c\in\mathbb{R},\lambda\in\mathbb{R}^d;\lambda^\top \bar{1}=1} \left[u(c) + \delta\mathbb{E}[J(t+1,(w-c)(\lambda^\top R_{t+1}))]\right] \tag{3.46}$$

Note that because the processes are i.i.d. there is no need to track the state value of the returns process. We show by induction:

- The optimal consumption is $C_t = \xi_t W_t$ with

$$\frac{1}{\xi_t} = \sum_{s=0}^{T-t} (\delta B^{1-\rho})^{\frac{s}{\rho}} \tag{3.47}$$

for B the certainty equivalent of the best investment.

- The optimal portfolio is the same as in the one-period case.

- The optimal utility satisfies for all $w \geq 0$,

$$J(t,w) = \frac{w^{1-\rho}\xi_t^{-\rho}}{1-\rho} \tag{3.48}$$

Indeed, by induction on t, we have:

- If $t = T$, the boundary condition $W_T = C_T$ implies that $\xi_T = 1$ which is consistent with (3.47). Moreover,

$$J(T,w) = u(C_t) = u(W_T) = \frac{(W_T)^{1-\rho}}{1-\rho},$$

which is consistent with (3.48).

- Assuming the claims hold for $t+1$, let us prove them for t: by replacing (3.48) in (3.46) we get:

$$\begin{aligned}
J(t,w) &= \max_{c\in\mathbb{R},\lambda\in\mathbb{R}^d;\lambda^\top\bar{1}=1} \left[u(c) + \delta\mathbb{E}\left[\frac{\{(w-c)\lambda^\top R_{t+1}\}^{1-\rho}\xi_t^{-\rho}}{1-\rho} \right]\right] \\
&= \max_{c\in\mathbb{R},\lambda\in\mathbb{R}^d;\lambda^\top\bar{1}=1} \left[\frac{c^{1-\rho}}{1-\rho} + \delta(w-c)^{1-\rho}\xi_{t+1}^{-\rho}\mathbb{E}\left[\frac{\{\lambda^\top R_{t+1}\}^{1-\rho}}{1-\rho} \right]\right]
\end{aligned}$$

Using the positivity of $(w-c)$ and ξ and the definition of B, we get

$$\begin{aligned}
J(t,w) &= \max_{c\in\mathbb{R}} \left[\frac{c^{1-\rho}}{1-\rho} + \delta(w-c)^{1-\rho}\xi_{t+1}^{-\rho} \max_{\lambda\in\mathbb{R}^d;\lambda^\top\bar{1}=1} \mathbb{E}\left[\frac{\{\lambda^\top R_{t+1}\}^{1-\rho}}{1-\rho} \right]\right] \\
&= \max_{c\in\mathbb{R}} \left[\frac{c^{1-\rho}}{1-\rho} + \delta(w-c)^{1-\rho}\xi_{t+1}^{-\rho}\frac{B^{1-\rho}}{1-\rho} \right] \\
&= \max_{\xi\in\mathbb{R}} \left[\frac{(\xi w)^{1-\rho}}{1-\rho} + \delta\{(1-\xi)w\}^{1-\rho}\xi_{t+1}^{-\rho}\frac{B^{1-\rho}}{1-\rho} \right]
\end{aligned}$$

where the last line is obtained by making the change of variables $c = \xi w$. By applying the first order condition and calling ξ_t the optimal value, we get

$$\begin{aligned}
&\xi_t^{-\rho} = \delta(1-\xi_t)^{-\rho}\xi_{t+1}^{-\rho}B^{1-\rho} && \Rightarrow \quad \xi_t^{-1} = \xi_{t+1}^{-1}(1-\xi_t)^{-1}\gamma \text{ where } \gamma = \delta^{\frac{1}{\rho}}B^{\frac{1-\rho}{\rho}} \\
\Rightarrow \quad &(1-\xi_t) = \xi_t\xi_{t+1}^{-1}\gamma && \Rightarrow \quad 1 = \xi_t(1+\xi_{t+1}^{-1}\gamma)
\end{aligned}$$

which together with the induction assumption implies (3.47). Finally, replacing the first equality in the value function, we get that

$$\begin{aligned}
J(t,w) =& \frac{w^{1-\rho}}{1-\rho}\left(\xi_t^{1-\rho} + \delta(1-\xi_t)^{1-\rho}\xi_{t+1}^{-\rho}B^{1-\rho}\right) \\
=& \frac{w^{1-\rho}}{1-\rho}\left(\xi_t^{1-\rho} + (1-\xi_t)\xi_t^{-\rho}\right) \\
=& \frac{w^{1-\rho}}{1-\rho}\xi_t^{-\rho}
\end{aligned}$$

as claimed. □

Remark 3.7. We have argued before that under our assumptions on utility functions, it is impossible to find an optimal investment if an arbitrage is available in the market: we argued that in such a case, it is always possible to improve the wealth by applying the arbitrage, and this implies an increase of utility due to the strict monotonicity. Thus, if one is able to optimise a utility, there must be no-arbitrage on the market. This type of relationship is used to great effect by several authors (see for example **?**) to obtain a proof of the first theorem of asset pricing beyond the simple finite probability setting we studied in Chapter 1.

3.6 Other multi-period problems in mathematical finance

3.6.1 The pricing and hedging problem

As we have discussed in Chapter 1, in complete markets it is possible to find the arbitrage-free price of a contingent payoff X to be paid at time T by means of an SDF.

Indeed, we have from Exercise that the price of such a contingent claim would be

$$p^X = \mathbb{E}[M_T X].$$

Another approach is the 'minimal super hedging price', that it is, finding the minimal price that allows to form a super-hedging self-financing strategy. This can be written in terms of optimisation by

$$\begin{aligned}
\min_{\boldsymbol{\theta}} \quad & w_0 \\
\text{s.t.} \quad & S_0^{\boldsymbol{\theta}} = w_0 \\
& S_T^{\boldsymbol{\theta}} \geq X \\
& \boldsymbol{\theta}_t \boldsymbol{S}_t = \boldsymbol{\theta}_{t+1}\boldsymbol{S}_t \text{ for all } t = 1, \ldots, T-1
\end{aligned}$$

where w_0 is the wealth available to produce a hedge: the result of such a problem would produce the minimal wealth allowing for a super-hedge, which is taken as the minimal super-hedging price.

In the complete case, by observing that every claim can be reproduced and that any excess is costly, this problem can be reduced to

$$\begin{aligned}\min_{\boldsymbol{\theta}} \quad & w_0 \\ \text{s.t.} \quad & S_0^{\boldsymbol{\theta}} = w_0 \\ & S_T^{\boldsymbol{\theta}} = X \\ & \boldsymbol{\theta}_t \boldsymbol{S}_t = \boldsymbol{\theta}_{t+1} \boldsymbol{S}_t \text{ for all } t = 1, \dots, T-1\end{aligned}$$

Note that the constraints are connected to the constraints of the optimal consumption-investment problem if we interpret $W_T := X$, $I_t = 0$ for all $t = 1, \dots, T$, and $C_t = 0$ for all $t = 0, \dots, T-1$. If we have an SDF M on the market, we can apply Lemma 3.17. We then obtain that the minimal wealth needed is

$$w_0 = \mathbb{E}[M_T X],$$

which coincides precisely with the no-arbitrage price above. One advantage of this method is that it can be used to price claims that are not only European (i.e. that pay only at a fixed time T), but also those with intermediate payments (say $(X_0, \dots, X_T)$). Clearly the expression is then

$$w_0 = \sum_{t=0}^{T} \mathbb{E}[M_t X_t],$$

which generalises the above claim.

Let us return for now to the European case for simplicity. Once the seller has found the price, it needs to know how to hedge the potential risk on the market by trading: hence, we need to find a self-financing strategy $\boldsymbol{\theta}$ that, starting from $S_0^{\boldsymbol{\theta}}$ has terminal wealth X. To find it we can solve

$$\begin{aligned}\min_{\boldsymbol{\theta}} \quad & \mathbb{E}[(S_T^{\boldsymbol{\theta}} - X)^2] \\ \text{s.t.} \quad & S_0^{\boldsymbol{\theta}} = \mathbb{E}[M_T X] \\ & \boldsymbol{\theta}_t \boldsymbol{S}_t = \boldsymbol{\theta}_{t+1} \boldsymbol{S}_t \text{ for all } t = 1, \dots, T-1\end{aligned}$$

which can be seen as an instance of the consumption-investment problem with 'utility' [4] given by $u(C) = -(C_t - X)^2$. Tools like the DPP are then readily available.

3.6.2 Pure-investment problem with transaction costs(*)

One of the main assumptions we have worked here up to this point is that trading is costless in this market. An interesting problems arises when transaction costs are added to the problem.

In this section, we consider briefly the problem of a pure-investors (i.e. that does not devote any wealth to consumption). To simplify we consider the problem with no extra-income. We assume that there are transaction costs whenever the risky assets are traded, that they are proportional to the changes in inventory of each stock and homogeneous among assets, and that they are paid at the beginning of the period. We encode this mathematically as follows: we define a process $(L_t)_{0 \le t < T}$ denoting the transaction costs paid at each time, that satisfies

$$L_t = \alpha \sum_{i=1}^{n} |(\theta_t^i - \theta_{t+1}^i) S_t^i| = \alpha |W_t| \sum_{i=1}^{n} |\pi_t^i - \pi_{t+1}^i|$$

[4]Note this is not a true utility as it is not strictly increasing. However, the quadratic case is very tractable mathematically

for some $\alpha > 0$.

The goal is to choose the optimal investment strategy $\boldsymbol{\theta}$ such that the final expected utility is maximised and full-reinvestment budget constraints are satisfied including payment of losses incurred as transaction costs. This reads

$$\begin{aligned}
&\max_{\boldsymbol{\pi}} \ \mathbb{E}\left[u(W_T)\right] \\
&\text{s.t.:} \\
&\qquad W_{t+1} = (W_t - L_t)\boldsymbol{\pi}_{t+1}\boldsymbol{R}_{t+1}; \\
&\qquad L_t = \alpha|W_t| \sum_{i=1}^{n} |\pi_t^i - \pi_{t+1}^i| \\
&\qquad \boldsymbol{\pi}_t \mathbf{1} = 1.
\end{aligned}$$

Interestingly, this problem can be understood as a consumption investment problem where the utility function only depends on final wealth, and where there is an additional constraint stating the consumption is equal to the transaction costs.

Considering a dynamic programming approach to this problem, one major difference between this framework and the consumption-investment problem is that the transaction cost needs as in put the control on the previous step: we would introduce the value function $J : \{0, \ldots, T\} \times \mathbb{R}^n \times \mathbb{R}^{n+1} \to \mathbb{R}$ such that

- $J(T, w, \lambda) = u(w)$
- $J(t, w, \lambda) = \max_{\bar{\lambda} \in \mathbb{R}^{n+1}} \mathbb{E}_t[J(t+1, w\{1 - \alpha \operatorname{sign}(w) \sum_{i=1}^n |\bar{\lambda}^i - \lambda^i|\}\bar{\lambda}\boldsymbol{R}_{t+1}, \bar{\lambda})]$.

The equation is harder to solve that the traditional investment problem. The interested reader can find a detailed analysis of this problem in continuous time in Davis and Norman (1990).

3.7 Exercises - Multiperiod case

Exercise 3.10. Assuming that there exists a martingale measure $\mathbb{Q}$ show (3.42).

Exercise 3.11. Show Lemma 3.19.

Exercise 3.12. Let us assume that the stochastic process of gross returns of market assets, $R = (R_1, R_2, \ldots)$ is a process in $\mathbb{R}^d$ with i.i.d. marginals (that is, each R_i is distributed identically and is independent of the others). For any portfolio $\pi \in \mathbb{R}^{n+1}$, let us denote by $S_t(\pi)$ the value of the dividend-reinvested portfolio with weights $\pi_t = \pi$ for all $t \geq 0$.

Assume that

$$\max\left(\mathbb{E}[\log(\pi^\top R_1)]\right) > -\infty,$$

and let $\pi^* \in \mathbb{R}^{n+1}$ be the portfolio achieving the maximum.

Show that for any portfolio $\pi \in \mathbb{R}^{n+1}$ with $-\infty < \mathbb{E}[\log(\pi^\top R)] < \mathbb{E}[\log(\pi^{*\top} R_1)]$, there exists T large enough such that with probability one $S_t(\pi) < S_t(\pi^*)$ for all $t > T$.

Hint: Calculate $\frac{1}{t}\log(S_t(\pi))$ and use the law of large numbers.

Remark 3.8. The previous exercise is connected to the concepts of Kelly portfolio and the Benchmark approach to pricing (see for example Platen (2011))

Exercise 3.13. Let us consider the portfolio choice problem in discrete time finite horizon T, assuming no endowments and, as in the preceding exercise, that the returns are i.i.d. Assume the investor has a time additive utility function with discount factor δ and marginal CRRA utility function with $\rho = 1$.

Using the dynamic programming principle and induction show that:

- The optimal consumption at time t is given by $\xi_t W_t$, where

$$\xi_t = \frac{1-\delta}{1-\delta^{T+1-t}}.$$

- The optimal investment portfolio is the solution of the one-period investing portfolio for a CRRA utility function with $\rho = 1$.
- The value function satisfies

$$J(t, w) = \frac{1}{\xi_t} \log(w) + J(t, 1),$$

where

$$J(t, 1) = \sum_{s=t+1}^{T} \delta^{s-t-1} \left(\frac{\delta \log \left[B(1-\xi_{s-1}) \right]}{\xi_s} + \log(\xi_{s-1}) \right)$$

and B is the certainty equivalent of the optimal return for the one-period problem with initial wealth 1.

3.8 Summary

- Optimisation theory includes a set of results on existence, uniqueness and characterisation of solutions to maximisation (or minimisation) problems. It plays an essential role in financial mathematics.
- In concave maximisation problems on convex domains all local optima are global; strictly concave problems have a unique solution; if in addition the function is differentiable, optima can be found from First Order Constraints. In addition, optimal solutions in certain constrained problems can be associated with finding saddle points of a modified objective function (the Lagrangian).
- The optimal investor problem aims at finding the best strategy that an investor can implement in the market in order to maximise their expected utility, given an initial investment amount and a set of future endowments.
- The optimal consumption-investor problem extends this idea to investors that also consume part of their wealth. The budget constraints implied by the problem induce a conservation property.

- Solutions to the above problems are linked with SDFs via differentiation of optimal solutions (under technical conditions)
- Other problems that can be solved using optimisation techniques include the pricing and hedging problem of derivatives and problems of investment with transaction costs.
- Problems in multiple periods can be tackled in many cases using the Dynamic Programming Principle.

Chapter 4

Risk measures

> In a world that is changing really quickly, the only strategy that is guaranteed to fail is not taking risks.
>
> *Mark Zuckenberg*

We have seen that investors are averse to risk in the sense that they might be willing to pay a premium in return for further certainty in their returns. This is also true in a general sense for companies. However, every productive endeavour is subject to risks, so that it is practically impossible to avoid all of them.

For this reason, there are incentives for individuals and companies to understand and manage their (financial) risks. In the case of financial companies this is reinforced by the existence of regulatory requirements for risk management.

Risk management is a system comprising a set of policies, organisational structure, quantitative models and indicators aimed at understanding the risk sources and exposures of a company, deciding and monitoring when they are within what is acceptable, and taking action in cases when they are not. A key element in modern risk management is its quantitative nature. It is based in *measuring* the risk implied in a given investment or in the whole business of a company, so that this can inform actions to reduce it if needed.

The study of a full risk management system and regulation is outside the scope of this lecture. We will mainly focus on how to measure risks. However, the fact that we want to use risk measures to act on them, informs our construction. It is therefore relevant to mention some tools available to companies and investors to reduce financial risks:

- Capital (or another type of safe collateral) can be added to the position. Capital is understood as a direct injection of monies, for example, from the owners of a financial company. Moreover, regulators impose capital requirements and restrictions as one of their main tools to promote financial stability.
- Hedging, i.e., implementing strategies on the market to counterbalance negative results on a position.
- Diversifying, meaning increasing the heterogeneity of assets or instruments on which there

is investment. This is a strategy based on our understanding that there is less variability of results in large numbers.

4.1 Risk measures and their properties

As before, we work on the probability space $(\Omega, \mathbb{P}, \mathcal{F})$. We denote by $\mathcal{X}$ a set of random variables in this space representing all possible financial positions at the end of the period. For example, we might consider the discounted net gains (as defined in (1.7)) from performing any market strategy stating from a fixed amount. We assume that $\mathcal{X}$ is a convex set (a straightforward generalization of Definition), and sometimes we assume it is in addition a cone (see Definition 4.1 below). Moreover, we assume that any deterministic value belongs to the cone.

Definition 4.1. $\mathcal{X}$ a subset of a vector space is convex if $X_1, X_2 \in \mathcal{X} \Rightarrow \lambda X_1 + (1-\lambda)X_2 \in \mathcal{X}$, and it is a cone if $X_1 \in \mathcal{X} \Rightarrow \lambda X_1 \in \mathcal{X}$ for all $\lambda > 0$.

Example 4.2.

- *Consider back the example of $\mathcal{X}$ the unit ball in $\mathbb{R}^d$. $\mathcal{X} = \{x \in \mathbb{R}^d : |x| < 1\}$. This is a convex set, but not a cone.*
- *Take the set $\mathcal{X} = \{x \in \mathbb{R}^d : x^i > 0\}$, the vectors with positive entries. This is a convex cone.*

Most typically, in what follows, we will take $\mathcal{X} = L^p(\Omega, \mathbb{P}, \mathbb{R}) = \{X : \Omega \to \mathbb{R} : \mathbb{E}[|X|^p] < \infty\}$, for $p \in [1, \infty)$); or $\mathcal{X} = L^\infty(\Omega, \mathbb{P}, \mathbb{R}) = \{X : \Omega \to \mathbb{R} : \exists C > 0, \mathbb{P}[|X| < C] = 1\}$. These examples are all vector spaces and so, in particular, they are convex cones.

Definition 4.3. A (unidimensional) **risk measure** is a function $\rho : \mathcal{X} \to \mathbb{R}$ that assigns to a given random variable X representing a financial position a real number, representing its "riskiness".

(!) Note that we use the same letter ρ to denote coefficient of risk aversion and a generic risk measure. The meaning of the symbol should be clear from the context.

Mathematically, a risk measure is just a function: classical examples would include the expectation, standard deviation or worst case (when dealing with finite distributions). As we want to convey a precise financial interpretation we will identify properties that express what we would understand as reasonable risk measures.

To start, it would be intuitively clear that if $X_1 > X_2$ surely (i.e. under every possible outcome) then X_2 represents a riskier financial position.

Property 4.4 (Monotonicity)**.** *For all $X_1, X_2 \in \mathcal{X}$ such that $X_1 \leq X_2$ almost surely we have $\rho(X_1) \geq \rho(X_2)$.*

Another relevant property is related to our desire to render risk measures useful to determine regulatory capital. Imagine we set the required capital to be equal to the measured risk. We would expect that this capital should be sufficient to offset the perceived risk.

Property 4.5 (Cash invariance or translation invariance)**.** *For all $X \in \mathcal{X}$ and for every $a \in \mathbb{R}$, we have $\rho(X + a) = \rho(X) - a$.*

The monotonicity and cash invariance properties are the minimum properties needed for a risk measure to be used for capital determination.

Definition 4.6. A risk measure that satisfies both is called *monetary risk measure.*

Note that any translation of a monetary risk measure is still monetary. For this reason, unless otherwise state, we add the following property for convenience:

Property 4.7 (Normalisation)**.**

$$\rho(0) = 0.$$

If we want to encode also the fact that diversification reduces risks, we need to add some further properties.

Property 4.8 (Convexity)**.** *For all $X_1, X_2 \in \mathcal{X}$ and $\lambda \in [0,1]$ then*

$$\rho[\lambda X_1 + (1-\lambda)X_2] \leq \lambda\rho(X_1) + (1-\lambda)\rho(X_2).$$

Convexity simply states that considering a convex combination or a 'portfolio' of two financial positions should be less risky than simply combining the measured risks.

Definition 4.9. A monetary risk measure with the convexity property is called a *convex* risk measure.

Let us now assume that $\mathcal{X}$ is also a cone. One can postulate in this case that risks can also be assumed to scale linearly as they grow, i.e.

Property 4.10 (Positive homogeneity)**.** *For all $X \in \mathcal{X}$ and every $\lambda > 0$ we have $\rho(\lambda X) = \lambda\rho(X)$.*

Definition 4.11. A convex measure that is also positive homogeneous is called *coherent risk measure.*

Both Positive homogeneity and Convexity together deduce

Property 4.12 (Subadditivity)**.** *For all $X_1, X_2 \in \mathcal{X}$, we have $\rho(X_1 + X_2) \leq \rho(X_1) + \rho(X_2)$.*

In fact, as pointed out in Föllmer and Schied (2011), we have

Proposition 4.13. *Considering the convexity, positive homogeneity and subadditivity properties, any two of them imply the remaining.*

Historically, a first formalisation of the intuitive idea of a what a good risk measure for financial risk management was achieved by Artzner, Delbaen, Eber and Heath in their seminal paper Artzner et al. (1999), where they introduced coherent risk measures (defined as satisfying monotonicity, cash invariance, positive homogeneity and sub-additivity). Then, it was pointed out that the positive homogeneity property might not appropriate in certain contexts, for example when liquidity effects want to be considered. In those cases, one would like for $\lambda > 1$ that

$$\rho(\lambda X) > \lambda\rho(X)$$

This led Föllmer and Schied and Fritelli and Rosazza-Gianin to propose simultaneously the formalisation of convex risk measures.

Finally, let us say that we call a risk measure *law invariant* if

$$X_1 \sim X_2 \Rightarrow \rho(X_1) = \rho(X_2).$$

Let us illustrate the definition and properties of risk measures by reviewing the most classical example of risk measure: standard deviation. Recall that it is defined by

$$\mathrm{sd(X)} = (\mathrm{var(X)})^{1/2} = (\mathbb{E}[(\mathrm{X} - \mathbb{E}[\mathrm{X}])^2]^{1/2} = (\mathbb{E}[\mathrm{X}^2] - \mathbb{E}[\mathrm{X}]^2)^{1/2}$$

This is one of the first measures to be used in practice due to its simplicity. It is a good measure of dispersion, but as a risk measure, it is quite limited: it is blind to the sign, so it does not distinguish between large losses and large gains. Moreover, it is not defined for all random variables, only for $X \in L^2(\Omega, \mathbb{P})$.

Standard deviation can be shown to satisfy homogeneity and subadditivity (thanks to Minkowski's inequality). Hence, it is also convex. However, it is not monotonic or cash invariant (thus not monetary) in general. It is thus not an ideal measure to determine capital requirements.

Let us now look at more general families of risk measures.

4.2 Utility-based risk measures

We can also use the investor preference point of view to define a risk measure.

4.2.1 Simple loss

A first idea to use utility functions applied over losses to determine their risk. More specifically,

$$\rho_u^{sl}(X) := \mathbb{E}[-u(X)]$$

Thanks to the monotonicity of u, ρ_u^{sl} is monotonous and if u is concave, ρ_u^{sl} is risk measure satisfying convexity. However, it is not in general a monetary risk measure, as cash invariance only holds if u is an affine modification of the identity.

4.2.2 Certainty equivalent

A modification of simple losses uses instead their certainty equivalents. Assume that u is continuous strictly increasing and concave. Then, we can define

$$\rho_u^{ce}(X) = -u^{-1}(\mathbb{E}[u(X)])$$

Again, we can easily verify the monotonicity property in the sense of risk measures. The convexity property must be checked on a case by case basis. Moreover, we have the following proposition

Proposition 4.14. *ρ_u^{ce} is cash invariant if and only if u has constant risk aversion.*

Exercise. Prove Proposition (4.14)

Here is an example of a convex risk measure:

Example 4.15 (Entropic risk measure)**.** *For some $\theta > 0$, the* entropic risk measure *with parameter θ (that we denote $(\rho_\theta^{\exp})$) is defined by*

$$\rho_\theta^{\exp}(X) := \frac{1}{\theta}\log\left(\mathbb{E}[e^{-\theta X}]\right)$$

Note that the entropic risk measure is in general **not** coherent, since for $\lambda > 1$

$$\rho_\theta^{\exp}(\lambda X) = \frac{1}{\theta}\log\mathbb{E}[e^{-\theta\lambda X}] \geq \frac{1}{\theta}\log\mathbb{E}[e^{-X\theta}]^\lambda = \lambda\rho_\theta^{\exp}(X)$$

with a strict inequality for certain distributions (take for example a standard Gaussian and $\lambda = 2$).

Convexity can be shown using Hölder inequality and the properties of exponentials and logarithms:

$$\begin{aligned}
\alpha\rho_\theta^{\exp}(X) + (1-\alpha)\rho_\theta^{\exp}(Y) &= \frac{1}{\theta}\log\left(\mathbb{E}[e^{-\theta X}]^\alpha\mathbb{E}[e^{-\theta Y}]^{(1-\alpha)}\right)\\
&= \frac{1}{\theta}\log\left(\mathbb{E}[(e^{-\alpha\theta X})^{\frac{1}{\alpha}}]^\alpha\mathbb{E}[(e^{-(1-\alpha)\theta Y})^{\frac{1}{1-\alpha}}]^{(1-\alpha)}\right)\\
&\geq \frac{1}{\theta}\log\left(\mathbb{E}[e^{-\alpha\theta X}e^{-(1-\alpha)\theta Y}]\right)\\
&= \rho_\theta^{\exp}(\alpha X + (1-\alpha)Y).
\end{aligned}$$

4.2.3 Shortfall risk

As we pointed out in Chapter 2, the certainty premium can be seen as a measure of risk. We generalise this idea. Let u be a concave utility function. Let $\varphi \in \text{Range}(u)$ denote some fixed utility benchmark. We then define

$$\rho_u^{SR}(X) := \inf\{z : \mathbb{E}[u(X+z)] \geq \varphi\}.$$

The shortfall risk associated to the utility u is then the minimal deterministic amount that is missing in order to obtain a pre-determined utility.

The shortfall risk here defined is a convex risk measure. It is not in general coherent.

Exercise. Show that the shortfall risk is a convex risk measure.

4.3 Tail-based risk measures

All the examples of utility functions we have presented up to now either miss the asymmetry between losses and gains, or average out positive and negative results.

To avoid any of these conditions, we can consider measures that focus on the lower tail of the distribution of X, where the worst results lie. One way of doing so is by means of a *quantile.*

Definition 4.16. The quantile function associated to a random variable X, $q_X : (0,1) \to \mathbb{R}$ given by

$$q_X(\lambda) = \inf\{x : \lambda \leq F_X(x)\} = \inf\{x : \lambda \leq P(X \leq x)\} \tag{4.1}$$

where F_X is the cumulative distribution function. It represents the minimal value whose cumulative distribution function is at least λ.

Remark 4.1. Note that two random variables that have the same distribution share also the same quantile function. We say that the quantile function is *law-invariant*.

The quantile function inherits some properties: because the cumulative distribution function is continuous to the right and has a limit to the left[1], we can show that the quantile function is continuous to the left with limits to the right. In particular, if the CDF is continuous, so it is the quantile function.

A simpler way to characterise (4.1) is as follows:

Proposition 4.17. *The quantile function is the only left continuous with right limits function defined from* $(0,1)$ *to* $\mathbb{R}$ *such that*

$$q_X(\lambda) \leq x \Leftrightarrow \lambda \leq F_X(x) \text{ for all } \lambda \in (0,1). \tag{4.2}$$

The proof is immediate from the definition of "infimum" as the minimal value of the set of upper bounds.

In the case where the cumulative distribution function has an inverse in a given domain, we have that from (4.2) that $q_X = F_X^{-1}$. We can then consider the quantile function as a 'generalised left continuous inverse' of the cumulative distribution function F_X.

Let us illustrate with a couple of examples:

Example 4.18. *Let* X *represent the outcome of one fair dice throw. We have that*

$$F_X(x) = \begin{cases} 0 & \text{if } x < 1 \\ \frac{\lfloor x \rfloor}{6} & \text{if } 1 \leq x \leq 6 \\ 1 & \text{if } x > 6 \end{cases},$$

where the floor *function* $\lfloor x \rfloor$ *is the largest integer less or equal than* x*. Then, we can find that the quantile function is given by*

$$q_X(\lambda) = \lceil 6\lambda \rceil,$$

where the ceiling *function* $\lceil 6\lambda \rceil$ *is the smallest integer greater or equal than* 6λ*. Indeed, note that for any* $\lambda \in (0,1)$,

$$\lceil 6\lambda \rceil \leq x \Leftrightarrow 6\lambda \leq \lfloor x \rfloor \Leftrightarrow \lambda \leq \frac{\lfloor x \rfloor}{6}$$

so that (4.2) *is verified. This is illustrated graphically in Figure 4.1*

Example 4.19. *Let* $X \sim U[-3,3]$, *uniformly distributed. Then we have,*

$$F_X(x) = \frac{1}{6}(x+3)\mathbb{1}_{\{-3<x\leq 3\}} + \mathbb{1}_{\{x>3\}}$$

and we find for $\lambda \in (0,1)$

$$q_X(\lambda) = 6\lambda - 3,$$

as we verify by noticing that

$$6\lambda - 3 \leq x \Leftrightarrow \lambda \leq \frac{1}{6}(x+3) \text{ for all } \lambda \in (0,1).$$

In fact, in this case we have that for $\lambda \in (0,1)$ *and* $x \in (-3,3)$ $F_X \circ q_X(\lambda) = \lambda$ *and* $q_X \circ F_X(x) = x$*. As before, we illustrate this graphically in Figure 4.2*

[1]Functions satisfying this property are usually known as *càdlàg*, a French acronym for *continue à droite, limite à gauche*

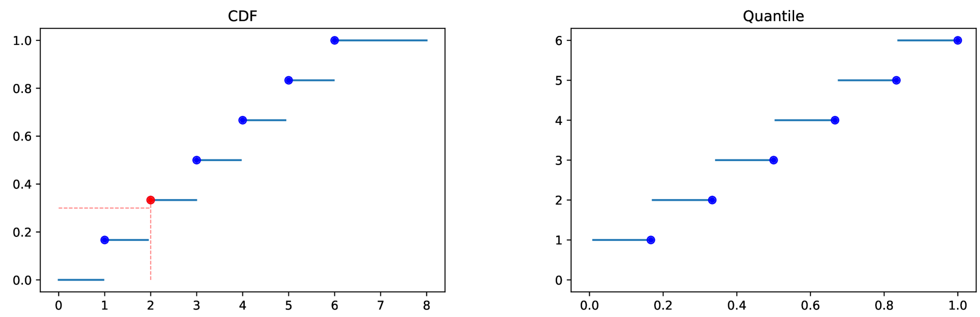

Figure 4.1: CDF (left) and quantile function (right) for the case of the distribution of a fair dice. The quantile for 0.3 (i.e., 2) is obtained from the CDF plot: it is the smallest x having $F_X(x) \geq 0.3$.

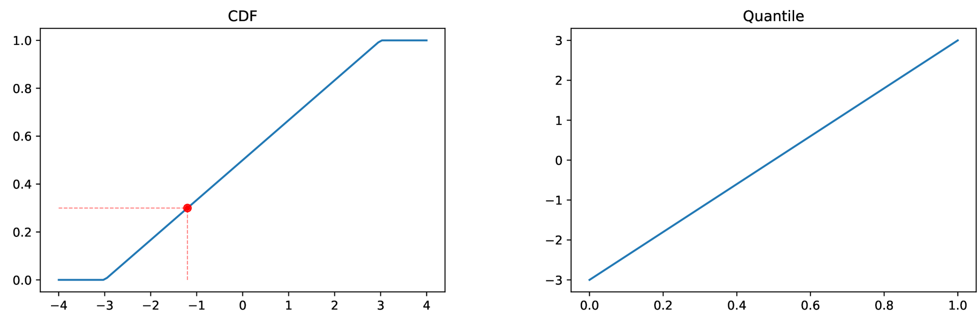

Figure 4.2: CDF (left) and quantile function (right) for the case of U[-3,3]. The quantile for 0.3 (i.e. -1.2) is obtained from the CDF plot: it is the smallest x having a CDF larger than or equal to the given value.

The uniform example before that if the cumulative distribution function has an inverse in a given domain, then the quantile function coincides with this inverse on this domain. This confirms that the quantile generalises the inverse of a function.

The following lemma shows another interesting property of the quantile function.

Lemma 4.20. *Let X be a random variable with c.d.f. F_X and quantile function q_x. Let U be a random variable distributed uniformly in $[0,1]$. Then, $q_X(U)$ and X have the same distribution.*

Proof. We need to show that the c.d.f. of $Y := q_X(U)$ and that of X coincide. Indeed, we have that for any $y \in \mathbb{R}$

$$F_Y(y) = \mathbb{P}[Y \leq y] = \mathbb{P}[q_X(U) \leq y] = \mathbb{P}[U \leq F_X(y)] = F_X(y),$$

where the third inequality holds since $\{\omega \in \Omega : q_X(U(\omega)) \leq y\} = \{\omega \in \Omega : U(\omega) \leq F_X(y)\}$ as a consequence of (4.2), and the last inequality follows from the fact that U is uniform in $[0,1]$ and $F_X(y) \in [0,1]$. $\square$

4.3.1 Value at risk

Probably the most popular risk measure is value at risk. Informally, it means to represent the lower value of the 'worst possible losses' with a given confidence probability.

Definition 4.21 (Value at risk)**.** Value at risk at level $\alpha \in (0,1)$ of a wealth X, that we denote by $\mathrm{V@R}^\alpha(\mathrm{X})$, is defined by

$$\mathrm{V@R}^\alpha(\mathrm{X}) = \mathrm{q}_{-\mathrm{X}}(\alpha) = \inf\{\mathrm{x} \in \mathbb{R} : \mathbb{P}(-\mathrm{X} \leq \mathrm{x}) \geq \alpha\} = \inf\{\mathrm{x} \in \mathbb{R} : \mathbb{P}(\mathrm{X} + \mathrm{x} \geq 0) \geq \alpha\}$$

If we assume that X is a random variable representing profit and losses (P&L), we can interpret value at risk as the minimal value such that when added to our current $P\&L$ would guarantee no losses with a probability α. Typical values used in market risk management practice are: $\alpha = 95\%$ or $\alpha = 99\%$.

(!). *Different references have different conventions. Note that the convention here is to calculate the value at risk of a random variable that represents, for example, the profile of results (with positive values meaning profits). The value at risk would have positive value if there is a risk of having negative results. Moreover,* α *has the interpretation of a 'confidence' level of large losses, so it is typically taken to be close to 1. Some old versions of these notes had a different convention.*

Example 4.22. *Assume that a given investment either generates a profit of* 100 *with probability* 0.75*, or a loss of* 150 *with probability* 0.25*. What is the value at risk at level* $\alpha \in (0,1)$ *of the P& L of this investment?*

We call X *the random variable representing the P&L. Let* $L = -X$*, representing the losses of this investment. We would have*

$$F_L(x) = 0.75\mathbb{1}_{x \geq -100} + 0.25\mathbb{1}_{x \geq 150}$$

By following the same procedure as in Example 4.18, we have that

$$\mathrm{V@R}^\alpha(\mathrm{X}) = \mathrm{q}_\mathrm{L}(\lambda) = -100\mathbb{1}_{0<\lambda\leq 0.75} + 150\mathbb{1}_{0.75<\lambda<1}$$

In particular the 99% *value at risk is 150 (the maximum loss!) but the value at risk at level* 75% *is -100 (the negative sign indicates that no losses are expected at that level). This example shows some limitations of* V@R *as a risk measure.*

Example 4.23. *Assume that* $X \sim U[-1,3]$ *represents the net profits of a given investment. What is its value at risk at level* $\alpha \in (0,1)$*?*

By following the same procedure as in Example 4.19, we have that

$$\mathrm{V@R}^\alpha(\mathrm{X}) = \mathrm{q}_{-\mathrm{X}}(\alpha) = \mathrm{q}_{[-3,1]}(\alpha) = 4\alpha - 3$$

In particular the 99% *value at risk is 0.96. So, if we added 0.96 to our initial profit and loss profile, we would not experience losses in* 99% *of the cases.*

Properties of value at risk

$\mathrm{V@R}^\alpha$ is extremely popular: it has a very simple economic interpretation, and it can be, in general easily approximated and tested (we will explore about this in later chapters). It satisfies monotonicity, translation invariance (hence monetary) and the positive homogeneity properties. The proofs are all very similar. Let us show, for instance that V@R is cash invariant.

Proposition 4.24. $\mathrm{V@R}^\alpha$ *is a cash invariant risk measure.*

Proof. From the definition of V@R we have

$$\begin{aligned}
\mathrm{V@R}^\alpha(\mathrm{X}+\mathrm{a}) &= \inf\{x \in \mathbb{R} : \mathbb{P}(-(X+a) \le x) \ge \alpha\} \\
&= \inf\{x \in \mathbb{R} : \mathbb{P}(-X \le x+a) \ge \alpha\} \\
&= \inf\{z-a \in \mathbb{R} : \mathbb{P}(-X \le z) \ge \alpha\} \\
&= \inf\{z \in \mathbb{R} : \mathbb{P}(-X \le z) \ge \alpha\} - a \\
&= \mathrm{V@R}^\alpha(\mathrm{X}) - \mathrm{a}.
\end{aligned}$$

□

Value at risk is also a law-invariant risk measure, which is very convenient when applied to market data. However, V@R is **not** convex or subadditive. Recall that this means that, in general, V@R does not properly recognise diversification as a risk reduction tool.

$\mathrm{V@R}^\alpha$ is, nevertheless, subadditive in some cases where the loss vector L, comes from a portfolio composed as linear combination of assets following an *elliptical distribution*: we define them using their characteristic function. Recall that the characteristic function of a d-dimensional distribution is its Fourier transform, i.e, for $t \in \mathbb{R}^d$

$$\varphi_X(t) := \mathbb{E}[e^{it^\top X}],$$

and it completely characterises a distribution.

Definition 4.25. We say that a family of probability distributions in dimension d is elliptical if there exists a characteristic function of a scalar variable ψ (called the generator), a vector $\mu \in \mathbb{R}^d$ and a matrix $\Sigma \in M^{d\times d}$ such that

$$\varphi_X(t) = e^{it^\top \mu}\psi(t^\top \Sigma t).$$

where φ_X denotes the probability density function of X.

Some examples include the (multivariate) Gaussian, $t-$Student, Laplace and logistic distributions.

4.3.2 Expected shortfall

There are some drawbacks of using value at risk. On the one hand, since it focuses on only one quantile, it ignores any information beyond that quantile. In other words it does not inform of how severe the uncommon losses are. Moreover, we remarked that value at risk is not convex. These two shortfalls, together with the eagerness of financial companies to reduce their capital without reducing their average returns, lead to strategies that were barely diversified and carried huge risks with low probability.

Expected shortfall was proposed to deal with these issues. Assuming that X *is a random variable with finite mean*, instead of reporting one quantile which can be seen as the "lower bound of extreme losses", one can report the average of all the quantiles above a certain margin. This is what is called Expected Shortfall[2] at level α.

[2]In this lecture notes we identify expected shortfall, average value at risk and conditional value at risk, which sometimes are distinguished in the literature, particularly for discrete random variables.

Definition 4.26 (Expected shortfall)**.** For a random variable $X \in L^1(\Omega, \mathbb{P})$, *expected shortfall at level* α, denoted $\mathrm{ES}^\alpha(\mathrm{X})$ is defined by the expression

$$\mathrm{ES}^\alpha(\mathrm{X}) = \frac{1}{1-\alpha}\int_\alpha^1 \mathrm{V@R^u(X)du}.$$

Note that this definition captures exactly the notion of averaging value at risks at different levels.

(!). *Pay attention to the factor in front of the integral limits of the integral and its sign. In references where a different convention for value at risk is used, they change for consistency.*

Other representations

Expected shortfall has other representations that can be very useful to understand its role and properties.

Proposition 4.27. *Let* $\tilde{\alpha} := F_{-X}(\mathrm{V@R}^\alpha(\mathrm{X}))$*. Then,*

$$\mathrm{ES}^\alpha(\mathrm{X}) = \frac{1}{1-\alpha}\left\{\mathbb{E}[-\mathrm{X}\cdot \mathbb{1}_{\{-\mathrm{X}>\mathrm{V@R}^\alpha(\mathrm{X})\}}] + (\tilde{\alpha}-\alpha)\mathrm{V@R}^\alpha(\mathrm{X})\right\}. \tag{4.3}$$

Proof. We have that

$$\mathrm{ES}^\alpha(\mathrm{X}) = \frac{1}{1-\alpha}\int_\alpha^1 \mathrm{V@R^u(x)du} = \frac{1}{1-\alpha}\int_\alpha^1 \mathrm{q_{-X}(u)du} = \frac{1}{1-\alpha}\mathbb{E}[\mathrm{q_{-X}(U)}\mathbb{1}_{\{\mathrm{U}\geq\alpha\}}]$$

for a random variable U on $[0,1]$. Now, let $\tilde{\alpha}$ be given as in the statement of the proposition. From (4.1), it is clear that $0<\alpha\leq\tilde{\alpha}\leq 1$. We can then write

$$\mathrm{ES}^\alpha(\mathrm{X}) = \frac{1}{1-\alpha}\mathbb{E}[q_{-X}(U)(\mathbb{1}_{U\geq\tilde{\alpha}} + \mathbb{1}_{\alpha\leq U\leq\tilde{\alpha}})] = \frac{1}{1-\alpha}\mathbb{E}[q_{-X}(U)\mathbb{1}_{U\geq\tilde{\alpha}}] + \frac{1}{1-\alpha}\mathbb{E}[q_{-X}(U)\mathbb{1}_{\alpha\leq U\leq\tilde{\alpha}}]$$

Note that $\tilde{\alpha} := F_{-X}(\mathrm{V@R}^\alpha(\mathrm{X}))$ and by (4.2) it follows that $\mathbb{1}_{U\geq\tilde{\alpha}} = \mathbb{1}_{q_{-X}(U)\geq\mathrm{V@R}^\alpha(\mathrm{X})}$. Hence, using in addition Lemma (4.20), it follows that

$$\begin{aligned}\mathrm{ES}^\alpha(\mathrm{X}) &= \frac{1}{1-\alpha}\mathbb{E}[q_{-X}(U)\mathbb{1}_{q_{-X}(U)\geq\mathrm{V@R}^\alpha(\mathrm{X})}] + \frac{1}{1-\alpha}\mathbb{E}[q_{-X}(U)\mathbb{1}_{\alpha\leq U\leq\tilde{\alpha}}]\\ &= \frac{1}{1-\alpha}\mathbb{E}[-X\mathbb{1}_{-X\geq\mathrm{V@R}^\alpha(\mathrm{X})}] + \frac{1}{1-\alpha}\mathbb{E}[q_{-X}(U)\mathbb{1}_{\alpha\leq U\leq\tilde{\alpha}}].\end{aligned}$$

Finally, note that from the definition of quantile, it follows that $q_{-X}(\lambda) = q_{-X}(\alpha)$ for all $\lambda \in [\alpha, \tilde{\alpha}]$. Thus,

$$\begin{aligned}\mathrm{ES}^\alpha(\mathrm{X}) &= \frac{1}{1-\alpha}\mathbb{E}[-X\mathbb{1}_{-X\geq\mathrm{V@R}^\alpha(\mathrm{X})}] + \frac{1}{1-\alpha}\mathbb{E}[\mathrm{V@R}^\alpha(\mathrm{X})\mathbb{1}_{\alpha\leq\mathrm{U}\leq\tilde{\alpha}}]\\ &= \frac{1}{1-\alpha}\mathbb{E}[-X\mathbb{1}_{-X\geq\mathrm{V@R}^\alpha(\mathrm{X})}] + \frac{\tilde{\alpha}-\alpha}{1-\alpha}\mathrm{V@R}^\alpha(\mathrm{X}).\end{aligned}$$

□

Using the definition of the expectation conditional to the event $-X \leq \mathrm{V@R}^\alpha(\mathrm{X})$ we immediately deduce

Corollary 4.28. *Let $\beta = (1-\tilde{\alpha})/(1-\alpha)$ where $\tilde{\alpha}$ is defined as in Proposition 4.27. Then*

$$\mathrm{ES}^{\alpha}(\mathrm{X}) = \beta\mathbb{E}[-\mathrm{X}| - \mathrm{X} > \mathrm{V@R}^{\alpha}(\mathrm{X})] + (1-\beta)\mathrm{V@R}^{\alpha}(\mathrm{X}). \tag{4.4}$$

Clearly, the second term in both (4.3) and (4.4) vanishes when $\tilde{\alpha} = \alpha$ (for example if the distribution has an inverse).

Corollary 4.29 below expresses expected shortfall in yet another way. It can be interpreted as adding the expected value plus the expected value of a put option with strike minus the expected value.

Corollary 4.29.

$$\mathrm{ES}^{\alpha}(\mathrm{X}) = \mathrm{V@R}^{\alpha}(\mathrm{X}) + \frac{1}{1-\alpha}\mathbb{E}[(-\mathrm{X} - \mathrm{V@R}^{\alpha}(\mathrm{X}))^{+}] = \mathrm{V@R}^{\alpha}(\mathrm{X}) + \frac{1}{1-\alpha}\mathbb{E}[(\mathrm{X} + \mathrm{V@R}^{\alpha}(\mathrm{X}))^{-}]. \tag{4.5}$$

where $a \wedge b = \min\{a, b\}$, $a \vee b = \max\{a, b\}$, $(a)^{+} := a \vee 0$ and $(a)^{-} := (-a)^{+}$.

Exercise. Show Corollaries 4.28 and 4.29

In fact we can strengthen Corollary 4.29 as follows

Proposition 4.30.

$$\mathrm{ES}^{\alpha}(\mathrm{X}) = \inf_{\mathrm{r}\in\mathbb{R}} \left\{ \mathrm{r} + \frac{1}{1-\alpha}\mathbb{E}[(\mathrm{X}+\mathrm{r})^{-}] \right\}.$$

The proof is based on the fact that the function to minimise is convex (so every local optimal is global) and can be worked out by perturbing the proposed optimal argument $r^{*} = \mathrm{V@R}^{\alpha}(\mathrm{X})$, and then concluding from Corollary 4.29.

Exercise. Show Proposition 4.30.

Examples

Example 4.31. *If we calculate value at risk with the data of example* (4.22)*, we get that*

$$\mathrm{ES}^{\alpha}(\mathrm{X}) = \frac{1}{1-\alpha}\int_{\alpha}^{1} \mathrm{V@R}^{\mathrm{u}}(\mathrm{X})\mathrm{du} = \frac{1}{1-\alpha}(-100(0.75-\alpha)^{+} + 150(1-(\alpha \vee 0.75))).$$

For example, we get that $\mathrm{ES}^{0.5}(\mathrm{X}) = 2(-25+37.5) = 25$, $\mathrm{ES}^{0.75}(\mathrm{X}) = 150$ *and* $\mathrm{ES}^{0.99}(\mathrm{X}) = 150$.

Example 4.32. *If we calculate value at risk with the data of example* (4.23)*, we get that*

$$\mathrm{ES}^{\alpha}(\mathrm{X}) = \frac{1}{1-\alpha}\int_{\alpha}^{1} \mathrm{V@R}^{\mathrm{u}}(\mathrm{X})\mathrm{du} = \frac{1}{1-\alpha}\int_{\alpha}^{1}(4\mathrm{u}-3)\mathrm{du} = \frac{2}{1-\alpha}(1-\alpha^{2}) - 3 = 2\alpha - 1$$

In particular, we get that $\mathrm{ES}^{0.99}(\mathrm{X}) = 0.98$.

ES^{α} is a coherent (and hence also a convex) risk measure. It is also very popular, and has progressively replaced V@R as the de facto risk measure for risk management.

The main drawbacks of expected shortfall are that it can only applied to random variables with finite mean (which exclude some heavy tail distributions), it is more difficult to estimate and to validate than V@R.

Properties

It is easy to show that the fact that expected shortfall inherits the monotonicity, cash invariance and positive homogeneity properties from value at risk, thanks to the monotonicity and linearity properties of the integral. Expected shortfall is also law-invariant. However, expected shortfall is in addition subadditive (and thus convex).

Proposition 4.33. ES^α *is subadditive.*

Proof. Let $X, Y \in L^2$. Then it follows from Proposition 4.30 that

$$\begin{aligned}\mathrm{ES}^\alpha(\mathrm{X}) + \mathrm{ES}^\alpha(\mathrm{Y}) &= \inf_{r\in\mathbb{R}} \left\{ r + \frac{1}{1-\alpha}\mathbb{E}[(X+r)^-] \right\} + \inf_{z\in\mathbb{R}} \left\{ z + \frac{1}{1-\alpha}\mathbb{E}[(Y+z)^-] \right\} \\ &= \inf_{r\in\mathbb{R}, z\in\mathbb{R}} \left\{ (r+z) + \frac{1}{1-\alpha}\mathbb{E}[(X+r)^- + (Y+z)^-] \right\}.\end{aligned}$$

But since for any $a, b \in \mathbb{R}$

$$(a)^- + (b)^- = \frac{|a|-a}{2} + \frac{|b|-b}{2} \geq \frac{|a+b|-(a+b)}{2} = (a+b)^-,$$

it follows that

$$\mathrm{ES}^\alpha(\mathrm{X}) + \mathrm{ES}^\alpha(\mathrm{Y}) \geq \inf_{\mathrm{r}\in\mathbb{R}, \mathrm{z}\in\mathbb{R}} \left\{ (\mathrm{r}+\mathrm{z}) + \frac{1}{1-\alpha}\mathbb{E}[([\mathrm{X}+\mathrm{Y}] + [\mathrm{r}+\mathrm{z}])^-] \right\} = \mathrm{ES}^\alpha(\mathrm{X}+\mathrm{Y}).$$

□

4.4 Summary of properties

	var	sd	$\rho_\theta^{\exp}$	$\mathrm{V@R}^\alpha$	ES^α
Monotonicity	$\star$	$\star$	✓	✓	✓
Translation invariance			✓	✓	✓
Subadditivity		✓	✓	†	✓
Positive homogeneity		✓		✓	✓
Convexity		✓	✓	†	✓
Normalisation	✓	✓	✓	✓	✓

$\star$ Variance and standard deviation are monotonous when restricting to losses with the same mean.
† V@R is subadditive (and hence convex) when considering linear combinations of a multidimensional elliptic function.

4.5 Robust representation(*)

Coherent and convex risk measures have a robust representation property in the sense that they can be expressed in terms that do not require a probability $\mathbb{P}$ to be fixed, provided that they satisfy the following regularity result.

Definition 4.34 (Continuity from above)**.** A risk measure is continuous from above if for any sequence of random variables such that $X_n \downarrow X$ implies $\rho(X_n) \uparrow \rho(X)$.

Let $\mathcal{M}$ be the set of probability measures such that for all $Q \in \mathcal{M}$, $\mathbb{E}^Q[X]$ is well-defined for all $X \in \mathcal{X}$. We have the following robust representation (see for example Föllmer and Schied (2011)).

Theorem 4.35. *Assume that ρ is a coherent risk measure continuous from above. Then, there exists a set $\mathcal{Q} \subset \mathcal{M}$ such that*

$$\rho(X) = \sup_{Q \in \mathcal{Q}} \left\{\mathbb{E}^Q[-X]\right\} \tag{4.6}$$

Moreover, $\mathcal{Q}$ can be chosen as a convex set for which the supremum is attained.

Intuitively, the theorem implies that a coherent risk measure takes into account uncertainty of the probability model of the possible outcomes: its value is the worst average loss that can be obtained amongst all the possible distributions in the set $\mathcal{Q}$.

As an example, assuming that $X \in L^1(\Omega, \mathbb{P})$, we get

$$\mathrm{ES}^\alpha(\mathrm{X}) = \sup_{\mathrm{Q} \in \tilde{\mathcal{Q}}_\alpha} \left\{\mathbb{E}^{\mathrm{Q}}[-\mathrm{X}]\right\}$$

where

$$\tilde{\mathcal{Q}}_\alpha = \{Q : Q \ll \mathbb{P}, \frac{\mathrm{d}Q}{\mathrm{d}P} \leq (1-\alpha)^{-1} a.s.\} \tag{4.7}$$

Convex risk measures

In the case of a convex risk measure, a similar representation exists. This time there is an extra penalisation term.

Theorem 4.36. *Let ρ be a convex risk measure continuous from above. Then there exist a set $\mathcal{Q} \subset \mathcal{M}$ and a function $\gamma : \mathcal{M} \to \mathbb{R}$ such that*

$$\rho(X) = \sup_{Q \in \mathcal{Q}} \left\{\mathbb{E}^Q[-X] - \gamma(Q)\right\}. \tag{4.8}$$

Furthermore, defining $\mathcal{A}_\rho = \{X : X \in \mathcal{X}; \rho(X) \leq 0\}$

$$\gamma(Q) = \sup_{X \in \mathcal{A}_\rho} \left\{\mathbb{E}^Q[-X]\right\}$$

Loosely speaking, the risk measure is once again an average on many possible measures on the set $\mathcal{Q}$, but where we have penalised (through the term γ) probability distributions that are considered "unrealistic".

Note that both robust representation theorems have converses, that is, risk measures defined using the right-hand side of expressions (4.6) and (4.8) are respectively coherent and convex.

4.6 Conditional and dynamic risk measure

We can extend the setting of one-period risk measures to study the perception of risks at different periods.

Definition 4.37. Let $t \in 0, \ldots, T$. A mapping $\rho_t : L^\infty(\Omega, \mathcal{F}, \mathbb{R}) \to L^\infty(\Omega, \mathcal{F}_t, \mathbb{R})$ is called a conditional risk measure.

The properties of static risk-measures generalise naturally to this definition:

- **Conditional cash invariance**: $\rho_t(X + m_t) = \rho_t(X) - m_t$ for all $m_t \in L_t^\infty$.
- **Monotonicity**: If $X \leq Y$ a.s. then $\rho_t(X) \geq \rho_t(Y)$.
- **Normalization**: $\rho_t(X) = 0$.
- **Conditional convexity**: For $\lambda \in: L_t^\infty$, $\lambda \in (0,1)$ we have $\rho_t(\lambda X + (1-\lambda)Y) \geq \lambda\rho_t(X) + (1-\lambda)\rho_t(Y)$.
- **Conditional positive homogeneity**: For $\lambda \in: L_t^\infty$, $\lambda > 0$ we have $\rho_t(\lambda X) = \lambda\rho_t(X)$.

We can obtain several examples of conditional risk measures from our one-period measures. For instance, we can define the following generalisations

- $\mathrm{V@R}_t^\alpha(\mathrm{X}) := \operatorname{ess\,inf}\{\mathrm{Z} \in \mathrm{L}_t^\infty : \mathbb{E}_t[\mathbb{1}_{-\mathrm{X} \leq \mathrm{Z}}] \geq \alpha\}$.
- $\mathrm{ES}_t^\alpha(\mathrm{X}) := \operatorname{ess\,inf}\{\mathrm{Z} + \frac{1}{1-\alpha}\mathbb{E}_t[(\mathrm{X}+\mathrm{Z})^-] : \mathrm{Z} \in \mathrm{L}_t^\infty\}$;

where for a measurable f, we say that $f_{e.inf}$ is the essential infimum of $f(X)$ for a random variable X if we have

$$f_{e.inf} = \sup\{a \in \mathbb{R} : \mathbb{P}[a \leq f(X)] = 1\},$$

i.e., the supremum of the 'almost sure' lower bounds of the function.

Remark 4.2. There is not a unique way of generalising one-period risk measures to several periods. To see other examples in the literature check Acciaio and Penner (2011).

A sequence of conditional risk measures $\{\rho_t\}_{t=0,\ldots,T}$ is denoted a *dynamic risk measure*. For further information about conditional and dynamic risk measures see Acciaio and Penner (2011).

4.7 Risk mitigation

A notion that is frequently used in risk management is that of acceptable positions. Mathematically, we model these positions as a subset of the set of all possible economic positions $\mathcal{A} \subset \mathcal{X}$. Investors and companies will avoid any position that falls outside this set. Risk mitigation is the process of using financial tools to render an economic position acceptable.

A simple way to determine an acceptability set is based on risk measures

Definition 4.38. The acceptability set associated to a risk measure ρ is defined by

$$\mathcal{A}_\rho = \{X : X \in \mathcal{X}; \rho(X) \leq 0\}.$$

A position X is **acceptable** if $X \in \mathcal{A}_\rho$.

In the following we examine some tools of risk mitigation under this framework.

4.7.1 Collateral addition

Collateral is understood as a financial deposit, usually required to be given in terms of a very secure asset, provided to mitigate losses associated to some type of default. *Note that the deposit is not used unless the risk is materialised.*

The amount of collateral required to render a position acceptable depends on the *numeraire* on which it is provided:

- If collateral is provided in a risk-less numéraire and ρ is **monetary** $\kappa = (\rho(X))^{+}$, that is the amount $\rho(X)$ if it is positive (which in practice is usually the case).

- If collateral is deposited using a defaultable numéraire N,

$$\kappa(X) = \inf_{r \in \mathbb{R}^{+}} \{r : \rho(X + rN) \leq 0\}.$$

  If ρ is strictly convex , there is at most a unique solution. If the solution is infinity (or infinity in practice) the numéraire is not adequate for collateral provision.

Capital as a type of collateral

One of the pillars of financial regulation is the need to provide capital to support any risky operation, as a way to reduce the likeliness of financial crises. In practice, financial regulation also establishes the types of instruments in which capital can be provided (for example AAA sovereign bonds with maturity 30 years or corporate AA+ bonds with maturity 10 years) and in which proportions (for example 70% and 30% respectively). See the notes and the end of the chapter for more information on capital regulation.

4.7.2 Hedging

Another way to mitigate risks is to *hedge* those risk, that is, to make a sequence of operations in the market to obtain payoffs that offset the initial risks.

With the convention that X denotes *discounted net position*, that is $X = \frac{\tilde{X}_1}{S_1^0}$ for some pay-off $\tilde{X}_1$ at final time, we look for a strategy $\boldsymbol{\theta}$ with the minimal initial cost such that $\rho(\frac{\boldsymbol{S}_1^{\boldsymbol{\theta}}}{S_1^0} - X) \leq 0$, i.e.

$$\boldsymbol{\theta}^* \in \arg\min\{\boldsymbol{\theta} \cdot \boldsymbol{S}_0 : \rho(\frac{S_1^{\boldsymbol{\theta}}}{S_1^0} - X) \leq 0\}$$

If such a $\boldsymbol{\theta}^*$ exists, we say that the strategy $\boldsymbol{\theta}^*$ partially hedges X. Of course, if there is a replication strategy i.e.,

$$X = \frac{\boldsymbol{S}_1^{\boldsymbol{\theta}}}{S_1^0},$$

the set on the right-hand side is not empty, but it can be seen that lower initial strategy costs are obtained if we limit ourselves to partial hedging.

4.8 Exercises

Exercise 4.1. Assume that $X \sim \mathcal{N}(\mu, \sigma^2)$ and $\alpha \in (0,1)$. Check that

$$\mathrm{V@R}^\alpha(X) = \sigma \mathbf{\Phi}^{-1}(\alpha) - \mu,$$

where $\mathbf{\Phi}$ denotes the standard Normal CDF.

Exercise 4.2. Assume that $X \sim \mathcal{N}(\mu, \sigma^2)$ and $\alpha \in (0,1)$. Check that

$$\mathrm{ES}^\alpha(X) = -\mu + \sigma \frac{\phi(\mathbf{\Phi}^{-1}(\alpha))}{1-\alpha},$$

where ϕ is the standard Normal PDF and $\mathbf{\Phi}$ denotes the standard Normal CDF.

Exercise 4.3. Compute $\mathrm{V@R}_\alpha(X)$ and $\mathrm{ES}_\alpha(X)$ when X is

- uniform
- log-normally distributed; that is, $X = e^{\sigma Z + m}$, where Z is standard normally distributed and $\sigma, m \in \mathbb{R}$.

Exercise 4.4. Give an example of two different distributions that have the same value at risk and expected shortfall at a 99% level. Can you define the example in such a way that one of the distributions would pose a bigger threat to the survival of a company? Justify your answer.

Exercise 4.5. Consider a setup with $d = 100$ bonds. These bonds can default. Each bond has price 100. Assume that defaults are independent and each default probability equals 2%. If there is a default, the bond pays 0, otherwise 105.

Consider two portfolios. Portfolio A consists of 100 copies of bond '1', Portfolio B consists of one unit of each of the bonds.

Compute $\mathrm{V@R}^{0.95}$ for each Portfolio A and Portfolio B (For the second computation, you may use any computer language, and find the correct value by trial-and-error). What do you observe?

Exercise 4.6. Give an alternative proof for (4.3) in the case where X has a strictly positive probability density function f_X. *Remark:* Note that this case, $\tilde{\alpha} = \alpha$.

Exercise 4.7. Show that the negative expectation $\rho(X) = \mathbb{E}[-X]$ is a coherent risk measure [3].

Exercise 4.8. Suppose that the probability space is finite, i.e. $\Omega = \{\omega_1, \dots, \omega_k\}$ Show that the worst-case risk measure $\rho(X) = \sup_i -X(\omega_i)$ is a coherent risk measure by verifying its properties. Find the set $\mathbb{Q}$ associated to the robust representation of this measure.

Exercise 4.9. Assume that ρ is a convex risk measure. (This is, ρ satisfies monotonicity, cash invariance, and convexity.) Assume also that ρ is normalised; that is, $\rho(0) = 0$. Show that

$$\rho(\lambda X) \le \lambda \rho(X), \quad \text{for all } \lambda \in [0,1]$$
$$\rho(\lambda X) \ge \lambda \rho(X), \quad \text{for all } \lambda \in [1,\infty)$$

[3] In fact, as observed in Artzner et al. (1999) all coherent measures are of this form, up to a change of probability measure.

Exercise 4.10. Consider buying a risky asset with price $S_0 > 0$ at the beginning of the period. Assume its price at the end of the period "S_1" is log-normally distributed with

$$S_1 = S_0 \exp(\xi); \quad \text{for} \quad \xi \sim \mathcal{N}(m, \sigma^2).$$

How much capital (in riskless numéraire and assuming no interest rate), would be needed to cover the risk of the profits or losses ($X = S_1 - S_0$) of the operation, when using:

1. $\mathrm{V@R}^\alpha$
2. ES^α

Exercise 4.11. Assume you already own the risky asset of exercise 4.10. Assume that it is possible to buy any nominal value of

- A call option with strike $K_c = S_0$ and price $p_c > 0$, that is an asset with pay-off $(S_1 - S_0)^+$
- A put option with strike $K_p = S_0$ and price $p_p > 0$, that is an asset with pay-off $(S_0 - S_1)^+$.

Can you partial hedge, replicate and super hedge your current $P\&L$ risk by buying one of those assets? Justify your answers giving the quantities of each option needed.

(*) Exercise 4.12. Let ρ be a coherent risk measure. Use the robust representation to show that ρ is additive, i.e.,

$$\rho(X + X') = \rho(X) + \rho(X')$$

if and only if the class $\mathcal{Q}$ reduces to a single probability measure $\mathbb{Q}$, i.e., ρ is simply the expected loss with respect to $\mathbb{Q}$.

(*) Exercise 4.13. Let $(X_n)_n$ denote a sequence of independent and identically distributed random variables such that $\mathbb{E}[|X_1|] < \infty$. Show that

$$\mathrm{V@R}^\alpha \left(\frac{1}{n} \sum_{i=1}^{n} X_i \right) \to -\mathbb{E}[X_1] \quad (n \uparrow \infty)$$

(Hint: Assume first that $\mathbb{E}[X_1] = 0$, and use the law of large numbers.)

4.9 Notes

On risk measures

Let us remark that as was shown in Exercise (4.4), there might be *blind spots*, for a risk measure that is, risky positions that might potentially harm the subsistence of a company, but that do not appear on a risk measure calculation.

There are many additional interesting properties that risk measures can have (ellicitability, robustness, ...), some of which we will examine when considering applications. The desire to have as many of these properties as possible has lead to many papers proposing new types of families of risk measures, but they sometimes sacrifice intuitive understanding. Some well-known families include expectiles and distortion risk measures.

On risk management and mitigation

The best source to learn about financial risk management regulation is to go to the website of the local regulator (The Bank of England, for example in the UK). However, an interesting overview of the regulation framework can be found on the documents of the Basel Committee on Banking Supervision (BCBS). The BCBS describes its objective as to *"enhance understanding of key supervisory issues and improve the quality of banking supervision worldwide"*. It is a joint effort of regulators of 28 countries to coordinate and improve banking/financial regulation, functioning more as a respected but informal forum on how to make regulation converge.

The most important contribution of the Basel Committee is the construction of the so-called Basel Accords, frameworks that state the general guidelines and principles that local regulators later develop. The most recent one, Basel III developed between 2010 and 2011, was introduced in the aftermath of the last financial crisis. Of particular importance is the chapter denoted Fundamental Review of the Trading Book.

Very schematically, the current paradigm for regulation(see for example **?**) allows for two approaches for capital calculation:

- The standardised approach for capital requirements is based on the linearisation-normal approximation. The idea is as follows: financial companies calculate sensitivities and exposures and the regulation provides the coefficients associated to them.

- In the internal model approach, an institution proposes a model for capital calculation (which, as shown, includes calculating risk measures) that is adjusted to their business, theoretically sound and empirically shown to work (not fail, in fact). It is subject to the approval of the regulator.

The agreement documents and the discussions to modify it on time are available and even open to comment on the Committees web-site: https://www.bis.org/bcbs/

4.10 Summary

- Risk measures are functions defined to estimate the riskiness of a position.
- A set of properties render these risk measures suitable for defining capital needs (monetary risk measures) and some extra assumptions incorporate risk mitigation obtained by diversification (convex and coherent risk measures).
- Important examples of risk measures can be defined from utility functions and from the tail of distributions (focusing on large losses). The most notable of the latter are Value at Risk and Expected Shortfall.
- Coherent and convex risk measures find a robust representation: they can be interpreted as the expected loss under alternative probability measures.
- (Static) risk measures can be extended to conditional and dynamic risk measures, which incorporate the changes of perception of risk as time evolves.
- The outcome of risk measurement is used to take decisions on whether to take measures to reduce it by avoiding or mitigating the risk.

Chapter 5

Equilibrium and CAPM

> The reality is that financial markets are self-destabilizing; occasionally they tend toward disequilibrium, not equilibrium.
>
> *George Soros*

5.1 Introduction to equilibrium

The prices of assets traded in a market are formed as market participants interact.

There are several models that try to understand this price formation: for example *market microstructure models* model the construction and evolution of prices as a result of the dynamics of the arrival of sell and buy orders on the market. Another view that occupies here is that of *equilibrium models*.

In this type of models, one understand the prices available for each available market instrument to be optimal when considering the aggregation needs and preferences of all participants. These are useful to understand the effects of changes in the market: starting from the equilibrium solution, and changing a rule (for example introducing taxes), one can study the impact of the new rule on the market.

In the following we give a taste of the equilibrium terminology and methods, following closely Back (2010). We study a simple investor setup, where there are no commodities to be consumed, and no endowment. We limit ourselves also to a case when all participants share the same family of utility functions but with different parameters, such that participants only respond to expected return and variance. This gives rise to the well-known CAPM model.

5.1.1 Setup of the equilibrium problem

Some common assumptions made to solve equilibrium problems are:

- Market instruments can be shorted and fractioned at will

- Information is available to all participants at the same time
- There are no transaction costs (taxes, fees, ...)
- Investors are small (price takers).
- All market participants have the same probability view and it coincides with the true probability.
- All investors consider the same time period

For the particular instance of equilibrium problem we are going to study, we assume in addition that:

- There are $n+1$ assets, one riskless assets and the other risky.
- Risky assets have jointly normal return with non-degenerate volatility matrix
- There are K investors, with initial wealth w_k each equipped with a utility $u_k(\cdot)$, for all $k = 1, \cdots, K$.

Remark 5.1. In this chapter we drop the subscript 1 associated to the period, and include instead a subscript (for example k) denoting the number of the investor. Thus the kth investor will hold a portfolio $\boldsymbol{\phi}_k$ with

$$\boldsymbol{\phi}_k = \begin{pmatrix} \phi_k^0 \\ \phi_k^1 \\ \vdots \\ \phi_k^n \end{pmatrix}, \qquad k = 1, 2, \dots, K.$$

which means that she places a wealth ϕ_k^i in the the ith asset at time 0. As before, we denote by $\hat{\boldsymbol{\phi}}_k$ the vector in $\mathbb{R}^n$ denoting the amounts invested by participant k in the risky assets only.

The central idea of equilibrium analysis is that *prices are produced as a consequence of utility maximisation and the fact that market clears*[1]. We have already worked with the notion of utility maximisation, so let us introduce market clearing.

For convenience, let us suppose that each risky stock with random (normally distributed) value S^i, for $i = 1, \dots, n$, is in **unit net supply**, that is, there is a total of one share of each risky asset available. This is not a restriction, as all assets are considered infinitely divisible, and so assuming there is one share of each asset available is just a normalisation.

Concerning the riskless asset, let us suppose it is in ***zero* net supply**. Riskless borrowing requires one investor to promise another to pay a named sum at time 1, and the total of all such promises made must equal the total of all such promises held.

Definition 5.1 (Market clearing)**.** The market is said to **clear** when the total stock holdings of all agents, for each stock, is equal to the total supply of each stock. In other words, when

$$\sum_{k=1}^{K} \phi_k^i = \begin{cases} p^i & \text{if } i = 1, \dots n \\ 0 & \text{if } i = 0 \end{cases}. \tag{5.1}$$

Note that this vector appears in isolation on the RHS of (5.1) since we have assumed that a total of one share of each stock is available, except in the case of the risk-less asset, for which there is zero net supply.

[1]That is, the supply and of the market must be matched

With these concepts, let us now define the notion of equilibrium.

Definition 5.2. A market is in equilibrium if:

- The market clears; and
- Any change in strategy of market participants that still clears the market will reduce the expected utility of some participants.

The second condition means that no trade would be possible: there are no two participants that will both have a benefit in trading (as per the condition one of them would be worse off). In particular, this second constraint is achieved if *all participants hold to the strategies maximising their utilities.*

Example 5.3. *Assume that all investors have CARA utility, the kth one having risk aversion α_k. Then, using that for each investor, the optimal investment is given by* (3.33)*, we can see that there is equilibrium if and only if the two inequalities*

$$\hat{\boldsymbol{p}} = \left(\sum_{k=1}^{K} \alpha_k^{-1}\right) \bar{\bar{\Sigma}}^{-1}(\boldsymbol{\mu} - R^0\mathbf{1})$$

and

$$\hat{\boldsymbol{p}} \cdot \mathbf{1} = \sum_{k=1}^{K} w_k$$

hold.

It is worth noticing that in equilibrium theory, price dynamics are an effect of changes in the probabilities of results for following periods as a result of information change: participants then adjust their strategies to new information until they reach a new equilibrium.

5.1.2 CAPM and its proof in the case of CARA utilities

CAPM is a result connecting the risk premia of assets with those of a reference *market model.* It can be deduced from an equilbrium argument in the framework of the example with all investors having CARA utility function, each with a possibly different risk aversion α_k, $k = 1, \dots, K$.

A key element in this result is the definition of a *market portfolio.*

Definition 5.4 (Market portfolio)**.** The *market portfolio* is one that holds each risky asset in a proportion identical to its value as a proportion of the total stock market value.

The initial wealth of the market portfolio, holding exactly one share of each stock, is

$$p_m = \sum_{i=1}^{n} p^i.$$

The proportion of wealth placed in asset i in the market portfolio is thus π_m^i, given by

$$\pi_m^i := \frac{p^i}{p_m}, \quad i = 1, \dots, n,$$

and we collect these weights in the vector $\boldsymbol{\pi}_M$, which satisfies the usual budget constraint $\boldsymbol{\pi}_m \cdot \mathbf{1} = 1$. With these weights, the return on the market portfolio is

$$R_m = \frac{S_m}{p_m} = \pi_m \cdot \hat{\boldsymbol{R}},$$

where, as usual, $\hat{\boldsymbol{R}}$ denotes the vector of returns of the risky stocks and $S_m = \mathbf{1} \cdot \hat{\boldsymbol{S}}$. Thanks to the Gaussian assumption, the return and the value are also Gaussian. The expected return is

$$\mu_m := \mathbb{E}[R_m] = \boldsymbol{\pi}_m \cdot \boldsymbol{\mu},$$

where, as usual, $\boldsymbol{\mu}$ denotes the vector of expected returns of the risky stocks. Moreover, the variance of the market portfolio is then

$$\mathrm{var}[R_m] = \boldsymbol{\pi}_m^\top \bar{\bar{\Sigma}} \boldsymbol{\pi}_m.$$

Theorem 5.5 (Capital Asset Pricing Model). *In equilibrium, with K agents each having CARA utility with risk aversion α_k for $k = 1, \ldots, K$, and with risky assets having normally distributed returns, $\hat{\boldsymbol{R}} \sim N(\boldsymbol{\mu}, \bar{\bar{\Sigma}})$, the risk premia satisfy*

$$\mu^i - R^0 = \beta^i(\mu_m - R^0), \quad i = 1, \ldots, n, \tag{5.2}$$

where β^i is the (so-called) beta *of asset i, defined by*

$$\beta^i = \frac{\mathrm{Cov}(R^i, R_m)}{\mathrm{var}[R_m]}, \quad i = 1, \ldots, n,$$

R_m is the return on the market portfolio, and $\mu_m = \mathbb{E}[R_m]$ is the expected return on the market portfolio.

Proof. As pointed out in the example (5.3), the equilibrium condition is

$$\bar{\bar{\Sigma}}^{-1} \left(\bar{\mu} - R^0 \mathbf{1}\right) \sum_{k=1}^{K} \frac{1}{\alpha_k} = \sum_{k=1}^{K} \boldsymbol{\phi}^k = \boldsymbol{p} = p_m \boldsymbol{\pi}_m.$$

Writing, for brevity,

$$\Gamma := \frac{1}{p_m} \sum_{k=1}^{K} \frac{1}{\alpha_k},$$

and solving for $\boldsymbol{\pi}_m$, gives

$$\boldsymbol{\pi}_m = \Gamma \bar{\bar{\Sigma}}^{-1} \left(\boldsymbol{\mu} - R^0 \mathbf{1}\right). \tag{5.3}$$

Now, using $R_m = \sum_{i=1}^{n} \pi_m^i R^i$,

$$\begin{aligned}
\mathrm{Cov}(R^i, R_m) &= \mathbb{E}\left[(R^i - \mu^i) \sum_{j=1}^{n} \pi_m^j (R_j - \mu^j)\right] \\
&= \sum_{j=1}^{n} \Sigma_{i,j} \pi_m^i \\
&= (\bar{\bar{\Sigma}} \boldsymbol{\pi}_m)_i \\
&= \Gamma(\mu^i - R^0)
\end{aligned}$$

Moreover,

$$\begin{aligned}\operatorname{var}[R_m] &= (\boldsymbol{\pi}_m)^\top \bar{\bar{\Sigma}} \boldsymbol{\pi}_m \\ &= \Gamma\,(\boldsymbol{\pi}_m)^\top \left(\boldsymbol{\mu} - R^0\mathbf{1}\right) \\ &= \Gamma\left(\mu_m - R^0\right).\end{aligned}$$

Hence

$$\beta_i = \frac{\operatorname{Cov}(R^i, R_m)}{\operatorname{var}[R_m]} = \frac{\mu^i - R^0}{\mu_m - R^0}.$$

This yields (5.2). □

Remark 5.2. Recalling that $\phi_m = \pi_m p_m$ and setting $\Gamma' = \Gamma p_m$, we have, from (5.3),

$$\phi_m = \Gamma' \bar{\bar{\Sigma}}^{-1}\left(\boldsymbol{\mu} - R^0\mathbf{1}\right).$$

Comparing with (3.33), we see that each CARA agent invests his risky assets in a multiple of the market portfolio.

The conclusion is very strong: if the CAPM model holds, it gives ground to the existence of reference portfolios (the proxy of this in practice are indexes), and essentially says that for the type of investors we examined, the optimal action is simply to invest in a multiple of the index.

The solution can also be read in another light: CAPM predicts a linear relationship between excess returns on assets and the excess return on the market portfolio. Its fame owes much to the fact that it makes a strong empirically verifiable prediction between, on the one hand, the mean rates of returns of individual assets and of the market portfolio, and on the other, the variances and covariances of asset returns. In principle these quantities could all be estimated from market data (the market portfolio would be taken to be a major share index), thereby providing a test of the CAPM theory. It is rare to find such a verifiable prediction from economic theory. Unfortunately, however, it turns out to be virtually impossible to make reliable estimates of expected returns. Furthermore, attempts to provide those estimates seem to suggest, especially the ones including more recent information. Many researchers have then tried to propose modifications to the theory to obtain better predictions (for example the works of Fama and French). We will see some of these alternative models when discussing factor models.

5.2 Exercises

Exercise 5.1. Prove a modified CAPM as in Theorem 5.5, but now assume that each of the K investors evaluates its utility with the function $U_k(X) = \zeta_k \mathbb{E}[X] - \operatorname{var}[X]/2$, for all $k = 1, \cdots, K$ for any random wealth X.

Exercise 5.2. Show that if all the other assumptions are kept, CAPM still holds with end of period endowment if these endowments are replicable and they are in zero net supply.

Chapter 6

Performance measurement and efficient frontiers in a one-period market model

> If my future were determined just by my performance on a standardized test, I wouldn't be here.
>
> *Michelle Obama*

6.1 Risk and return

As we have presented before, investors with different utility functions, might have different opinions on how risky a position is. However, let $0 < r$ be a deterministic positive number and R be a random return. Then, for any utility function we have

$$R \leq r + R \Rightarrow \mathbb{E}[u(R)] < \mathbb{E}[u(r+R)].$$

This very simple equation can be understood as follows: *if two assets have the same randomness but different mean, then a rational investor will prefer the one with the highest mean.*

Similarly, consider two positive numbers $0 < \sigma_1 < \sigma_2$ and two random returns $R_1, \Delta R_2$ distributed as $R_1 \sim \mathcal{N}(m, \sigma_1^2); \Delta R_2 \sim \mathcal{N}(0, \sigma_2^2 - \sigma_1^2)$, with R_1 and ΔR_2 independent. Let $R_2 = R_1 + \Delta R_2$ (so that R_2 is normal with variance σ_2^2). If we assume that our investor has concave utility function, we get from Jensen's inequality that

$$\begin{aligned}
\mathbb{E}[u(R_2)] =& \mathbb{E}[u(R_1 + \Delta R_2)] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} u(x+y) f_{\Delta R_2}(y) f_{R_1}(x) dy dx \\
&\leq \int_{-\infty}^{\infty} u\left(x + \int_{-\infty}^{\infty} y f_{\Delta R_2}(y)\right) f_{R_1}(x) dx = \int_{-\infty}^{\infty} u(x) f_{R_1}(x) dx \\
&= \mathbb{E}[u(R_1)]
\end{aligned}$$

In other words, between two assets with pay-off distributed as normal variables with the same mean, *every rational investor with a concave utility function (ergo risk-averse) will choose the one with smallest variance.*

These two simple facts lie behind the work by H. Markowitz on portfolio choice that gave rise to modern portfolio theory and was recognised with the (shared) Nobel price in economics in 1990. He proposes the mean and variance of the pay-off of an asset as indicators of the return and risk of the asset respectively. He observes that, accepting these two indicators, there are some assets that *dominate* other assets:

Definition 6.1. We say that a portfolio with return R_1 dominates (in mean-variance sense) another with return R_2 if either $\mathbb{E}[R_1] > \mathbb{E}[R_2]$ and $\operatorname{var}(R_1) \leq \operatorname{var}(R_2)$ or if $\mathbb{E}[R_1] \geq \mathbb{E}[R_2]$ and $\operatorname{var}(R_1) < \operatorname{var}(R_2)$.

As we have argued before (at least for the case of normal variables), every rational and risk-averse investor[1] will prefer R_1 over R_2 in the two described cases. This means that a rational investor, having the choice, will prefer portfolios that are not dominated by any other portfolio. Moreover, it would only invest in a portfolio belonging to a subset of all the possible portfolios available to her, that is, on the portfolios that are not dominated by any other portfolio.

Definition 6.2. The *mean-variance frontier of risky assets* is the set of portfolios that have minimum variance among all portfolios with the same expected return.

In other words, the mean-variance frontier is comprised by all portfolios which are not dominated.

We study the calculation of the efficient frontier theory in the following. This chapter is based mainly on Chapter 5 in Back (2010).

6.2 Preliminaries

With the notation of previous chapters we have the following assumptions:

- The covariance matrix $\bar{\bar{\Sigma}}$ of the risky assets is non-singular; that is, no portfolio that contains risky assets is risk free, and there are not redundant portfolios.
- At least two risky assets have different expected returns.

6.3 The calculus approach in the case with no risk-free assets

For the moment, we assume that there is *no* risk-free asset, that is, we assume that $\mathbf{1}^T \hat{\boldsymbol{\pi}} = 1$.

To compute the mean-variance frontier we need to solve the following optimisation problem, for fixed μ_p:

$$\min_{\hat{\boldsymbol{\pi}}} \frac{1}{2} \hat{\boldsymbol{\pi}}^\top \bar{\bar{\Sigma}} \hat{\boldsymbol{\pi}} \tag{6.1}$$

[1] More precisely an investor with strictly increasing and concave utility function.

subject to

$$\begin{aligned}\boldsymbol{\mu}^\top\hat{\boldsymbol{\pi}} &= \mu_p;\\ \mathbf{1}^\top\hat{\boldsymbol{\pi}} &= 1.\end{aligned}\tag{6.2}$$

Remark 6.1. Note that we do not satisfy all the assumptions we introduced before to guarantee the uniqueness of the solution: notably, the function that we are minimising is not strictly increasing.However, this is a problem that is known to have a unique solution: quadratic with linear constraints

To solve it we first introduce the Lagrange coefficients δ, γ and compute the Lagrangian

$$\mathcal{L}(\hat{\boldsymbol{\pi}}, \delta, \gamma) = \frac{1}{2}\hat{\boldsymbol{\pi}}^\top\bar{\bar{\Sigma}}\hat{\boldsymbol{\pi}} - \delta(\boldsymbol{\mu}^T\hat{\boldsymbol{\pi}} - \mu_p) - \gamma(\mathbf{1}^T\hat{\boldsymbol{\pi}} - 1)$$

This yields the first-order condition

$$0 = \nabla_{\hat{\boldsymbol{\pi}}}\mathcal{L}(\hat{\boldsymbol{\pi}}, \delta, \gamma) = \bar{\bar{\Sigma}}\hat{\boldsymbol{\pi}} - \delta\boldsymbol{\mu} - \gamma\mathbf{1}.\tag{6.3}$$

plus the two additional conditions related to satisfying the constraints.

Exercise. Check the last equality in (6.3)!

This then yields

$$\hat{\boldsymbol{\pi}} = \delta\bar{\bar{\Sigma}}^{-1}\boldsymbol{\mu} + \gamma\bar{\bar{\Sigma}}^{-1}\mathbf{1}\tag{6.4}$$

and the constraints can be written as

$$\mu_p = \delta\boldsymbol{\mu}^\top\bar{\bar{\Sigma}}^{-1}\boldsymbol{\mu} + \gamma\boldsymbol{\mu}^\top\bar{\bar{\Sigma}}^{-1}\mathbf{1};\tag{6.5}$$

$$1 = \delta\mathbf{1}^T\bar{\bar{\Sigma}}^{-1}\boldsymbol{\mu} + \gamma\mathbf{1}^T\bar{\bar{\Sigma}}^{-1}\mathbf{1}.\tag{6.6}$$

We now need to solve for δ and γ.

Rearranging the last equality yields

$$\delta = \frac{1 - \gamma\mathbf{1}^T\bar{\bar{\Sigma}}^{-1}\mathbf{1}}{\mathbf{1}^T\bar{\bar{\Sigma}}^{-1}\boldsymbol{\mu}}\tag{6.7}$$

Exercise. Solve the optimisation problem (6.1)–(6.2) explicitly.

6.4 Two-fund spanning

Given (6.4)–(6.6), we could also solve the optimisation problem (6.1)–(6.2) in the following way:

First, define two portfolios of risky assets:

$$\bar{\pi}_{\boldsymbol{\mu}} = \frac{1}{\mathbf{1}^T\bar{\bar{\Sigma}}^{-1}\boldsymbol{\mu}}\bar{\bar{\Sigma}}^{-1}\boldsymbol{\mu} \qquad \Rightarrow \qquad \mathbf{1}^T\bar{\pi}_{\boldsymbol{\mu}} = 1;\tag{6.8}$$

$$\bar{\pi}_{\mathbf{1}} = \frac{1}{\mathbf{1}^T\bar{\bar{\Sigma}}^{-1}\mathbf{1}}\bar{\bar{\Sigma}}^{-1}\mathbf{1} \qquad \Rightarrow \qquad \mathbf{1}^T\bar{\pi}_{\mathbf{1}} = 1.\tag{6.9}$$

From (6.4) and (6.6), we obtain

$$\begin{aligned}\hat{\boldsymbol{\pi}} &= \delta\bar{\bar{\Sigma}}^{-1}\boldsymbol{\mu} + \gamma\bar{\bar{\Sigma}}^{-1}\mathbf{1} = (\delta\mathbf{1}^T\bar{\bar{\Sigma}}^{-1}\boldsymbol{\mu})\bar{\pi}_{\boldsymbol{\mu}} + (\gamma\mathbf{1}^T\bar{\bar{\Sigma}}^{-1}\mathbf{1})\bar{\pi}_{\mathbf{1}} \\ &= \lambda\bar{\pi}_{\boldsymbol{\mu}} + (1-\lambda)\bar{\pi}_{\mathbf{1}}\end{aligned}$$

with $\lambda = \delta\mathbf{1}^T\bar{\bar{\Sigma}}^{-1}\boldsymbol{\mu}$. To compute λ, use

$$\mu_p = \boldsymbol{\mu}^\top\bar{\pi} = \lambda\boldsymbol{\mu}^\top\bar{\pi}_{\boldsymbol{\mu}} + (1-\lambda)\boldsymbol{\mu}^\top\ \bar{\pi}_{\mathbf{1}}$$

yielding

$$\lambda = \frac{\mu_p - \boldsymbol{\mu}^T\bar{\pi}_{\mathbf{1}}}{\boldsymbol{\mu}^T\bar{\pi}_{\boldsymbol{\mu}} - \boldsymbol{\mu}^T\bar{\pi}_{\mathbf{1}}} \tag{6.10}$$

Therefore, the mean-variance optimal portfolio is

$$\hat{\boldsymbol{\pi}} = \frac{\mu_p - \boldsymbol{\mu}^T\bar{\pi}_{\mathbf{1}}}{\boldsymbol{\mu}^T\bar{\pi}_{\boldsymbol{\mu}} - \boldsymbol{\mu}^T\bar{\pi}_{\mathbf{1}}}\bar{\pi}_{\boldsymbol{\mu}} + \frac{\boldsymbol{\mu}^T\bar{\pi}_{\boldsymbol{\mu}} - \mu_p}{\boldsymbol{\mu}^T\bar{\pi}_{\boldsymbol{\mu}} - \boldsymbol{\mu}^T\bar{\pi}_{\mathbf{1}}}\bar{\pi}_{\mathbf{1}}. \tag{6.11}$$

Exercise. At the beginning of this section we have made the assumption that at least two risky assets have different expected returns. Argue that this assumption guarantees that λ is well defined, that is, that $\boldsymbol{\mu}^T\bar{\pi}_{\boldsymbol{\mu}} - \boldsymbol{\mu}^T\bar{\pi}_{\mathbf{1}} \neq 0$.

Observation: The portfolio $\bar{\pi}^*$ is a linear combination of the two portfolios $\bar{\pi}_{\mathbf{1}}$ and $\bar{\pi}_{\boldsymbol{\mu}}$. Therefore, any investor that invests on the mean-variance frontier is content to allocate her wealth among two *mutual funds*. This is sometimes called a **two-fund theorem**.

Note that any two funds in the mean-variance frontier span the whole frontier (see Exercise 6.1)

6.5 Mean-standard deviation trade-off

Variance of **any** frontier portfolio:

$$\sigma_p^2 = (\hat{\boldsymbol{\pi}})^\top\bar{\bar{\Sigma}}\hat{\boldsymbol{\pi}} = (\lambda\hat{\boldsymbol{\pi}}_{\boldsymbol{\mu}} + (1-\lambda)\hat{\boldsymbol{\pi}}_{\mathbf{1}})^\top\bar{\bar{\Sigma}}(\lambda\hat{\boldsymbol{\pi}}_{\boldsymbol{\mu}} + (1-\lambda)\hat{\boldsymbol{\pi}}_{\mathbf{1}}) \tag{6.12}$$

Fix the following constants:

$$A = \boldsymbol{\mu}^T\bar{\bar{\Sigma}}^{-1}\boldsymbol{\mu}, \quad B = \boldsymbol{\mu}^T\bar{\bar{\Sigma}}^{-1}\mathbf{1}, \quad C = \mathbf{1}^T\bar{\bar{\Sigma}}^{-1}\mathbf{1} \tag{6.13}$$

Exercise. Show that, with λ given in (6.10),

$$\lambda = \frac{BC\mu_p - B^2}{AC - B^2} \tag{6.14}$$

and σ_p^2, given in (6.12) satisfies

$$\sigma_p^2 = \frac{A - 2B\mu_p + C\mu_p^2}{AC - B^2}. \tag{6.15}$$

Therefore, the variance is a quadratic function of the mean μ_p. Graphically, we usually plot the (standard deviation, mean) pairs:

$$\left(\sqrt{\frac{A - 2B\mu_p + C\mu_p^2}{AC - B^2}}, \mu_p\right) \tag{6.16}$$

Note that in most cases there are two portfolio with the same variance, but the efficient frontier is only composed by the ones with higher mean.

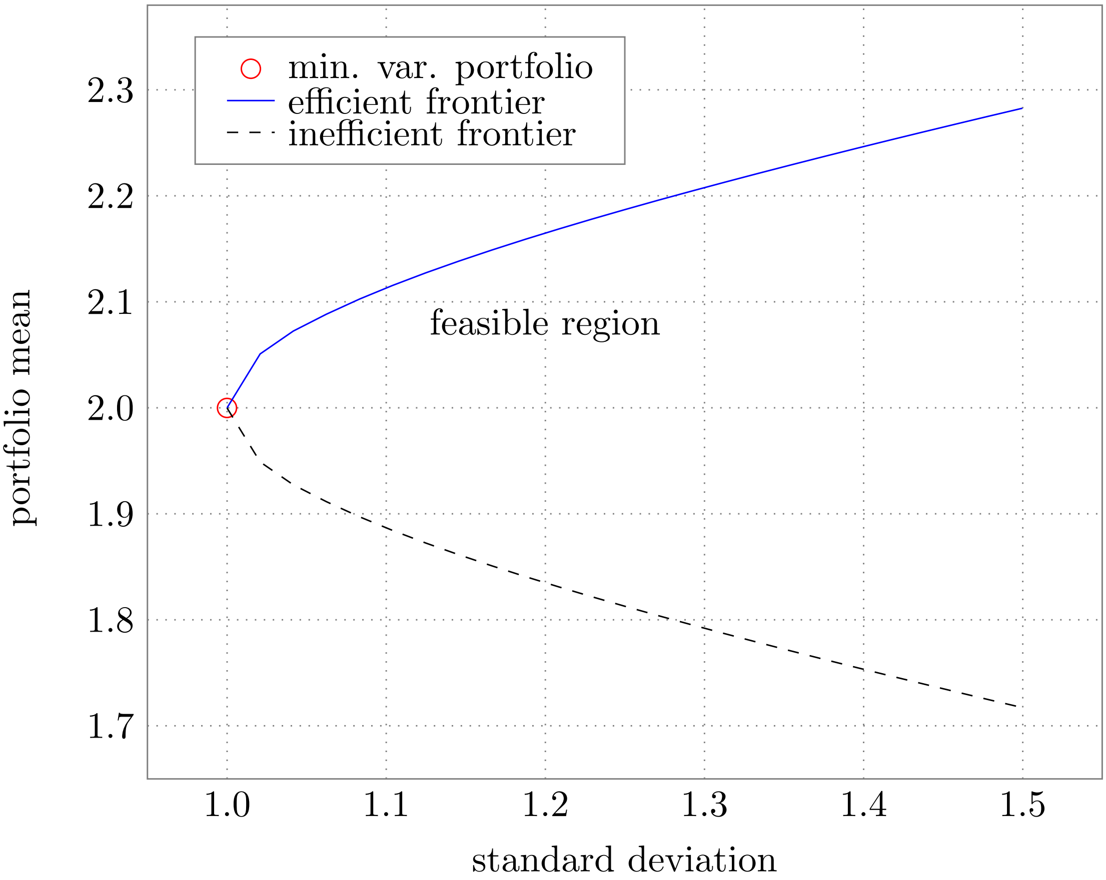

Figure 6.1: Markowitz efficient frontier, in terms of portfolio mean versus portfolio standard deviation (a hyperbola).

6.6 Global minimum mean–variance portfolio

If we minimise σ_p^2 in (6.15) as a function of μ_p we obtain $\mu_p = B/C$ and $\sigma_p^2 = 1/C$. The corresponding λ in (6.14) satisfies $\lambda = 0$.

Therefore, $\hat{\boldsymbol{\pi}}_1$ is the *global minimum variance portfolio.* That is, $\hat{\boldsymbol{\pi}}_1$ has the smallest possible variance among all portfolios $\hat{\boldsymbol{\pi}}$ that satisfy $\hat{\boldsymbol{\pi}}^\top \mathbf{1} = 1$.

6.7 Mean–variance optimization with risk-free asset

We now assume that a risk-free asset is available, that is, we do NOT assume any more that $\mathbf{1}^T \hat{\boldsymbol{\pi}} = 1$ necessarily.

To compute the mean-variance frontier we need to solve the following optimisation problem, for fixed μ_p:

$$\min_{\hat{\boldsymbol{\pi}}} \frac{1}{2}\hat{\boldsymbol{\pi}}^\top \bar{\bar{\Sigma}} \hat{\boldsymbol{\pi}} \tag{6.17}$$

subject to

$$\hat{\boldsymbol{\pi}}^\top \boldsymbol{\mu} + (1 - \mathbf{1}^\top \hat{\boldsymbol{\pi}}) R^0 = \mu_p \tag{6.18}$$

We again consider the Lagrangian:

$$\mathcal{L}(\hat{\boldsymbol{\pi}}, \delta) = \frac{1}{2}\hat{\boldsymbol{\pi}}^\top \bar{\bar{\Sigma}} \hat{\boldsymbol{\pi}} - \delta(\hat{\boldsymbol{\pi}}^\top \boldsymbol{\mu} + (1 - \mathbf{1}^\top \hat{\boldsymbol{\pi}}) R^0 - \mu_p)$$

This yields the first-order condition

$$0 = \nabla_{\boldsymbol{\pi}} \mathcal{L}(\hat{\boldsymbol{\pi}}, \delta) = \bar{\bar{\Sigma}} \hat{\boldsymbol{\pi}} - \delta(\boldsymbol{\mu} - R^0 \mathbf{1}). \tag{6.19}$$

Exercise. Show that we obtain the frontier portfolios

$$\hat{\boldsymbol{\pi}} = \frac{(\mu_p - R^0)}{(\boldsymbol{\mu} - R^0\mathbf{1})^T \bar{\bar{\Sigma}}^{-1} (\boldsymbol{\mu} - R^0\mathbf{1})} \bar{\bar{\Sigma}}^{-1} (\boldsymbol{\mu} - R^0\mathbf{1}). \tag{6.20}$$

Moreover, check that we have the following.

$$\sqrt{\hat{\boldsymbol{\pi}}^T \bar{\bar{\Sigma}} \hat{\boldsymbol{\pi}}} = \frac{|\mu_p - R^0|}{\sqrt{(\boldsymbol{\mu} - R^0\mathbf{1})^T \bar{\bar{\Sigma}}^{-1} (\boldsymbol{\mu} - R^0\mathbf{1})}}.$$

Definition 6.3. The **Sharpe ratio** is the ratio of risk premium to standard deviation:

$$\frac{\mu_p - R^0}{\sqrt{\hat{\boldsymbol{\pi}}^T \bar{\bar{\Sigma}} \hat{\boldsymbol{\pi}}}}$$

Remark 6.2. Sharpe ratios are not always suitable to compare different investment opportunities. For example, one opportunity has a (deterministic return) of 100%, implying a Sharpe ration of infinity. Another investment opportunity might have a return of 100% or 200%, each with probability 1/2. The second opportunity is clearly much better than the first one, but has a finite Sharpe ratio.

Any frontier portfolio has Sharpe ratio

$$\pm\sqrt{(\boldsymbol{\mu} - R^0\mathbf{1})^T \bar{\bar{\Sigma}}^{-1} (\boldsymbol{\mu} - R^0\mathbf{1})}.$$

There are two types of frontier portfolios:

- Mean-variance "*efficient*" portfolios are those with positive risk premia $\mu_p \geq R^0$.
- Mean-variance *inefficient*" portfolios are those with negative risk premia $\mu_p < R^0$.

There are three possible cases, with the notation of (6.13):

- $R^0 < B/C$;
- $R^0 > B/C$;
- $R^0 = B/C$.

6.7.1 Tangency portfolio

Assume $\mathbf{1}^\top \bar{\bar{\Sigma}}^{-1}(\boldsymbol{\mu} - R^0\mathbf{1}) \neq 0$, and define

$$\hat{\boldsymbol{\pi}}_* = \frac{1}{\mathbf{1}^T \bar{\bar{\Sigma}}^{-1}(\boldsymbol{\mu} - R^0\mathbf{1})} \bar{\bar{\Sigma}}^{-1}(\boldsymbol{\mu} - R^0\mathbf{1}). \tag{6.21}$$

Then, obviously $\mathbf{1}^\top \pi_* = 1$.

Note that $\hat{\boldsymbol{\pi}}_*$ is of the form (6.20) with μ_* satisfying the following equality:

$$\frac{\mu_* - R^0}{(\boldsymbol{\mu} - R^0\mathbf{1})^T \bar{\bar{\Sigma}}^{-1}(\boldsymbol{\mu} - R^0\mathbf{1})} = \frac{1}{\mathbf{1}^T \bar{\bar{\Sigma}}^{-1}(\boldsymbol{\mu} - R^0\mathbf{1})}. \tag{6.22}$$

Thus, $\hat{\boldsymbol{\pi}}_*$ is on the mean-variance frontier which involves the risk-free asset (a cone formed by two rays starting at $(0, R^0)$. It is also on the mean-variance frontier without the risk-free asset. Moreover, it can be seen that the tangency portfolio is on the efficient part of the frontier (that is, it has positive risk premium)if and only if $\mathbf{1}^T \bar{\bar{\Sigma}}^{-1}(\boldsymbol{\mu} - R^0\mathbf{1}) > 0$: indeed, this is deduced (6.22) and the fact that the variance-covariance matrix is positive definite.

Exercise. Show that $\mathbf{1}^\top \bar{\bar{\Sigma}}^{-1}(\boldsymbol{\mu} - R^0\mathbf{1}) \neq 0$ if and only if $R^0 \neq B/C$, with the notation of (6.13).

6.8 Two-fund spanning revisited

Assume risk-free asset and tangency portfolio $\hat{\boldsymbol{\pi}}_*$ exist. Then any portfolio on the mean-variance frontier can be constructed by combining the risk-free asset and the tangency portfolio $\hat{\boldsymbol{\pi}}_*$.

Exercise. How would Figure 6.2 look in the case $R^0 > B/C$? How in the case $R^0 = B/C$! (Recall lecture)

<u>Observation:</u> Recall the two portfolios $\hat{\boldsymbol{\pi}}_{\bar{\mu}}$ and $\hat{\boldsymbol{\pi}}_{\mathbf{1}}$ from (6.8) and (6.9). Now, note that

$$\hat{\boldsymbol{\pi}}_* = c_1 \hat{\boldsymbol{\pi}}_\mu + c_2 \hat{\boldsymbol{\pi}}_1$$

for some constants c_1 and c_2 which are independent of μ_p. Thus, the portfolio $\hat{\boldsymbol{\pi}}_*$ combines these two portfolios. Therefore, if there is a risk-free asset, then any investor who invests on the mean-variance efficient frontier chooses a combination of a position in the risk-free asset and a position in the portfolio $\hat{\boldsymbol{\pi}}_*$, where $\hat{\boldsymbol{\pi}}_*$ itself is a fixed combination of $\hat{\boldsymbol{\pi}}_{\bar{\mu}}$ and $\hat{\boldsymbol{\pi}}_{\mathbf{1}}$. This is sometimes called **one-fund theorem**.

6.9 Extension to other risk measures

We highlighted previously some limitations of standard deviation as a risk measure. We can then repeat a similar analysis than before using alternative risk measures. This will be explored during the practical part. In any case, we can make the following claim: If a risk measure is convex, then there is an efficient frontier.

Despite the previous fact, the shape of the attainable mean-risk space when considering only risky assets is not guaranteed to be a hyperbola.

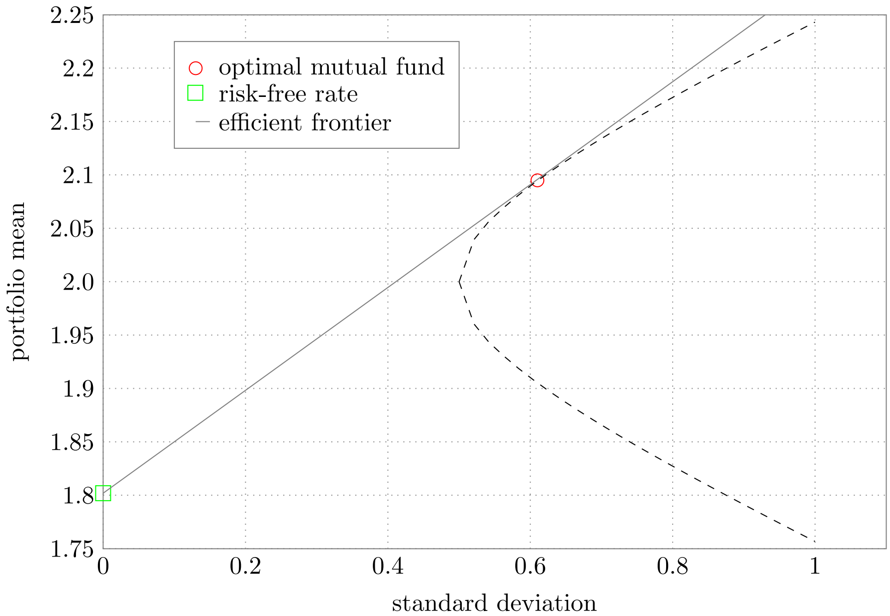

Figure 6.2: Efficient frontier with riskless asset in the case $R^0 < B/C$. The efficient frontier is sometimes also called the *capital market line*.

6.10 Risk allocation and performance comparison

We have explained how to compare the performance of different investment alternatives when considering them individually, by calculating their own mean and capital (risk measure). However, it is also useful to be able to evaluate the performance of each alternative as a *component of a portfolio*. Because the expectation is linear, this means that we need to understand the risk contribution of each position as a part of the portfolio: this is usually not equal to its individual risk, as we have joint evolution of investments, and the risk of each part is determined in conjunction with the other investments.

The determination of what is the risk contribution of each element as part of the portfolio is known as *risk (or capital) allocation*. Apart from the performance comparison, here are two examples where this might be useful:

- Loan pricing: The investor is a loan manager with a portfolio of n loans. She wants to understand how much capital should be allocated to each loan.
- Financial investment: An individual or institutional investor with a portfolio of n assets. He tries to establish what is the riskiness of each one considered as part of the portfolio.

Let us introduce some performance measures that generalise the Sharpe ratio.

6.10.1 RORAC

We can also define an indicator that plays an important role when using more general risk measures. We assume that there is no risk-free asset.

Definition 6.4. The RORAC (returns of risk-adjusted capital) of a position $\boldsymbol{\phi}$ (that represents the amount invested in each position) with respect to the risk-measure function r_ρ is given by

$$\mathcal{R}^\rho(\boldsymbol{\phi}) = \frac{\text{Expected profit of portfolio } \boldsymbol{\phi}}{\text{Risk capital of portfolio } \hat{\boldsymbol{\phi}}} = \frac{\mathbb{E}[\boldsymbol{\phi}^\top(\boldsymbol{R} - \mathbf{1})]}{(\rho(\boldsymbol{\phi}^\top \boldsymbol{R}))^+},$$

The RORAC is usually applied as a performance measure amongst investments, as it takes into account both the returns and the risk associated to them.

Note that the Sharpe ratio can be seen as the result of shifting the numerator of RORAC and taking as risk measure the standard deviation: the shifting comes from the fact that RORAC looks at profits and not excess of return (as we have assumed there is no risk-free asset).

6.10.2 Risk allocation

Let us consider a monetary and positive homogeneous risk measure ρ, the total amount of risk capital associated is $\kappa = \rho(R)$, and without loss of generality, assume an agent has invested in n different investing possibilities with gross return represented by $R_1, \ldots, R_n$ with weights π_0^i, that is its total return is

$$R = \sum_{i=1}^{n} \pi_0^i R_i.$$

Definition 6.5. A risk allocation is a set of weights $\kappa_1, \ldots, \kappa_n$ such that κ_i can be understood as the risk contribution of the asset i, and such that

$$\kappa = \sum_{i=1}^{n} \pi_0^i \kappa_i.$$

Here we focus on a per-unit capital allocation. Let $\Pi \subset \mathbb{R}_+^d \setminus \{\mathbf{0}\}$ represent the set of possible positive positions in the available alternatives. For a $\boldsymbol{\pi} \in \Pi$, the associated return is given by

$$R(\boldsymbol{\pi}) := \sum_{i=1}^{d} \pi_i R_i.$$

Note that the loss of the original portfolio is simply $R(\boldsymbol{\pi}_0)$.

Let us define a *risk-measure function* $r_\rho : \Pi \to \mathbb{R}$ by

$$r_\rho(\boldsymbol{\pi}) := \rho(R(\boldsymbol{\pi})).$$

We can interpret r_ρ as the required risk capital for a position $\boldsymbol{\pi}$ in the set of possible investments.

Definition 6.6. A per unit capital allocation principle associated with r_ρ is a mapping λ^{r_ρ} : $\Pi \to \mathbb{R}^n$ such that

$$\sum_{i=1}^{n} \pi_i \lambda_i^{r_\rho}(\boldsymbol{\pi}) = r_\rho(\boldsymbol{\pi}); \qquad \text{for all } \boldsymbol{\pi} \in \Pi.$$

In other words, a unit capital allocation is a rule that assigns a part of the total risk capital to each unit of R_i, when the overall position is $R(\boldsymbol{\pi})$.

An application of a result by Euler shows that if r_ρ is positive homogeneous and differentiable at $\boldsymbol{\pi}$ with non-negative entries (not all zero), then

$$r_\rho(\boldsymbol{\pi}) = \sum_{i=1}^{n} \pi_i \frac{\partial r_\rho}{\partial \pi_i}(\boldsymbol{\pi})$$

This suggests the following rule

Definition 6.7 (Euler capital allocation)**.** If r_ρ is a positive homogeneous risk measure function on the set Π, then the per-unit Euler allocation is associated to r_ρ is defined by

$$\pi^{r_\rho} : \Pi \to \mathbb{R}^n, \quad \lambda_i^{r_\rho}(\boldsymbol{\pi}) = \frac{\partial r_\rho}{\partial \pi_i}(\boldsymbol{\pi})$$

Example 6.8. *We can calculate the Euler capital allocation for the most popular risk measures we have presented before. In the case of value at risk and expected shortfall, we assume that the loss random variables are continuous with a regular joint density. In this case, the allocation for the asset i is*

sd: $\lambda_i^{\mathrm{sd}}(\boldsymbol{\pi}) = \frac{\mathrm{cov}(R_i, R(\boldsymbol{\pi}))}{\sqrt{\mathrm{var}(R(\boldsymbol{\pi}))}}$
V@R$^\alpha$**:** $\lambda_i^{\mathrm{V@R}^\alpha}(\boldsymbol{\pi}) = \mathbb{E}[-R_i | -R(\boldsymbol{\pi}) = \mathrm{V@R}^\alpha(\mathrm{R}(\boldsymbol{\pi}))]$
ES$^\alpha$**:** $\lambda_i^{\mathrm{ES}^\alpha}(\boldsymbol{\pi}) = \mathbb{E}[-R_i | -R(\boldsymbol{\pi}) \geq \mathrm{V@R}^\alpha(\mathrm{R}(\boldsymbol{\pi}))]$

See the exercises for a justification.

If the capital allocation is to be used for performance comparison, we expect that if we increment a position in an asset with better RORAC than the one of the whole portfolio, then the overall RORAC should increase. Let us formalise this property.

Definition 6.9. Let r_ρ be differentiable for all $\boldsymbol{\pi} \in \Pi$. We say that λ^{r_ρ} is *suitable for performance measurement* if for all $\pi \in \Pi$,

$$\frac{\partial \mathcal{R}^{r_\rho}(\boldsymbol{\pi})}{\partial \pi_i} \begin{cases} > 0 & \text{if } \frac{\mathbb{E}[R_i - 1]}{\lambda_i^{r_\rho}(\boldsymbol{\pi})} > \mathcal{R}^{r_\rho}(\boldsymbol{\pi}) \\ < 0 & \text{if } \frac{\mathbb{E}[R_i - 1]}{\lambda_i^{r_\rho}(\boldsymbol{\pi})} < \mathcal{R}^{r_\rho}(\boldsymbol{\pi}) \end{cases}$$

Tasche (1999) proposed the previous definition and proved that under the assumptions of Definition 6.9, the only per-unit capital allocation principle suitable for performance measurement is the Euler principle. Thus, the Euler principle can be used to allocate capital for performance comparison purposes.

6.11 Exercises

Exercise 6.1. Considering the case without risk-free asset, show that any two different portfolios $\bar{\pi}_a$ and $\bar{\pi}_b$ on the mean-variance frontier can serve as mutual funds for spanning the whole frontier: that is, show that any portfolio on the mean-variance frontier can be written as the combination of $\bar{\pi}_a$ and $\bar{\pi}_b$.

Exercise 6.2. Considering only risky assets, find the global minimum variance portfolio directly by solving the minimisation problem

$$\min_{\hat{\boldsymbol{\pi}}\in\mathbb{R}^d} \boldsymbol{\pi}^\top \bar{\bar{\Sigma}}\boldsymbol{\pi}; \quad \text{s.t. } \boldsymbol{\pi}\cdot\mathbf{1} = 1$$

Exercise 6.3. Suppose that the risk-free return is equal to the expected return of the global risky minimum variance portfolio (see the previous question). Show that there is no tangency portfolio.

Exercise 6.4. Suppose that the law of one price holds (see exercise 1.6), and that the assumptions of this chapter hold (finite assets, with finite variance and invertible variance-covariance matrix).

- Show there is a unique (not necessarily positive) SDF in the span of the asset **pay-offs**. Call it M^p
- Write the associated price for this pay-off (R_p) and its return.
- Assume that there is a risk free-asset. Show that this portfolio is inefficient.

Exercise 6.5. Let $\bar{\bar{\Sigma}}$ denote the covariance matrix of $\bar{R} = (R_1, \cdots, R_n)$, and $\hat{\mu}$ its mean. Consider the *standard deviation risk measure* $\rho(R) = \text{STD}(R) = \sqrt{\mathbf{1}^\top\bar{\bar{\Sigma}}\mathbf{1}}$.

1. Compute the optimal capital allocation $(C_1, \cdots, C_n)$ according to the Euler principle.
2. Check that indeed $\rho(R(\pi)) = \sum_{i=1}^n \pi_i C_i$.
3. Calculate the RORAC at the optimal. Calculate the gradient of the RORAC with respect to π_i for each i at the optimal capital allocation, and verify that this allocation is suitable for performance measurement.

Exercise 6.6. Let R_1 and R_2 be two independent continuous random variables in $\mathbb{R}$, and let f_1, F_1 and f_2, F_2 be their respective pdf and cdf functions.

As before, let $R(\lambda) := \lambda_1 R_1 + \lambda_2 R_2$, for $\lambda_1, \lambda_2 \neq 0$. Since R_1, R_2 are continuous, $R(\lambda)$ is also continuous. Let us denote by F^λ, f^λ its cdf and pdf respectively.

1. Show that

$$F^\lambda(x) = \mathbb{E}\left[F_2\left(\frac{1}{\lambda_2}(x - \lambda_1 R_1)\right)\right] = \mathbb{E}\left[F_1\left(\frac{1}{\lambda_1}(x - \lambda_2 R_2)\right)\right]$$

   and calculate f^λ.

2. Recall that for a couple of continuous random variables X,Y with positive density f_X, f_Y we can define the conditional expectation $\mathbb{E}[X|Y = y]$ to be given by

$$E[X|Y = y] = \int_{-\infty}^{\infty} x f_{X|Y}(x|y)dx \tag{6.23}$$

   where

$$f_{X|Y}(x, y) = \frac{f_{X,Y}(x, y)}{f_y(y)}.$$

   Show that $\mathbb{E}[R_1|R(\lambda) = x] = \frac{1}{f^\lambda(x)}\mathbb{E}\left[\frac{1}{\lambda_2}R_1 f_2\left(\frac{1}{\lambda_2}(x - \lambda_1 R_1)\right)\right]$.

3. Euler allocation of value at risk in this context: Given that $R(\lambda)$ is continuous, we have that $F^{\lambda}(-\mathrm{V@R}^{\alpha}(R(\lambda))) = \alpha$. Use the chain rule and the results in a. and b. to show that the Euler allocation of value at risk is

$$\pi_i^{\mathrm{Euler}}(R(\lambda)) = \mathbb{E}[-R_i| - R(\lambda) = \mathrm{V@R}^{\alpha}(R(\lambda))]$$

6.12 Summary

- Investors face a tension between returns and risks.
- Within the compromise of this tension, it is possible to construct a Pareto set (or efficient frontier). That is, all portfolios that are not dominated by another portfolio (i.e. a portfolio that is better in every feature).
- The efficient frontier can be explicitly be found in the case of mean-standard deviation: it forms a hyperbola in the space of return vs. standard deviation of returns.

Chapter 7

A brief overview on asset types (*)

Up to this point, we have presented a theoretical framework that allows us to measure risk, performance, or construct portfolios designed to answer to the needs of a given investor.

There is, however, an important gap to be able to apply the methods we have developed: we need to be able to estimate the (joint distribution) of random variables representing the asset value S^i or its return R^i.

In this short chapter we take the first step in this direction: we include here a brief review of the main asset classes in financial markets. This will help us understand some features of asset classes in practice, and to clarify the use in practice of some of the concepts we have presented.

7.1 Equity

Equity denotes in general investment on companies either private or public. Since we focus on market risk and portfolio theory, we consider only public equity (i.e., shares traded in a stock exchange).

In the public market system, shareholders have a limited responsibility over the losses that a company can cumulate. As a consequence, equity prices are **non-negative**. It is then common to model directly their **log-returns**.

Remark 7.1. Note that although share prices cannot be negative, they can reach zero.

Recall that in our modelling we need to consider the *total value of assets*. In the case of equities, this total value is driven by two factors: *the price of the asset in the market*, and the *dividend payments*. Therefore, assume that P is the stochastic process representing the price of the asset and D is the process of dividend payments. Suppose also that the number of shares is kept fixed (more on this below). Then, we can set

$$R_t := \frac{P_t + D_t}{P_{t-1}}; \qquad t > 0$$

and equivalently,

$$S_0 = P_0; \qquad S_{t+1} = S_{t-1}\frac{P_t + D_t}{P_{t-1}}.$$

Whenever the number of shares is not constant throughout the period, we need to adjust the above equation. In particular, suppose that the ratio of shares from time $t-1$ to time t is the process $\phi_t = \frac{N_t}{N_{t-1}}$. We then have

$$R_t := \phi_t \frac{P_t + D_t}{P_{t-1}}; \qquad t > 0$$

and equivalently,

$$S_0 = P_0; \qquad S_t = S_{t-1}\phi_t \frac{P_t + D_t}{P_{t-1}}.$$

To understand the above relation, note that $N_t P_t$ returns the market capitalisation of the company (that is, the total value of the company on the market). Similarly, $N_t D_t$ is the total dividend paid. Hence, this is equivalent to calculating the total return of the whole company, and then supposing that any dividends are re-invested.

Remark 7.2. In many cases, there are different types of shares for a given company: for example, in some cases dividends are first paid to some assets (like preferred assets) and then to some *common* shares. Likewise, there are shares with no voting rights. More complex situations as these might require custom adjustments, depending on the model purpose.

7.2 Fixed income

Fixed income assets owe their name to the fact that the interest paid on investments are fixed in advance. Compare by opposition with equity as explained before, where dividends are uncertain.

Fixed income assets include *bonds*, *certificates of deposit*, among others. We focus here on *bonds*, as they are the most common type of instruments. A *bond* is a debt obligation: the issuer borrows money from investors and pays a **fixed** interest. The *maturity* of a bond is the date at which the *principal* (and any remaining interest) are paid and the obligation ends.

Some bonds pay interest periodically (say annually). Each of these payments is a *coupon*. Other bonds all interest along the principal: these are called *zero-coupon* bonds.

Government and corporate bonds are the most common bond issuers. Investors can acquire the bond directly from issuers at the *face value* in what is called the *primary market*. Then, they can trade with them in the *secondary market*, where prices are set by the market.

Remark 7.3. Let us remark that as highlighted in the previous section, in the event of a default, fixed-income investors are paid before stockholders.

7.2.1 Zero-coupon bonds

We start by looking at zero-coupon bonds. In what follows, we assume zero-coupon bonds pay one unit at maturity. Note there is no loss of generality, since we can then simply multiply prices by the total value.

Without credit risk

Let us first consider the case *without credit risk.* In an arbitrage-free market model, we find from our pricing equations that a bond with maturity T has price

$$P_{0,T} = \mathbb{E}[M_T 1] = \mathbb{E}[M_T] = \mathbb{E}^{\mathbb{Q}}\left[\frac{1}{S_T^0}\right]$$

The price of a zero-coupon bond with maturity T at a future time t is $\mathcal{F}_t$ measurable and equals

$$P_{t,T} = \mathbb{E}_t\left[\frac{M_T}{M_t}\right] = \mathbb{E}_t^{\mathbb{Q}}\left[\frac{S_t^0}{S_T^0}\right]$$

Including credit risk

Now, let us consider the case *with credit risk.* Let τ be a random variable denoting the time of default of the issuer. We get

$$P_{0,T} = \mathbb{E}[M_T \mathbb{1}_{\tau > T}]. \tag{7.1}$$

In many contexts, the random variable τ is assumed to be an exponential random variable. Practitioners then use equation (7.1) to determine the rate of decrease of this exponential variable "perceived by the market" and use it elsewhere to correct for credit risk.

Similarly, the price of a zero-coupon bond with maturity T at a future time t is $\mathcal{F}_t$ measurable and equals

$$P_{t,T} = \mathbb{E}_t\left[\frac{M_T}{M_t}\mathbb{1}_{\tau > T}\right].$$

7.2.2 Coupon paying bonds

Assume now that a bond without credit risk pays coupons x annually (and assume that the periodicity of the market model is also annual). We then have

$$P_{0,T} = x\sum_{i=1}^{T}\mathbb{E}[M_i] + \mathbb{E}[M_T].$$

The other cases are adjusted similarly.

7.2.3 Yields

The bond's *yield to maturity* $y_{0,T}$ is the per-period (typically annualised) rate of return if the bond is held to maturity, that is,

$$P_{0,T} = \frac{1}{(1+y_{0,T})^T}$$

in the case of a zero-coupon bond, while in the case of coupon paying bonds we get

$$P_{0,T} = x \sum_{i=1}^{T} \frac{1}{(1+y_{0,i})^i} + \frac{1}{(1+y_{0,T})^T}.$$

Frequently market information providers provide information on yields rather than bond prices, since this allows to use information on bonds regardless of how coupons are paid. Note, though, that the above equation involves different yields, and as such a system of equations need to be solved. This process is typically called *curve stripping*.

Remark 7.4. The yields defined above corresponds to an annualised compound discrete rate. There are other *conventions* corresponding to different ways of calculating interest rates. For example, a continuously compounded yield would solve

$$P_{0,T} = e^{-y(0,T)T}.$$

7.2.4 Forward rates

We call *forward rate* to the rate implied by market prices of instruments like zero-coupon bonds to be valid in the future. In terms of zero-coupon bonds, we get

$$F_{t-1} = \frac{P_{0,t-1}}{P_{0,t}} - 1.$$

Indeed, both strategies below bring a payoff 1 at t:

- Buy a zero-coupon with maturity t
- Buy $\frac{P_{0,t}}{P_{0,t-1}}$ zero-coupon bonds with maturity $t-1$ and then invest with a rate of return given by the forward rate.

7.3 Foreign Exchange (FX)

We have assumed that all assets are denoted in a unique *numeraire*. In practice, several currencies are used. The terms of their exchange are called *exchange rates*.

In many markets, exchange are *floating*, that is, they are determined by offer and demand. Instruments like *forward exchanges*, under which future exchange rates are fixed are thus needed.

An important observation is that when exchange rates are *floating* a money market account in one currency might look risky in the other.

7.3.1 Forward exchange rate

Consider two currencies. We are going to deduce an expression for the forward exchange rate in terms of the prices of zero-coupon bonds with no credit risk.

- Consider two currencies: the base currency and the alternative one. Denote by a $\star$ superindex the prices in the latter.

- Let x denote the exchange rate (that is, the number of units of the base currency to buy one unit of the alternative currency) and call the forward exchange rate F_t^x with the same convention.

Consider the following two investments:

- Buy F_t^x units of the zero-coupon bond with maturity t in the base currency. This means that at time t there is precisely F_t^x units of the base currency, which is the amount needed to get 1 unit of the alternative currency (since the rate was fixed). This has a cost of $F_t^x P(0,t)$ in the basis currency.
- Buy 1 unit of a zero-coupon bond with maturity t in the alternative currency. This has a price of $xP^*(0,t)$ in the basis currency.

By no arbitrage, both costs should be the same. Hence,

$$P^*_{0,t} = P_{0,t}\frac{F_t^x}{x}; \quad \Rightarrow \quad F_t^x = x\frac{P^*_{0,t}}{P_{0,t}}.$$

7.4 Derivatives

It is possible to define derivatives on most of the asset classes defined before, including futures, options, and swaps.

A few of these derivatives have a market of themselves (for example, call options on some stocks) that trade on some reference every day (for some strikes and maturities), but the majority are traded OTC. Whenever they are traded in the market, they are treated as before.

7.5 Commodities and other asset classes

Here we include some asset classes that require some specialised modelling that goes beyond what we can include in this short summary.

These include:

- Commodities, which contains basic goods that are exchangeable for other of the same type. Typical examples are gold, oil, natural gas, coal, grains, among others.
- Digital currencies (like Bitcoin and Ethereum).
- Wines and art brokerage
- Real-state
- Non-fungible tokens (NFT's) digitally signed digital artworks.

Chapter 8

Statistics review

In this chapter we examine some tools in statistics that will prove useful when applying in practice the techniques we have studied before.

8.1 Hypothesis testing

In hypothesis testing, we use statistical techniques to answer a question of interest based on observed data. To do so, we formulate a probabilistic assumption that we put to challenge with the help of statistical techniques applied on an observed sample.

The general template for hypothesis testing is the following:

1. State the *null hypotheses* H_0 and additional statistical assumptions on the sample.
2. Choose a statistic and the test parameters
3. Compute the test statistic and its p-value
4. Decide whether to reject the null assumption or not.

8.1.1 State the null hypothesis H_0 and additonal assumptions

In this step, we turn a question of practical interest into a probabilistic question.

Example 8.1. *We would like to test whether the net return of a given investment has on average a 2% mean-return per period. The observed mean return is* 1.8% *while the observed unbiased sample standard deviation is 1% from a 36 sample size.*

By calling X *the random variable denoting the returns in a given period, we would like to test the claim* $\mathbb{E}[X] = 2\%$.

$$(H_0)\text{: The random variable } X \text{ has mean } \theta_0 = 2\%.$$

However, we need to add some additional statistical assumptions in order to be able to compute a result. For example:

- X *has finite (even if unknown variance).*

- *The available data $X_1, \ldots X_n$ are i.i.d*

Remark 8.1. The hypothesis test is based not only on the null assumption but also on the statistical assumptions. Hence, for example, a failure in the test might be caused by a failure in one of the additional assumptions rather than the one we want to test.

8.1.2 Choose statistics and test parameters

In this step, we select a test statistic based on the null assumption we would like to test, and the additional statistical assumptions made.

Here are some examples of test statistics:

- **Wald test**: This test compares directly the values of the estimated parameter and the parameter in the null-assumption, modulo a normalisation with the observed parameter volatility.
  Denoting $\hat{\theta}$ the *observed* value of the parameter in the test, we get

  $$Z_{wald} = \frac{\hat{\theta} - \theta_0}{\mathrm{sd}(\hat{\theta})}$$

  Under the assumption that the parameter has finite variance, and that the samples are i.i.d., the law of large numbers this statistic is asymptotically Gaussian when H_0 is true. Moreover, we can have sharper results if the estimated parameter is known a priori to follow a Gaussian distribution; in this case the estimator follows a t-distribution with $n-1$ degrees of freedom. This is useful for tests with small values.
  Sometimes, it is more convenient to use the related test

  $$S_{wald} = \frac{(\hat{\theta} - \theta_0)^2}{\mathrm{var}(\hat{\theta})}$$

  which is asymptotically χ_1^2.
- **Score test** This test uses a scoring function based on the log likelihood function of the parameter given the data, so that we do not need to directly estimate the desired parameter. Let $f(x|\theta)$ be the probability density function of the data assuming the parameter is θ. Recall that we define the likelihood function $\mathcal{L}$ so that

  $$\mathcal{L}(\theta|x) = f(x|\theta).$$

  With this notation, the score test is defined by

  $$S_{score} = \left. \frac{(\partial_\theta \log \mathcal{L}(\theta|X_1, \ldots, X_n))^2}{-\mathbb{E}[\partial_\theta^2 f(X_1, \ldots, X_n|\theta)|\theta]} \right|_{\theta=\theta_0} ;$$

  it is asymptotically χ_1^2 when H_0 is true.
- **Likelihood ratio test**: In the case when the null assumption is of the type $H_0 : \theta \in \Theta_0$, we can define a statistic of the form

  $$\lambda_{LT} = 2 \log \left(\frac{\sup_{\theta \in \Theta} \mathcal{L}(\theta)}{\sup_{\theta \in \Theta_0} \mathcal{L}(\theta)} \right).$$

  In the case where Θ_0 is of the form 'the last ℓ entries are fixed', the statistic is asymptotically χ_ℓ^2.

- **χ-test for multinomial**:
  A particular case of the above that is useful for multinomial distribution is the $\chi-$test. It uses the test statistic
$$\sum_{j=1}^{k} \frac{(X_k - np_{0j})^2}{np_{0j}}$$
  which is asymptotically χ^2_{k-1}. This test can also be adapted for goodness-of-fit and independence.

Choosing test parameters

When performing the test, we encounter a confusion matrix that relates the outcome of the test with the actual truth value of the null hypothesis as shown in table 8.1: we can be right (rejecting when it is false and not rejecting when it is true) or wrong, or we can be wrong in two ways. Rejecting when it was true or not rejecting when it was false.

		Decision	
		Retain null	Reject null
Truth in population	True	Correct: $(1-\alpha)$	Type I error: α
	False	Type II error: β	Correct (power) : $(1-\beta)$

Table 8.1: Confusion matrix relating outcome of test and truth of H_0

Unless an alternative is considered, hypothesis testing places its focus on obtaining evidence to reject the null assumption, not on obtaining evidence to support it. As seen before, it is frequently the case that we know the distributional properties of the test statistics given the null assumption. We can then use this information to estimate the value of the Type I error. We set a priori a value of acceptable Type I error that determines our choice of rejecting or not. This is call the level of the test and it is usually denoted by α. Typical values are 5% and 1%. This level can be informally interpreted as the likelihood of having observed a value of the test even further away from the ideal given that the null assumption was true.

If an alternative assumption is available, we can also aim to reduce the type II error by choosing the number of samples and by including choosing the test-statistic.

Example 8.2 (Continuation of Example 8.1)**.** *In the given setting we have not made an assumption on the distribution of the data. This limits the possibility to apply tests based on the likelihood function. We then choose to use the Wald test, and we fix the level of the test to be 5%.*

8.1.3 Compute test statistic and p-value

Once we have chosen the test statistic and the level of the test, we can compute the observed value of the test statistic and its corresponding *p-value*. The p-value is the probability of observing results even more extreme than the one we observed. This can be done manually using statistical tables of distributions or with the help of statistical software like the packages *Statsmodel* or *pengouin* in Python.

Example 8.3 (Continuation of Example 8.2). *In the given setting we get:*

$$Z_{wald} = \frac{0.018 - 0.02}{\frac{0.01}{\sqrt{36}}} = -\frac{12}{10} = -1.2$$

By definition, given we are testing equality, the corresponding p-value is written

$$\mathbb{P}[|Z_{wald}| > 1.2] \approx 2(1 - \Phi^{-1}(1.2)) \approx 0.23.$$

8.1.4 Decide whether to reject the null assumption or not

We compare the p-value and the defined level of the test. We reject in those cases where the p-value is smaller than the defined level.

Note that in some cases[1], it might be easier not to calculate a p-value, but rather to estimate a cut-off level of acceptance or rejection from the distribution of the estimator and the chosen level of the test.

Example 8.4 (Continuation of Example 8.3). *In the given setting we compare the p-value of 0.23 with the level of the test 5% and conclude that we do not reject the null assumption.*

8.2 Linear regression

Regression is a method to study the relation between a variable of interest or response variable (say $Y \in \mathbb{R}^m$) and some covariate factors (say $F \in \mathbb{R}^n$). In linear regression one proposes a linear model for this relationship, that is

$$Y = \alpha + \beta F + \epsilon,$$

where $\alpha \in \mathbb{R}^m$ and $\beta \in \mathbb{R}^{m \times n}$. Naturally, we are interested in *the best* possible model of this type. In the Ordinary Least Squares (OLS) approach, this is understood as choosing coefficients α, β so that $\mathbb{E}(|\epsilon|^2)$ is minimal.

For the purpose of finding the optimal coefficients for the model, we can rewrite the problem as

$$Y = \tilde{\beta}\tilde{F}$$

where

$$\tilde{F} = \begin{pmatrix} 1 \\ F \end{pmatrix}; \qquad \tilde{\beta} = \begin{pmatrix} \alpha & \boldsymbol{\beta} \end{pmatrix}$$

with the property that $\mathbb{E}[Y|\tilde{\boldsymbol{F}}] = 0$. Hence, without loss of generality, we can avoid using explicitly the intercept coefficient α to the cost of increasing the dimension of the problem and restricting to the case $\mathbb{E}[Y|\tilde{\boldsymbol{F}}] = 0$.

Alternatively, we can account for the intercept directly by noticing that we can set

$$\alpha = \bar{Y} - \beta\bar{F}$$

where $\bar{Y} = Y - \frac{1}{N}Y\mathbf{1}_N$, $\bar{F} = F - \frac{1}{N}F\mathbf{1}_N$ and β solves the linear regression (without intercept)

$$\bar{Y} = \beta\bar{F} + \epsilon.$$

[1]Particularly when tables of values are used instead of computer software

8.2.1 Coefficient estimation

In a practical setting, we want to obtain an estimator for the value of the coefficients from a sample (of size N) of the couples (Y, F). Let us then treat F and Y as matrices, $F \in \mathbb{R}^{n\times N}$ and $Y \in \mathbb{R}^{m\times N}$, containing i.i.d. samples of the original distribution.

Theorem 8.5. *Assume that $FF^\top$ is invertible. Then, the least-square estimator for the regression coefficient in*

$$Y = \hat{\beta}F + \epsilon \tag{8.1}$$

is given by

$$\hat{\beta} = YF^\top(FF^\top)^{-1}$$

Proof. Our goal is to choose $\hat{\beta}$ to minimise the ℓ^2 norm of the sample error, i.e.

$$\begin{aligned} &\min_{\beta\in\mathbb{R}^{m\times n}} |\epsilon|_2^2 \\ &s.t. \qquad Y = \hat{\beta}F + \epsilon. \end{aligned}$$

By replacing the constraint in the objective, this is equivalent to solving

$$\min_{\beta\in\mathbb{R}^{m\times n}} \operatorname{tr}\left[(Y - \hat{\beta}F)^\top(Y - \hat{\beta}F)\right],$$

where tr stands for the trace of the resulting matrix. Applying first order conditions, we get that the optimal must satisfy

$$-2YF^\top + 2\hat{\beta}(FF^\top) = 0$$

from where the claim follows. □

8.2.2 Generalized Least-Squares (GLS)

From a numerical point of view, we might encounter problems to obtain satisfactory coefficients whenever the matrix $FF^\top$ is ill-conditioned [2]. In this case, the coefficient $\hat{\beta}$ might not have good statistical properties for our system.

Statistically, the matrix $FF^\top$ will be ill-conditioned when the there is heteroscedasticity in the model and/or when errors are not independent

In order to improve the efficiency of the estimator, one can use a modified estimator

$$\hat{\beta} = Y\Sigma^{-1}F^\top(F\Sigma^{-1}F^\top)^{-1}$$

where $\Sigma \approx \operatorname{cov}(\epsilon|F)$. From a statistical point of view, this is equivalent to applying the linear system to a re-weighted group of factors, thus reducing the heteroscedasticity. Interestingly, from a numerical point of view, this can be seen as using a preconditioner to improve stability (to know more about preconditioning in numerical analysis see for example Chen (2005)).

[2] Recall that the condition number of a matrix A measures how an error in 'b' induces an error in 'x' when solving the linear system $Ax = b$.

8.3 Model selection

We discussed the problem of finding the coefficients that would produce the best linear model to express a dependent variable with respect to a given set of factors. Sometimes, however, the precise set of factors to consider is not completely defined: for example, we might start with a *larger* set of factors with the understanding that only a subset of those factors should actually be considered.

In theory, including more factors than needed does not affect the solution, since in principle the coefficient for those factors would then be zero. However, since we only obtain estimators from samples, it is possible that some of these factors may *appear* not to be zero even if they *are*. For this reason, it is recommended to test the hypothesis that the factors are in fact zero using the Wald test. Factors for which we cannot reject the null hypothesis that they are zero can be excluded from the list. A new regression can then be done over the smaller sample.

This procedure can be complemented by calculating a measure reflecting the tension between the number of parameters used and the goodness-of-fit of the model they produce. Two possible estimators are:

- AIC (Akaike Information Criterion): Minimise $|S| - \ell_S$ where ℓ_S is log-likelood at the MLE.
- BIC (Bayesian Information Criterion): Minimise $\frac{|S|}{2}\log(n) - \ell_S$

where S is the set of factors considered. We start from the largest possible set of factors. Then, one by one, the factor with the largest p-value in the Wald test is dropped until the above criterion stops decreasing.

Chapter 9

Factor models

The London Stock Exchange has currently around 2292 listed companies while the New York Stock Exchange has around 2400, Euronext Paris 1078, Shanghai Stock Exchange 1041. There are also thousands of standardised futures and options on equity, interest rates and commodities available in different exchanges like the Chicago Mercantile Exchange (CME). On top of all of these standard instruments, financial institutions produce millions of customised over the counter operations.

Modelling individually each one of those securities is a formidable task because that would require to understand the simultaneous joint law of all of them, and the types of distributions that would allow this easily (for example jointly Gaussian) might be limiting.

Such complexity can be reduced by the introduction of factor models. The objective of such models is to produce relevant predictions on the general trends of the valuation of these market instruments in terms of a few relevant variables. The benefits are twofold: for risk managers, this allows to focus on a reduced number of risk sources; while for investors, the reduction in dimensionality might is in practice a form of regularization that might improve the statistical reliability of estimations.

In this chapter we introduce formally such models and study some of its properties.

9.1 Factor Models

A factor model for a random variable Y is an expression of Y in terms of other random variables $F_1, \ldots, F_n$ called *the factors* such that

$$Y = h(F_1, \ldots, F_n) + \epsilon$$

where ϵ is a random *model error*. Evidently a good model will generate small model errors.

Factor Models can be obtained in two main ways: they are sometimes deduced from existing mathematical formulas, for example as a byproduct of a pricing formula. In this case, the factor model acts like a transformation that allows us to express a variable of interest in terms of variables to which we can have better access or or more historical informatior.

Alternatively, factor models can be "statistically" driven: this means that they are observed to hold with statistical significance from historical data analysis. In many cases, we can give, a posteriori, an economic interpretation to statistically driven factors. However, this is nor guaranteed or even, depending on the goals of the factor model, needed.

We can classify the types of factor that can be chosen in two: observable and unobservable factors. Observable factors are the ones for which we can obtain reliable values from sources known to the market: an index value, some economic indices (GDP, inflation,...). Unobservable factors are abstract and are indirectly deduced from the data.)

9.1.1 Linear models

The simplest and more common type of factor model occurs when h is affine.

We say that $Y \in \mathbb{R}$ has a linear model in terms of the factors $F_1, \ldots, F_\kappa \in \mathbb{R}$, if there exist $\beta_0, \ldots, \beta_\kappa \in \mathbb{R}$ such that

$$Y = \beta_0 + \beta_1 F_1 + \ldots + \beta_\kappa F_\kappa + \epsilon \tag{9.1}$$

where ϵ is a random variable in $\mathbb{R}$ with zero expectation and such that $\mathbb{E}[\epsilon|F_k] = 0$ for each k.

A typical objective for a linear model is to minimise the error term, or, more precisely, its variance. If we call $\bar{\bar{\Sigma}}$ the variance covariance matrix of $\boldsymbol{F} = (F_1, \ldots, F_n)^\top$, and we assume it is (left) invertible, we can show that

$$\begin{bmatrix} \beta_1 \\ \vdots \\ \beta_n \end{bmatrix} = \bar{\bar{\Sigma}}^{-1} \begin{bmatrix} \mathrm{cov(Y, F_1)} \\ \vdots \\ \mathrm{cov(Y, F_\kappa)} \end{bmatrix} \tag{9.2}$$

and by taking expectations, we have that

$$\beta_0 = \mathbb{E}[Y] - \sum_{i=1}^{\kappa} \beta_i \mathbb{E}[F_i] \tag{9.3}$$

If we are allowed to modify the factor variables, we can also "avoid" the β_0:

$$Y = \sum_{i=1}^{\kappa} \beta_i (F_i + \eta_i) + \epsilon$$

where we choose $\boldsymbol{\eta}$ in such a way that $\sum_{i=1}^{\kappa} \beta_i \eta_i = \beta_0$, and the expression for each β_i is kept constant. A possible solution when $\sum_{i>0} \beta_i \neq 0$ is to take all $\eta_i = \beta_0 / (\sum_{j>0} \beta_j)$.

9.2 Beta pricing

As we have highlighted before, one of the main asset pricing problems consists in finding a suitable relationship between the pay-offs or returns and the prices. An alternative (but evidently related) to finding the SDF directly is inspired in linear modelling regression.

If a linear model exists for the random excess market returns with respect to some factors, we would have a linear relationship of risk premium and expectation of the factors. Indeed, if

$\boldsymbol{F} = (F_1, \ldots, F_k)^\top$ is a k–dimensional random vector with invertible covariance matrix $\bar{\bar{\Sigma}}_f$, we say there is a *multifactor* beta pricing model if there exists $\boldsymbol{\lambda} \in \mathbb{R}^k$ such that

$$\mathbb{E}[R] - R^0 = \boldsymbol{\lambda}^\top \bar{\bar{\Sigma}}_F^{-1} \mathrm{cov}(\boldsymbol{F}, R).$$

We call $\boldsymbol{\lambda}$ the *factor risk premium* and $\bar{\bar{\Sigma}}_F^{-1} \mathrm{cov}(\boldsymbol{F}, R)$ the vector of betas for some return R.

Interpretation: λ^i is the extra expected return of an asset for each unit increase in its ith-*beta.*

A particular case of interest is the one of a single factor: we say there is a *single-factor* beta pricing model with factor f for some random variable f if there exists a constant λ such that

$$\mathbb{E}[R] - R^0 = \lambda \frac{\mathrm{Cov}(f, R)}{\mathrm{var}[f]}. \tag{9.4}$$

Example 9.1. *Recalling Exercise 1.3 (in the case of a one-period model), we can say that there exists a single-factor linear model for the excess of return with factor M (any SDF for the market).*

9.2.1 Single-factor models with returns as factors

Assume that $f = R_*$ is the return of a certain asset. Then, (9.4) translates into

$$\mathbb{E}[R] - R^0 = \lambda \frac{\mathrm{Cov}(R_*, R)}{\mathrm{var}[R_*]}. \tag{9.5}$$

Plugging in $R = R_*$ we obtain

$$\mathbb{E}[R_*] - R^0 = \lambda \frac{\mathrm{Cov}(R_*, R_*)}{\mathrm{var}[R_*]} = \lambda.$$

Thus, the factor risk premium is the ordinary risk premium of R_*. In this case, (9.4) can be written as

$$\mathbb{E}[R] - R^0 = (\mathbb{E}[R_*] - R^0) \frac{\mathrm{Cov}(R_*, R)}{\mathrm{var}[R_*]}. \tag{9.6}$$

Theorem 9.2. *There is a beta pricing model with a return R_* as the single factor if and only if the return is on the mean-variance frontier (with risk-free asset) and does not equal the risk-free rate.*

Proof. Let R_* be a return on the mean-variance frontier and let R_* not equal the risk-free rate. Then the corresponding portfolio $\boldsymbol{\pi}$ of (6.20) can be written as

$$\boldsymbol{\pi} = \delta \bar{\bar{\Sigma}}^{-1} (\boldsymbol{\mu} - R^0 \mathbf{1})$$

for some $\delta \neq 0$. Then, for each asset i, we have

$$\mathrm{Cov}(R_*, R_i) = \delta(\mu_i - R^0).$$

Exercise 9.2 below shows that $\delta = \mathrm{var}[R_*]/(\mathbb{E}[R_*] - R^0)$, which implies that (9.6) holds.

To show the opposite direction, let us suppose that there is a portfolio with a return satisfying (9.6). Recalling the expression for the Sharpe ratio of any portfolio and replacing (9.6), we get

$$\mathcal{S}(R) = \frac{\mathbb{E}[R] - R^0}{\sqrt{\text{var}(R)}} = \frac{\mathbb{E}[R_*] - R^0}{\sqrt{\text{var}(R)}} \frac{\text{Cov}(R_*, R)}{\text{var}[R_*]} = \frac{\mathbb{E}[R_*] - R^0}{\sqrt{\text{var}(R_*)}} \frac{\text{Cov}(R_*, R)}{\sqrt{\text{var}(R_*)\text{var}(R)}}$$

But $\frac{\text{Cov}(R_*,R)}{\sqrt{\text{var}(R_*)\text{var}(R)}}$ is the correlation of R and R^*, which is a coefficient that is less than 1 in absolute value[1]. Hence

$$|\mathcal{S}(R)| \leq |\mathcal{S}(R_*)|$$

Since this is true for any portfolio, R_* must be in the mean-variance frontier with risk-free asset. □

Example 9.3 (Capital Asset Pricing Model)**.** *If the return R_* is taken to be the return of the* market portfolio[2] *, the one-factor beta pricing is the CAPM (Capital Asset Pricing Model) we introduced in 5.1.2.*

Note, though that the model is deduced here from a more data-driven approach rather than a utility maximisation one.

9.2.2 Fama-French models

An example of a multi-factor pricing model is provided by the Fama and French models presented in the papers **??**.

The authors found, via statistical analysis of historical series, that more factors beyond the market portfolio were needed to better capture the mean excess return. The additional factors reflect differences in the average excess return produced between companies that have different size, book-to-market ratio, profitability or disposition to invest in their own projects.

The initial 3-factor model, first proposed in 1992, created interesting debates on whether it invalidated market efficiency, understood as the overall tendency of the market to create a portfolio that would belong to the mean-variance frontier: indeed, if market efficiency holds, the market portfolio coincides with the tangency portfolio, and, as we showed, we should be able to find a beta pricing model using the unique factor. Hence, the need to add other factors or even a constant in front is a strong argument against this this supposed efficiency.

Those defending the efficiency assumption argued for an indirect explanation regarding higher capital risk associated to some of these factors (for example, small companies could be seen to be inherently 'riskier' on average than big ones). Others argued that the additional factors simply reflected incorrect pricing from market participants.

The 5-factor model was an update proposed to improve the predictability power of the model. However, it is important to highlight that factors seem to depend on country/market, so they are not universal. Here, we include the factors as stated in the original research, which correspond to the US market:

[1] It can be easily proved from Cauchy-Schwarz inequality

[2] defined by proportions $\pi^1, \ldots, \pi^n$ equal to the ratio of the market capitalisation of a company and the total cumulated market capitalisation

- 3-factor:
$$\mathbb{E}[R] - R^0 = \alpha + \beta_1(\mathbb{E}[R_M] - R^0) + \beta_2 SMB + \beta_3 HML + \epsilon$$
where SMB is difference of returns of two portfolios of 'small' and 'big' stocks; and HML difference in return of two portfolio with high and low Book-to-Market ratio.
- 5-factor:
$$\begin{aligned}\mathbb{E}[R] - R^0 =& \alpha + \beta_1(\mathbb{E}[R_M] - R^0) + \beta_2 SMB + \beta_3 HML \\ &+ \beta_4 RMW + \beta_5 CMA + \epsilon\end{aligned}$$
where RMW is difference of returns of two portfolios with 'strong' and 'weak' profitability; and CMA difference in return of two portfolio with low and high inner investment.

The composition of the portfolios used to calculate the factors can be obtained from Kenneth French's website French (2020).

Remark 9.1. Here, we have used the notation α for the β_0 term introduced in the linear model. This is in accordance to the convention in finance, where the letter is so well-known that the terms "search for alpha" (meaning searching for returns above the market) is current.

9.3 Exercises

Exercise 9.1. Considering (9.1), assume that ϵ has zero expectation and show (9.3). Then, show (9.2), assuming that the variance covariance matrix $\bar{\bar{\Sigma}}$ is invertible by solving the problem
$$\min_{\beta_1,\ldots,\beta_\kappa \in \mathbb{R}} \operatorname{var}(\epsilon).$$

Exercise 9.2. Check that in the proof of Theorem 9.2,
$$\delta = \operatorname{var}[R_*]/(\mathbb{E}[R_*] - R^0)$$
(Hint: Recall (6.20)!)

Chapter 10

Risk measures in practice

> Knowledge is of no value unless you put it into practice.
>
> *Anton Chekhov*

In this section, we present some considerations about the estimation and evaluation of risk measures in practice. In particular, we consider three questions: How to calculate risk measures for assets that might not be directly priced from market? How is the calculation actually done in practice? And, how can we verify that our estimation makes sense?

An essential result to apply risk measures in practice essentially means that an approximation to the risk measure of a P&L can be produced from the value of the risk measure calculated on an *approximation* of the P&L (in the sense of distribution). Indeed, we have the following Proposition:

Proposition 10.1. *Value at risk, expected shortfall, standard deviation, all utility-based risk measures we presented are continuous in distribution; that is, assume that $\{X_n\}$ converges in distribution to X, then $\rho(X_n) = \rho(X)$ for all the examples above.*

Proposition 10.1 justifies many procedures used in practice: for example we can use finite sampling (like in Monte Carlo methods) or Taylor approximation of certain mappings precisely because these procedures generate sequences that converge in distribution.

10.1 The measurement cycle

Market risk management in practice includes, as an essential task, the frequent estimation of the incurred risks in all market-related positions of a financial company. More specifically, they need to be calculated at the beginning of one period, using available market or historical information to be valid throughout the length of this period, then re-calculated and so on.

The process starts with gathering reliable data: this is crucial as conclusion in all stages will be wrong if data is not of high quality. The available data is then used to identify risk factors and find mappings from all P&L to these factors. Methods of estimation risk measure can then be applied and the solution is validated using methods like backtesting. The process is illustrated in Figure 10.1.

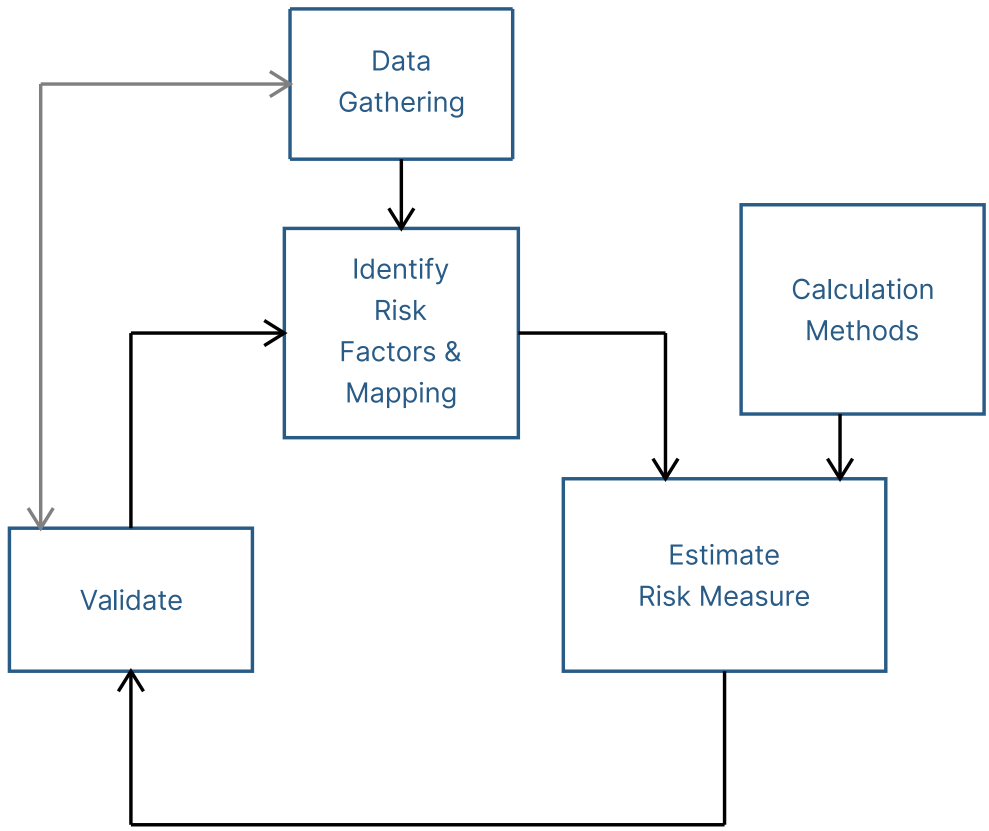

Figure 10.1: Risk measurement cycle

We discuss further each step below.

10.2 Risk factors and mapping of risks

The positions of a financial company can be extremely complex, including thousands of different types of assets, most of which are not traded liquidly in the market.

Hence, in practice, a model is introduced so that the overall value of a position is expressed in terms of a small(ish) number of (risk) factors (see the Chapter on factor models).

Example 10.2 (Stock portfolio)**.** *As before, set θ_i : to be the shares acquired on asset i, and π_i : the corresponding proportion of wealth in asset i.*

Take logarithmic prices of quoted stocks as risk factors. Then $X_i = \log(S_{i+1}/S_i)$ *and*

$$\Delta V_{k+1} = V_{k+1} - V_k = \sum_{i=1}^{n} \pi_i V_k(\exp(X^i_{k+1}) - 1)$$

Example 10.3 (European option portfolio on common stock)**.** *We take the price of the stock, interest rate and implied volatilities as risk factors.*

$$\Delta V_{k+1} = \sum_{i=1}^{n} \theta_i \left[C^{\mathrm{BS}}(t_{k+1}, S_{k+1}, r_{k+1}, \sigma_{k+1}, K_i, T_i) - C^{\mathrm{BS}}(t_k, S_k, r_k, \sigma_k, K_i, T_i)\right]$$

where $C^{BS}(t, S, r, \sigma, K, T)$ *is the Black-Scholes formula.*

Example 10.4 (Bond portfolio)**.** *Take as risk factor the continuously compounded yield curve* $y(s,T) := -(T-s)^{-1} \log p(s,T)$

$$\begin{aligned}
\Delta V_{k+1} &= \sum_{i=1}^{n} \theta_i (P(t_{k+1}, T_i) - P(t_k, T_i)) \\
&= \sum_{i=1}^{n} \theta_i \left(e^{-(T_i - t_{k+1}) y(t_{k+1}, T_i)} - e^{-(T_i - t_k) y(t_k, T_i)}\right) \\
&= \sum_{i=1}^{n} \theta_i e^{-(T_i - t_k) y(t_k, T_i)} \left(e^{-(T_i - t_{k+1}) y(t_{k+1}, T_i) + (T_i - t_k) y(t_k, T_i)} - 1\right) \\
&= \sum_{i=1}^{n} \pi^i_k V_k \left\{\exp[-(T_i - t_{k+1})[y(t_{k+1}, T_i) - y(t_k, T_i)] + y(t_k, T_i)(t_{k+1} - t_k)] - 1\right\}.
\end{aligned}$$

where we used that $\pi^i_k V_k = \theta_i P(t_k, T_i)$.

10.3 Risk measure estimation

Even if we have simplified how to obtain losses from some market variables, in general, the objective probability measure underlying the possible losses is unknown. Therefore we need to estimate the probability measures of the risk factors.

There are two main approaches: either we *explicitly* assume that the risk factors or their changes follow a known distribution; or we implicitly assume that we can deduce this distribution by observing past market information.

10.3.1 Explicit model assumption: The closed-form case

If we can safely assume that the P&L follows a distribution for which we can explicitly calculate the risk measure, then we simply use the closed-form formula in terms of the parameters of the solution. For example, if we want to calculate value at risk of a P&L that can be assumed to be Gaussian, it suffices to calculate mean and variance and replace in the closed-formed formula for the expressions for V@R and ES (see exercises 4.1 and 4.2).

In many cases, we cannot make such assumption for the P&L but we can do so for the *factors.* We can then make effective use of Proposition 10.1 to deduce an approxiamtion f the value of the risk measure. A typical example combines linearisation (also known as Δ approximation) and a Gaussian assumption. Assuming ΔX_i are jointly Gaussian, we can linearise the function f so that for some $a \in \mathbb{R}$, $b \in \mathbb{R}^n$

$$f(x_1 + \Delta x_1, \ldots, x_n + \Delta x_n) - f(x_1, \ldots, x_n) \approx \nabla f(x_1, \ldots, x_n) \cdot (\Delta x_1, \ldots, \Delta x_n) = \boldsymbol{\Delta x}^\top \bar{b}.$$

Exercise. Argue how we can obtain a linearisation result using Taylor's theorem.

Let us make this example more concrete.

Example 10.5 (Stock portfolio)**.** *With $\boldsymbol{F}_k$ the vector of log-returns we have using linearisation that*

$$V_{k+1} = V_k \sum_{i=1}^{n} \pi_i \exp(F^i_{k+1}) \Rightarrow \Delta V_{k+1} = V_k \sum_{i=1}^{n} \pi_i \Delta F^i_{k+1} + O(||\Delta F_{k+1}||^2)$$

If the data supports stationarity and Gaussian assumptions (say $\Delta \boldsymbol{F}_k \sim \mathcal{N}(\boldsymbol{m}, \bar{\bar{\Sigma}})$), and using the fact that a linear function of a multidimensional normal random variable is normal,

$$\mathrm{V@R}^\alpha_\mathrm{k}(\Delta \mathrm{V}_{\mathrm{k}+1}) \approx \mathrm{V}_\mathrm{k}(\boldsymbol{\pi}^\top \bar{\bar{\Sigma}} \boldsymbol{\pi})^{1/2} \Phi^{-1}(\alpha) - \mathrm{V}_\mathrm{k}(\boldsymbol{\pi} \cdot \boldsymbol{m})$$

Example 10.6 (One stock, GARCH model)**.** *Suppose that increments in log returns follow a GARCH(1,1)*

$$\Delta F_{k+1} = a_0 + a_1 \Delta F_k + \sigma_k \xi_{k+1}$$
$$\sigma_k^2 = b_0 + b_1 \sigma_{k-1}^2 + b_2 \xi_k^2$$

Where $(\xi_k)_{k \in \mathbb{N}}$ are i.i.d and standard Gaussian.

As long as we can estimate the parameters, we can define

$$m_k := a_0 + a_1 \Delta F_k$$

so that

$$\mathrm{V@R}^\alpha_\mathrm{k}(\Delta \mathrm{V}_{\mathrm{k}+1}) \approx \sigma_\mathrm{k} \mathrm{V}_\mathrm{k} \Phi^{-1}(\alpha) - \mathrm{m}_\mathrm{k} \mathrm{V}_\mathrm{k},$$

Note that to completely characterise the solution, we need to have values for the parameters m_i, σ_i. These can be estimated from available historical information, using statistical techniques (for example Maximum likelihood) from historical data.

The linear approximation used in conjunction with the normal assumption on risk factor changes justifies the use of sensibilities as risk management tools. Indeed, the best linear approximation makes appear the derivatives of the mapping function with respect to each one of the risk factors. Note, on the other hand, that the method is theoretically applicable to risk analysis for small changes in the risk factors, which means it is usually reserved to short time interval Δ.

This method is advantageous because it is inexpensive, analytical, and can be easily interpreted. But all these advantages are at the expense of two very important assumptions that are usually not verified in practice: that the linearisation is a good approximation and that the risk factors have a joint distribution for which we have closed form expressions (for example multivariate Gaussian).

10.3.2 Explicit model assumption: The Monte Carlo simulation case

In this method, we also assume that the risk factors follow a parametric distribution, but we no longer require the risk measure to have a closed form. As a consequence, we do not need to perform additional transformations (like linearisation), and can choose from a broader family of distributions.

The method can be summarised as follows:

- A parametric model is chosen for the value of the vector of risk-factors at time $k+1$ given the information at time k: $\boldsymbol{F}_{k+1}|\mathcal{F}_k$.
- The distribution is calibrated using market information, or possible shock considerations
- We generate a large number M of samples of the risk factors following the assumed calibrated distributions, given $\mathcal{F}_k$: $(F_{k+1}^{i,(1)},\ldots,F_{k+1}^{i,(M)})$.
- We calculate $\Delta V_k^{(j)} = f(t_{k+1}, F_{k+1}^{1,(j)}\ldots,F_{k+1}^{n,(j)}) - f(t_k, F_k^{1,(j)}\ldots,F_k^{n,(j)})$
- The risk measure for $\Delta V^{(j)}$ can then be estimated from the empirical distribution

The last step requires the definition of some estimators for the risk measure. Let us illustrate this step by considering V@R and ES:

Assume we have ordered the simulated samples, that is $\Delta V_k^{(i)} \leq \Delta V_k^{(j)}$ if $i \leq j$. Then, the following are estimators of V@R$^\alpha$:

$$\begin{aligned}
\mathrm{V@R}^\alpha_{\mathrm{k},+} &:= -\Delta \mathrm{V}_\mathrm{k}^{(\lfloor(1-\alpha)\mathrm{M}\rfloor)}; \qquad \mathrm{V@R}^\alpha_{\mathrm{k},-} := -\Delta \mathrm{V}_\mathrm{k}^{(\lceil(1-\alpha)\mathrm{M}\rceil)};\\
\mathrm{V@R}^\alpha_{\mathrm{k,mid}} &:= \lambda \mathrm{V@R}^\alpha_{\mathrm{k},+} + (1-\lambda)\mathrm{V@R}^\alpha_{\mathrm{k},-}; \quad \lambda := (\mathrm{M}\alpha - \lfloor \alpha \mathrm{M}\rfloor)\,.
\end{aligned}$$

Note that $\mathrm{V@R}^\alpha_{\mathrm{k},+}$ corresponds to the value at risk applied to the empirical distribution. Convergence of the estiamtion is then guranteed by the convergence in distribution of the samples and Proposition 10.1. Moreover, as M grows, all estimators converge if the original loss distribution is continuous.

Similarly, the following is an estimator for ES$^\alpha$

$$\mathrm{ES}^\alpha_{\mathrm{k},+} = \frac{-1}{\lfloor(1-\alpha)\mathrm{M}\rfloor} \sum_{\mathrm{i}=1}^{\lfloor(1-\alpha)\mathrm{M}\rfloor} \Delta \mathrm{V}_\mathrm{k}^{(\mathrm{i})}$$

The Monte Carlo method has the advantage of greater flexibility in the distribution choice with respect to the close form one, and permits the user to control the balance between accuracy and calculation time. It also allows for the inclusion of stress cases that may not have occurred in the past. However, it is very expensive and requires choosing form parametric distributions that can be easily simulated.

10.3.3 Historical simulation

This is essentially a version of the Monte Carlo simulation, where we use the **historical empirical distribution** obtained from the data. The key assumption is that the risk-factor are driven by a *stationary* factor.

Stationary processes

A stochastic process is said to be stationary if its law does not change when shifted in time. Mathematically, this means that X is stationary if for all $\tau \in \mathbb{N}$, and all $(t_1, t_2, \ldots, t_n) \in \mathbb{N}^n$,

$$\mathbb{L}(X_{t_1}, X_{t_2}, \ldots, X_{t_n}) = \mathbb{L}(X_{t_1+\tau}, X_{t_2+\tau}, \ldots, X_{t_n+\tau})$$

where $\mathbb{L}$ denote the law of the joint processes.

Example 10.7. *Let $(X_s)_{s\in\mathbb{N}}$ be i.i.d. Then X is stationary.*

In particular, one can imagine the previous example as if X were the output of throwing a sequence of dice independently. X_i would be the outcome of the ith die. Stationarity essentially tell us something that lies behind the frequentist approach of probability: that we can understand the distribution of one die by observing the outcomes of many others. This is a great property to have when trying to infer information from past observations.

A weaker form of stationarity limits this "equivalence" to the preservation in time of their first two moments: a stochastic process is wide-sense stationary if its mean and auto-covariance function are invariant with respect to shifts. Mathematically, X is wide-sense stationary if for all $\tau \in \mathbb{N}$, and all $(t_1, t_2) \in \mathbb{N}^2$,

$$\mathbb{E}[X_{t_1}] = \mathbb{E}[X_{t_1+\tau}];$$

and

$$\operatorname{cov}(X_{t_1}, X_{t_2}) = \mathbb{E}[(X_{t_1} - \mathbb{E}[X_{t_1}])(X_{t_2} - \mathbb{E}[X_{t_2}])])] = \operatorname{cov}(X_{t_1+\tau}, X_{t_2+\tau})$$

Example 10.8. *Let $(X_s)_{s\in\mathbb{N}}$ be i.i.d with $\mathbb{E}[X] = 0$. The process $\hat{X}_{t+1} = -a\hat{X}_t + \sigma\sqrt{1-a^2}X_t$ (with $0 < a < 1$) and $\hat{X}_0 = 0$ is wide-sense stationary.*

Using stationarity in practice

Given the above, we can see that assuming that a risk factors is stationary means that we can deduce from observed past performance an approximation to its probability distribution: this justifies our use of the empirical distribution for the calculation of risk measures.

As we have seen in our numerical experiments, not all processes are evidently stationary. But, even if the direct variables we are looking at are not stationary, we can still look for models that map other stationary risk factors to these non-stationary variables.

Example 10.9. *Assume that*

$$F_{k+1} = F_k + \xi_{k+1},$$

for stationary $(\xi_k)_{k\in\mathbb{N}}$ with unknown distribution.

Now, using stationarity we deduce the distribution of ξ from the data: essentially, past observations of ΔF play the same role of the simulated sample in the Monte Carlo method. We can thus apply steps 4-5 and use the estimators defined for the Monte Carlo approach.

Let us emphasise that we are not forced to assume directly that the increments are stationary. For example, if we want to reflect changes in the past circumstances, a time-weighting can be added to give more importance to recent values, that is, we introduce a weighting function w so that

$$\tilde{P}^w[L = x] := \frac{1}{\sum_{i=1}^N w(i)} \sum_{i=1}^N w(i) \mathbb{1}_{\{x = L^{(-i)}\}}$$

A typical example is the *exponentially weighted*, which for a fixed parameter $\lambda > 0$ uses $w(x) = \exp(-\lambda x)$. Further time structure can be added by assuming explicit models on how observations evolve as a time series: examples are ARMA, ARIMA, GARCH models (check the references for more information on these models).

Historical simulation can also be implemented or complemented by the use of *bootstrapping*. Bootstrapping belongs to the set of methods based on resampling: the essential idea is to sample with replacement from the empirical distribution either to calculate the risk measures (and avoid simply "replaying history") and to calculate tests of hypothesis (as in back-testing).

Arguably using historical information is appealing, as we make only two essential a priori assumptions: that the distributions of future observations can be deduced from past observations and that past observations follow a given time evolution (for example independent increments). Moreover, by construction, previous loss scenarios are included: in particular, one can (and should!) choose the observation window to include stress periods for the risk factors. Unfortunately this is a very demanding method both in the quantity and quality of historical information. For some applications, it is also demanding in the amount of calculations. In addition, it excludes changes in the environment that might modify previous history.

10.4 Back-testing

As we have seen, risk measure estimation in practice is subject to both numerical and model errors. An essential tool to mitigate those errors is back-testing

Back-testing comprises the use of some statistically meaningful tests to evaluate the goodness of our risk estimation. Different kinds of tests can be defined to establish different properties that are desired from our estimation:

- Coverage property: This kind of test tries to rule out a calculation method/model that evidently underestimates risks when evaluated to past observations.
- Comparative effectiveness: This is tested by ruling out risk measure estimations that are significantly less effective in describing the realised past risk than a reference one (for example the standard regulation model).
- Clustering: Any type of structural clustering assumptions of the model can be tested against the realised sequence of losses. For example, if the initial model assumes independence of losses between periods, the test is failed if there is enough evidence of joint dependence.

We centre our discussion on some usual statistics to test the coverage property of V@R and ES.

10.4.1 Framework description

In the following, we consider the results of T (passed) periods. By abusing of our notation, let us denote by $\mathrm{V@R}^{\alpha}_{-t}$ (respectively $\mathrm{ES}^{\alpha}_{-t}$), the value at risk (respectively expected shortfall) at level α for the period $-t, -t+1$, for $t = 1, \dots T$.

Let ΔV_{-t}, for $t = 1, \dots T$ denote the realised P& L at the **end** of period t. Define

$$I_{-t} := \mathbb{1}_{\{-\Delta V_{-t} \geq \mathrm{V@R}^{\alpha}_{-t}\}};$$

that is, the V@R failure indicator on the period $-t, -t+1$.

Every covering back-test entails the evaluation of a null assumption H_0 which usually contains (but may not limited) the fact that the risk measure is well estimated. For all the following tests, let us denote by κ the statistical significance of the test[1]. Recall that this means that we want to limit the Type I error (rejecting the null hypothesis provided it is true) to be at most $1-\kappa$. Equivalently, we reject the null hypothesis if the realised tested statistic falls in the $1-\kappa$ tail of the distribution,

10.5 Back-testing covering for V@R

10.5.1 Test 1: Assuming independence of failures

Let us define as our null hypothesis

$$H_0 : \text{The sequence } \{I_{-t}\}_{t=1,\ldots,T} \text{ is i.i.d. and } \mathbb{P}[-\Delta V_{-t} > \mathrm{V@R}^{\alpha}_{-\mathrm{t}}(\mathrm{L})] = (1-\alpha),$$

which is equivalent to assuming

$$H_0 : \text{The sequence } \{I_{-t}\}_{t=1,\ldots,T} \text{ is i.i.d. and } \mathbb{P}[I_{-t} = 1] = (1-\alpha),$$

in other words, we suppose that there is no clustering on V@R breaches and that the V@R is well calculated.

Under these assumptions, each variable I_{-t} is a Bernoulli random variable with mean $1-\alpha$. Hence, the statistic

$$Z_{\mathrm{V@R},1} = \sum_{t=1}^{T} \hat{I}_{-t}.$$

follows the binomial distribution. Its mean is $T(1-\alpha)$ and variance $T\alpha(1-\alpha)$. Given that we know this distribution, we can evaluate if the realised Z_1 falls within the κ bilateral confidence interval.

A similar estimator is obtained by dividing the previous estimator by the number of observations and using the central limit theorem, we get that

$$\tilde{Z}_{\mathrm{V@R},2} = \frac{Z_{\mathrm{V@R},1}}{T} = \frac{1}{T}\sum_{t=1}^{T} \hat{I}_{-t}.$$

under H_0 is asymptotically normal with mean $(1-\alpha)$ and variance $\frac{1}{T}\alpha(1-\alpha)$. Hence, the estimator

$$Z_{\mathrm{V@R},2} = \frac{\sqrt{T}}{\sqrt{\alpha(1-\alpha)}}\left(\tilde{Z}_{\mathrm{V@R},2} - (1-\alpha)\right)$$

follows asymptotically a standard Gaussian distribution.

Yet a third estimator can be defined using the previous two, to obtain a chi-square test. Let

$$Z_{\mathrm{V@R},3} = 2\log\left[\frac{(\tilde{Z}_{\mathrm{V@R},2})^{Z_{\mathrm{V@R},1}}(1-\tilde{Z}_{\mathrm{V@R},2})^{T-Z_{\mathrm{V@R},1}}}{(\alpha)^{T-Z_{\mathrm{V@R},1}}(1-\alpha)^{Z_{\mathrm{V@R},1}}}\right].$$

[1]Usually it is called α, a letter we have used to denote the level in V@R and ES.

This is an estimator coming from a likelihood ratio, and it follows asymptotically a chi-squared distribution with 1 degree of freedom χ^2_1.

Clearly, we can use all the previous statistics and distributions to test for under/over-estimation only, i.e.

$$H_0 : \text{The sequence } \{I_{-t}\}_{t=1,\dots,T} \text{ is i.i.d. and } \mathbb{P}[I_{-t} = 1] \leq (1-\alpha).$$

10.6 Backtesting ES

As we will see in the following, the main difference between back-testing V@R and ES is that the latter generally requires storing additional information to the calculated estimator for each day.

10.6.1 Z test on the quantile space

This test introduced in Costanzino and Curran (2015) is based on the observation that under the null hypothesis

$$H_0 : \text{The sequence } \{I_{-t}\}_{t=1,\dots,T} \text{ is i.i.d., } \mathbb{P}[-\Delta V > \mathrm{V@R^p(\Delta V)}] = (1-\mathrm{p}), \text{ for all } \mathrm{p} \geq \alpha,$$

the estimator

$$\tilde{Z}_{\mathrm{ES},1} = \frac{1}{T}\sum_{t=1}^{T} \frac{1}{1-\alpha}\int_{\alpha}^{1} \mathbb{1}_{\{-\Delta V_{-t} \geq \mathrm{V@R^u_{t-}}\}}\mathrm{du}$$

is asymptotically normal with mean and variance

$$\mathcal{N}\left(\frac{1-\alpha}{2}, \frac{(1-\alpha)(4+3\alpha)}{12T}\right),$$

and so an asymptotically standard normal estimator can be constructed.

Note that, somehow surprisingly, the estimator does not explicitly require to record the values of expected shortfall! On the contrary, the test is based on recording the values of violations on value at risk and their respective quantile. This is of course a consequence that the null hypothesis is stronger than before: not only we are assuming that we have independence and identically distributed samples, but also that the whole tail distribution is well estimated.

10.6.2 Z test on conditional expectation

This estimator proposed in **?** is only applicable to continuous distributions. It is based on the following null hypothesis

$$H_0 : \mathbb{P}[-\Delta V > \mathrm{V@R^p(\Delta V)}] = (1-\mathrm{p}), \text{ for all } \mathrm{p} \geq \alpha,$$

The estimator is:

$$Z_{\mathrm{ES},2} = 1 - \sum_{t=1}^{T} \frac{-\Delta V_{-t} I_{-t}}{T(1-\alpha)\mathrm{ES}^{\alpha}_{\mathrm{t}}}.$$

In this case, however, we do not have an asymptotic convergence result. Therefore, in order to perform the statistical test, it is necessary to perform a Monte Carlo simulation based on the full tail distribution available form the null assumption. See **?** for further details.

10.6.3 Ellicitability

An interesting property related to back-testing of risk measure is ellicitability.

Definition 10.10. Let $\mathcal{P}$ be a class of probability measures in Ω. A (law invariant) risk measure ρ is said to be *ellicitable* relative to the class $\mathcal{P}$ if there is a scoring function $s : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ such that

$$\rho(L) = \arg\min_{x\in\mathbb{R}} \mathbb{E}[s(x, L)] \text{ for all } L \text{ with law in } \mathcal{P}.$$

Some examples of ellicitable risk measures are: V@R and expectiles. Expected shortfall and standard deviation, on the other hand, are not ellicitable (see Ziegel (2016)). This result was understood by part of the financial community as a proof that the apparent difficulty to back-test unconditional covering for Expected Shortfall was a consequence of lacking an essential property.

We remark, however, that ellicitability is not necessary for unconditional covering, but might be necessary for to perform efficient comparison back-tests, that is, tests were the historical performance of a risk measure estimation is compared with that of a different estimator. Indeed, a scoring function gives an immediate way to test the best result amongst two possible risk measure estimations by computing and comparing an approximation of the scoring function.

10.7 Exercises

Exercise 10.1. Let X be an integrable continuous random variable with p.d.f. $f_X : \mathbb{R} \to \mathbb{R}_{++}$ (i.e. assume that $f_X > 0$).

i. Show that value at risk satisfies

$$\mathrm{V@R}^{\alpha}(\mathrm{X}) = \arg\min_{\mathrm{r}\in\mathbb{R}}\{\alpha\mathbb{E}[(\mathrm{X}+\mathrm{r})^{-}] + (1-\alpha)\mathbb{E}[(\mathrm{X}+\mathrm{r})^{+}]\}$$

Note: The continuity assumption is put only for convenience in the proof, but in fact the result holds in general. Compare with Proposition 4.30.

ii. Explain why the integrability assumption is necessary.

iii. Conclude that value at risk is ellicitable for integrable random variables and give a scoring function.

Exercise 10.2. One dimensional stock: Assume that $S_{n+1} = S_n \exp(\xi_{n+1})$, where all the $\xi_1, \xi_2 \ldots,$ are i.i.d. with $\mathcal{N}(m, \sigma^2)$. Set $\Delta S_i = S_i - S_{i-1}$

i. Show that the random variables $I_i = \mathbb{1}_{\{-\Delta S_i \geq \mathrm{V@R}^{\alpha}(\Delta \mathrm{S_i})\}}$, for $i = 1, \ldots$ are i.i.d. and give their distribution. Recall that $\mathbb{1}_A$ denotes the indicator function of the set A.

ii. In a set of 400 periods, the observed losses are seen to exceed the estimated value at risk at 99% confidence on 10 occasions. Perform a back-test to establish the covering property of a reserve made with this value at risk calculation. Fix the confidence of the test at 97.5%. You may use that $\Phi^{-1}(0.975) \approx 1.96$.

Exercise 10.3. Assuming that H_0 in Section 10.5.1 is true, show the claims on the distributions of $Z_{\mathrm{V@R},1}, \tilde{Z}_{\mathrm{V@R},2}$ and $Z_{\mathrm{V@R},2}$.

Some final comments

- When dealing with back-testing, we have assumed implicitly that there is no overlapping information between periods. In some applications, this would severely limit the number of available observation dates. An adjustment on the hypothesis testing would be necessary, but this is outside the scope of our course.
- A test on independence of the V@R infringements can be found in Christoffersen (1998). If this test is failed, the asymptotic results are in general not guaranteed, and some extra procedure (for example fitting a time dependence) and corresponding adjustments on the distribution of the estimator (that can require Monte Carlo simulations) might be necessary to perform the test.

Bibliography

Acciaio, B. and Penner, I. (2011). Dynamic Risk Measures. In Nunno, G. D. and Øksendal, B., editors, *Advanced Mathematical Methods for Finance*, pages 1–34. Springer Berlin Heidelberg.

Artzner, P., Delbaen, F., Eber, J.-M., and Heath, D. (1999). Coherent Measures of Risk. *Mathematical Finance*, 9(3):203–228.

Back, K. (2010). *Asset pricing and portfolio choice theory*, volume Financial Management Association survey and synthesis series. Oxford University Press, New York, New York.

Bernoulli, D. (1954). Exposition of a New Theory on the Measurement of Risk. *Econometrica*, 22(1):23–36.

Boyd, S. P. and Vandenberghe, L. (2004). *Convex optimization.* Cambridge University Press, Cambridge, UK ; New York.

Chen, K. (2005). *Matrix Preconditioning Techniques and Applications.* Cambridge University Press, 1 edition.

Christoffersen, P. F. (1998). Evaluating interval forecasts. *International economic review*, pages 841–862.

Clarke, F. (2013). *Functional analysis, calculus of variations and optimal control / Francis Clarke.* Graduate texts in mathematics ; 264. Springer, London.

Costanzino, N. and Curran, M. (2015). Backtesting General Spectral Risk Measures with Application to Expected Shortfall. SSRN Scholarly Paper ID 2514403, Social Science Research Network, Rochester, NY.

Davis, M. H. and Norman, A. R. (1990). Portfolio selection with transaction costs. *Mathematics of operations research*, 15(4):676–713.

Durrett, R. (2010). Probability: Theory and Examples Rick Durrett January 29, 2010. *Biometrics*, 49(3):497.

French, K. R. (2020). Kenneth R. French - Home Page.

Föllmer, H. and Schied, A. (2011). *Stochastic finance: an introduction in discrete time.* Walter de Gruyter.

Gilboa, I. (2009). *Theory of decision under uncertainty*, volume 45 of *Econometric Society Monographs.* Cambridge University Press, Cambridge.

Munkres, J. R. (2014). *Topology.* Pearson, Harlow, 2. ed., pearson new internat. ed edition.

Platen, E. (2011). A benchmark approach to investing and pricing. In *The Kelly capital growth investment criterion: Theory and practice*, pages 409–426. World Scientific.

Rockafellar, R. T. (1970). *Convex analysis.* Princeton university press.

Savage, L. J. (1954). *The foundations of statistics.* John Wiley & Sons, Inc., New York; Chapman & Hill, Ltd., London.

Tversky, A. and Kahneman, D. (1986). Rational choice and the framing of decisions. *Journal of business*, pages S251–S278.

Tversky, A. and Kahneman, D. (1992). Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and uncertainty*, 5(4):297–323.

Ziegel, J. F. (2016). Coherence and Elicitability. *Mathematical Finance*, 26(4):901–918.