PANTEION UNIVERSITY OF SOCIAL AND POLITICAL SCIENCES

ΠΑΝΤΕΙΟΝ ΠΑΝΕΠΙΣΤΗΜΙΟ ΚΟΙΝΩΝΙΚΩΝ ΚΑΙ ΠΟΛΙΤΙΚΩΝ ΕΠΙΣΤΗΜΩΝ

SCHOOL OF ECONOMIC SCIENCES AND PUBLIC ADMINISTRATION
DEPARTMENT OF PUBLIC ADMINISTRATION
POSTGRADUATE PROGRAMME
«PUBLIC ADMINISTRATION»
DIRECTION: ECONOMIC SCIENCE

Mean-Variance and Pessimistic Portfolio Allocation: A Comparative Study

MASTER THESIS

Stavroula Athanasakopoulou

Athens, 2017

Three-member Committee

Gregory Kordas, Assistant Professor of Econometrics, Panteion University (Supervisor)

George Petrakos, Associate Professor of Quantitative Methods, Panteion University

Anastasios Tasopoulos, Dean of School of Economic Sciences and Public Administration, Panteion University

Abstract

Mean-Variance and Pessimistic Portfolio Allocation: A Comparative Study

by Stavroula Athanasakopoulou
07115M057

The main objective of this thesis is to comparatively study two risk measures in Finance, Mean-Variance (MV) and Conditional Value-at-Risk (CVaR), taking into account their connection to the econometric process of regression. Trigger of this thesis topic was the extensive interest to elaborate and further study the issues discussed in a paper that my thesis advisor, Gregory Kordas, has co-authored, namely Bassett, Koenker, and Kordas (2004). Recent developments in the general theory of choice under uncertainty, of which financial portfolio selection is a special case, replace the classical Expected Utility maximization with a Choquet expectation that allows for pessimism in that it overweights the probabilities of unfavorable outcomes and underweights the probabilities of favorable ones.

In this master thesis, a substantial Monte Carlo examination of the properties of simple single-quantile portfolios has been undertaken for the cases when asset returns follow skewed and fat-tailed distributions. This is computationally possible, since there exists very good software that do single-quantile regressions in many statistical packages (`R`, `Stata`, etc.). Computer code in `R` has been structured and the results arising are very encouraging, considering that the computed portfolios do well both in the lower, as well as, in the upper tail of the distribution.

Bassett *et al.* (2004) also consider more general distortions that cannot be expressed as Quantile/CVaR-risk problems but are still Pessimistic, Choquet-expectation optimal choices. One very interesting distortion function is given in an Example of their paper, where the investor picks portfolios according to the *minimum* of several (say 2 or 3) realizations of the underling asset distribution. This leads to potentially very pessimistic behavior, that can be examined in future research.

Keywords: *Choquet capacity, Expected Utility, Mean-Variance Analysis, Conditional Value-at-Risk Analysis, OLS Regression, Quantile Regression*

Acknowledgements

First of all, I would like to express my gratitude to my family members for their support in every field during this two-year postgraduate program in Economic Science. I would also like to sincerely thank my supervisor, Gregory Kordas, for his valuable suggestions and guidance. Due to him I managed to deal with financial, statistical and econometric topics of great difficulty, which I later used during the elaboration of my thesis.

Furthermore, I would like to thank the Associate Professor George Petrakos and the Dean of School of Economic Sciences and Public Administration Anastasios Tasopoulos for the exceptionally helpful and informative conversations we had on my thesis topic. Finally, I could not omit to express my gratitude to all of my professors in the context of this postgraduate program, as well as to Panteion University broadly which has hosted me in its structures these two years.

Contents

List of Figures

List of Tables

List of Abbreviations

CDs	**C**ertificates of **D**eposit
CVaR	**C**onditional **V**alue **a**t **R**isk
EU	**E**xpected **U**tility
Forex	**F**oreign **e**xchange **M**arket
IPOs	**I**nitial **P**ublic **O**fferings
MBS	**M**otage-**b**acked **S**ecurities
MV	**M**ean-**V**ariance
NYSE	**N**ew **Y**ork **S**tock **E**xchange
OLS	**O**rdinary **L**east **S**quares
OTC	**O**ver-**t**he-**C**ounter
Repos	**R**epurchase **a**greements
SEC	**S**ecurities and **E**xchange **C**ommission
SEU	**S**ubjective **E**xpected **U**tility
SnP	**S**tandard a**n**d **P**oor's
U.S.	**U**nited **S**tates
VaR	**V**alue **a**t **R**isk

To my parents,

Efthimia and Stavros

Introduction

This master thesis presents a comparative study between two risk measures in Finance, Mean-Variance (MV) and Conditional Value-at-Risk (CVaR). It is divided into three parts; the first part (Chapters (2) to (5)) presents the identification, both theoretical and mathematical, of the Theory of Expected Utility and the Theory of Mean-Variance Analysis. The second part (Chapters (6) and (7)) discusses Choquet Expected Utility, as a generalized case of Expected Utility, and CVaR Analysis. The third part (Chapter (8)) discusses the comparison of MV and CVaR, using a Monte Carlo simulation in R. But, before entering in these specific topics, which are the main parts of this thesis, it is appropriate to explain certain characteristics of the world of Finance that will smoothly transfer the reader to the risk measure idea.

In Chapter (1), Financial Markets are categorized and examined, and the Stock Market is defined. After these first declarations, the Efficient Market Hypothesis of Fama (1969) is explained. In Chapter (2), the origins of the idea of the Expected Utility Theory are given based on Bernoulli (1738), Savage (1954) and von-Neumann & Morgenstern (1947). In Chapter (3), Expected Utility is analyzed and the definitions concerning attitudes towards risk are presented (von-Neumann & Morgenstern, 1947). In Chapters (4) and (5) the Mean-Variance Analysis topic is followed and its connection to the OLS regression is explained. Chapters (6) and (7) consist of the connection between the Choquet Expected Utility, the CVaR Analysis and the Quantile regression. Choquet Expected Utility, which is based on a distortion, introduces the idea of *pessimism* and leads to CVaR Analysis and Pessimistic risk measures. The idea of pessimism is based on the fact that anyone would like to get rich but no one would like to go bankrupt, so the probability of the least favorable events should be overweighted in order to avoid great losses. Finally, in Chapter (8), MV and CVaR Portfolios are counterbalanced in the context of a Monte Carlo experiment in R.

It can be said that this thesis follows two *paths*, that both lead to a regression analysis. The first path (Figure (1)) begins with the Classic Axioms of Choice of Savage (1954), goes to the Expected Utility Theory, the Mean-Variance Analysis and ends up with the OLS regression, which is proved to be an alternative

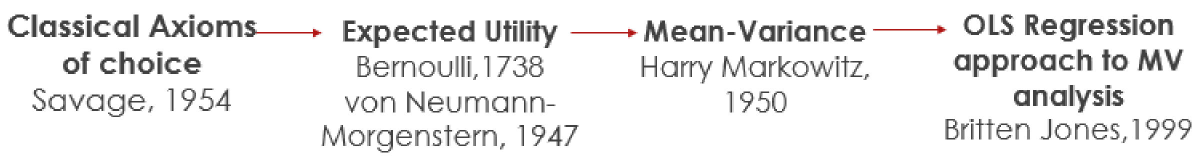

FIGURE 1: From Expected Utility to OlS regression

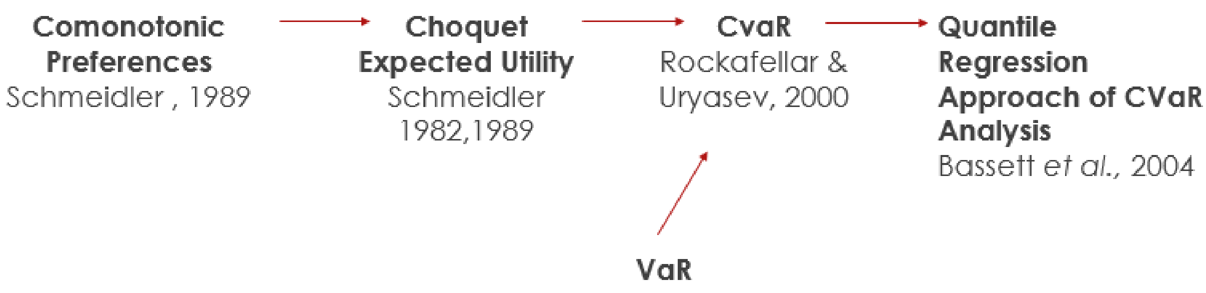

Figure 2: From Choquet Expected Utility to Quantile regression

approach to the Mean-Variance Analysis. The second path (Figure (2)) begins with the Comonotonic Preferences of Schmeidler (1989), proceeds with the Choquet Expected Utility and the CVaR Analysis, where the idea of *pessimism* is introduced. At this point, VaR seems a satisfactory method to use but, as it will be analyzed, VaR fails to establish a position in this path. The second path concludes with the Quantile regression as an identical process of the CVaR Analysis.

At first, the connections between Expected Utility- Mean-Variance Analysis and OLS regression and Choquet Expected Utility- CVaR Analysis and Quantile regression, as mentioned introductory, may seem quite confusing. The reader, though, will be able to understand the connections as he/she moves through the Chapters, given the extensive theoretical, as well as mathematical explanations provided. I hope that the reader will share the same feelings of enthusiasm with me, while the specific issues get unrolled and pessimism, even though it sounds sad, seems a happy choice to make when risk, money and preferences get involved.

Chapter 1

The world of Finance

1.1 Understanding the "Financial Market"

The term **"Financial Market"** is widely used to describe a marketplace of any kind where buyers and sellers "trade". Objects of this trade are the so-called **assets**, which are negotiated in different Financial Markets, based on their special features. The size of a Financial Market may vary depending on the number of its participants. Based on these parameters, the types of assets negotiated, as well as the size of the participants, the Financial Market consists of 3 different markets, the Money Market, the Capital Market and the Foreign exchange Market, as seen in Figure (1.1) and discussed below.

- **Money Market**
  Money Market is the type of Financial Market where assets with high liquidity and very short maturities are negotiated. The time horizon given by the term "*very short maturities*" indicates that assets in this market are negotiated for a short period of time that can be up to one year. Some of the most common Money Market assets are Bankers Acceptances, Commercial Papers, Eurodollar Deposits, Federal Funds, Municipal Notes, Negotiable Certificates of Deposit (CDs), Repurchase Agreements (Repos), state Treasury Bills. The transactions that take place in the Money Market concern participants of high credit, like banks, large companies as well as governments. In some cases, individuals might also be offered the chance to invest small amount of money in these assets by the Money Market Funds.

- **Capital Market**
  Capital Market is the type of Financial Market where Long-Term Debt and Equity-Backed Securities are negotiated. A Capital Market undertakes the issuing of assets for medium-term and long-term periods, that overcome the duration of one year. A Capital Market includes two main financial assets; Equity Securities or, in other words, Stocks, and Debt Securities or, in other words, Bonds. The participants of this market are numerous and may be individual investors, institutional investors, governments, companies, banks and financial institutions. In a Capital Market, contradictory preferences occur as suppliers of capital aim at the combination of the maximum possible return with the lowest possible risk, whereas capital takers aim at the maximum capital at the lowest cost.

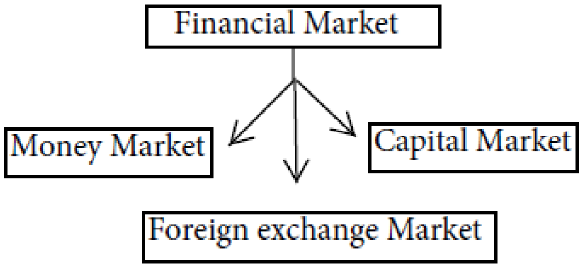

FIGURE 1.1: Financial Market

It can be said that the size of a country's Capital Market is proportional to the size of its economy. Capital Markets create a "*circle of welfare*" as, in their structures, money moves from economically stronger individuals to organizations that need to be productive. The ease or difficulty of this cyclic transaction depicts the "health" of markets around the world. As Capital Market products get widespread, ripples in a country can cause major waves in another place of the world. The most resent result of this interaction is the global financial crisis of 2007-09, which was triggered by the collapse of U.S. Mortgage-Backed Securities (MBS). This financial collapse expanded through Capital Markets all over the world, as banks and institutions in Europe and Asia held trillions of dollars of U.S. securities.

Both Money and Capital Markets are composed of two market sub-categories, the **Primary Market** and the **Secondary Market**. Primary Markets concern the issuing of primary titles such as shares, bonds, state Treasury Bills and government bonds. Secondary Markets concern the negotiation and the exchange of existing securities which were acquired in the Primary Market.

- **Foreign exchange Market**
  The Foreign exchange Market is the type of Financial Market where currencies are traded. The Foreign exchange Market, also known as "Forex", is considered to be the largest Financial Market. Its participants are banks, commercial companies, investment management firms, hedge funds, and retail forex brokers and investors. Participants can buy, sell, exchange and speculate on different currencies. The Forex Market presents some unique features which classify it as one of the most attractive market for investors who want to optimize their profit.

- Working 24 Hours a Day, 5 Days a Week

The transactions occurring in the Forex Market are continuous. Due to time differences, when a Forex Market closes, another Forex Market opens. Unlike stocks, the Forex Market operates 24 hours daily apart from weekends.

- High Liquidity

The Forex Market involves one of the largest asset categories, currencies. Currency trading provides participants of Forex with high liquidity, which is the ability of an asset to be diverted to cash quickly, with no price discount.

- Leverage

The leverage given in the Forex Market is one of the highest forms of leverage that traders and investors can use. Simply put, leverage is a loan given to an investor by his broker. For instance, if someone was to trade at 20:1 leverage, they could trade 20 € on the market for every 1 € on their account. This simply means that they could trade 20,000 € while owning just 1,000 €.

1.2 Stock Market

The term **Stock Market** is linked with the types of markets where equities, bonds and other kinds of securities are traded, either through Formal Exchanges or at Over-the-Counter Markets. The importance of the Stock Market is connected to the fact that it provides companies with capital when in lack, in exchange of a percentage of ownership, a fact that defines the Stock Market as one of the most significant players in the modern economy.

The Stock Market is composed of two basic sub-markets, the Primary Market and the Secondary Market. The Primary Market is where new assets are sold through Initial Public Offerings (IPOs). IPO prices are determined by the amount of shares that a company issues. These shares are mostly sold by banks and bought by institutions. Shares acquired in the Primary Market are then negotiated in the Secondary Market- as mentioned in Section (1.1)-, where individuals may participate as well. Two types of securities are mostly traded on Stock Markets, Over-the-Counter (OTC) and Listed Securities. Stock Market exchanges take place in central cities around the world, such as London and Tokyo.

In the United States, the greatest Stock Market exchanges take place in the **New York Stock Exchange** (NYSE, 1792) located on **Wall Street**, and in the **Nasdaq** (1971). *Wall Street* is the most powerful stock exchange worldwide with a capitalization that is larger than London's and Tokyo's combined. In the Stock Market, different types of specialists are employed, such as traders, stock analysts, stockbrokers, portfolio managers. Regulatory bodies are charged with the control of the Stock Markets. Securities and Exchange Commission (SEC) is such a body in the U.S..

The **Index of stocks** is a measure used to explain Stock Market movements. Indexes are composed of different kinds of stocks and represent the state of a certain market or the state of a certain market's part on a daily basis. A well-known stock index is the **Dow Jones Industrial Average**, which includes the 30 largest companies in the U.S.. Dow Jones is a price–weighted average based on the price of the stocks and it imprints these stock performances. Another significant index is **Standard and Poor's 500**. The SnP 500 includes the 500 largest stocks traded in the U.S.. Both Dow Jones and Standard and Poor's 500 are the most accepted measures of the U.S. Stock Market and are considered to be trustful for the presentation of this country's Economy on a certain period of time.

But why is the Stock Market so significant? The Stock Market gives a company the opportunity to raise capital through stock shares and corporate bonds. Individuals can take part in this financial gain of the company. Of course, investors face a certain risk in this transaction, as they can lose money during price share declines by having to sell the assets at a current lower price. From the aspect of a country's economy, rising stock prices of companies from a particular country imply a healthy and growing market, while an avoidance trend in certain stocks of a particular country indicates lack of confidence in the country's economic prospects. Historically, two periods of economic decline are connected to a crash in Stock Markets, the Great Depression of 1929 and the Great Recession of 2008.

1.3 Predicting the Unpredictable

In his classic book *A Random Walk Down Wall Street*, Burton G. Malkiel demolishes stock market chartists and their intriguing job. Malkiel, an academician who has also held several positions on the boards of some of the most prestigious U.S. mutual funds, sites a number of studies that prove beyond any doubt that charting has **no predictive ability**. To explain the "paradox" that chartists still hold their job places in many brokerage firms, Malkiel justifies it based on the fact that they achieve a remarkable amount of commission money for their employers by recommending many trades to the people that take their advise. Malkiel also gives many examples of "investment gurus" that have for a short time enjoyed fame in Wall Street. These "gurus" claim to have the ability to predict market trends but the most of them end up loosing the money they were trusted to invest. The idea that the Stock Market future performance can not be determined was first supported in 1965 by Eugene Fama who claimed that asset prices follow the **Random Walk Model**. One of the basic assumptions of the Random Walk Model is that we do not know and cannot predict tomorrow's asset prices. The past history of asset prices can be examined, but it cannot be used for forecasting the next movements of a price. Fama later used this specification to describe the content of the **Efficient Market**.

1.4 The Efficient Market Hypothesis

The term *Efficient Market* can be traced to Fama (1969) who first defined the market to be efficient if it "*adjusts rapidly to new information*"(Fama *et all*, 1969). The information element was afterwards adjusted to this definition, with Fama noting that asset prices in an Efficient Market *"fully reflect all available information"* (Fama, 1991).

A Financial Market is called efficient if it reflects all the information in the prevailing security prices correctly. The **Efficient Market Hypothesis** has 3 different forms, depending on the strength of the conditional information set :

- **Weak Form** : In this form, the asset prices include all the information available publicly, taken by the history of prices or returns themselves. According to the Efficient Market Hypothesis, there is no investor who could use past prices information to predict future moves. This means that asset prices do not follow a pattern, which could lead to excess profits. This also implies that asset prices follow a Random Walk, meaning that on average there is no correlation between price changes' sequence.
  Random Walk
  $P_t = P_{t-1} + u_t \Rightarrow$
  $P_t - P_{t-1} = u_t \Rightarrow$
  $\Delta\ P_t = u_t$, where P_t : the level of asset prices and u_t : the error term

- **Semi-strong Form**: In this form, the asset prices include any kind of information available to market participants. According to the Efficient Market Hypothesis, no publicly available information, which is reflected in the actual asset price, can be used to predict future moves. Investors are unable to earn excess profits using any information, whether publicly available or private information.

- **Strong Form** : In this form, asset prices include all information known to any market participant, whether is publicly available or not (internal information). In this form, the Efficient Market Hypothesis says that no information, neither public nor private, can be used to predict future moves, which means that future price moves are completely unpredictable.

In its Strong Form, insider trading has in many cases enabled market participants to find themselves with great amount of money, rising doubts to others regarding their forecast weakness. In its Weak Form, though, there is no underlying information that could help certain participants against others. Price history is publicly available but asset prices follow a Random Walk, making it impossible for someone to compute future price levels. If past price information had economically exploitable predictive power then everyone could get rich. This cannot be an equilibrium situation, so even if such profit opportunities arise from time to time, they are quickly wiped out by the market.

Figure (1.2) depicts the reactions of an Efficient Market to a set of new information. The Efficient Market should react immediately to the good news. This reaction leads to a rise in the asset prices, up to a certain level, where the effect of the new information stabilizes. A delayed Market reaction would

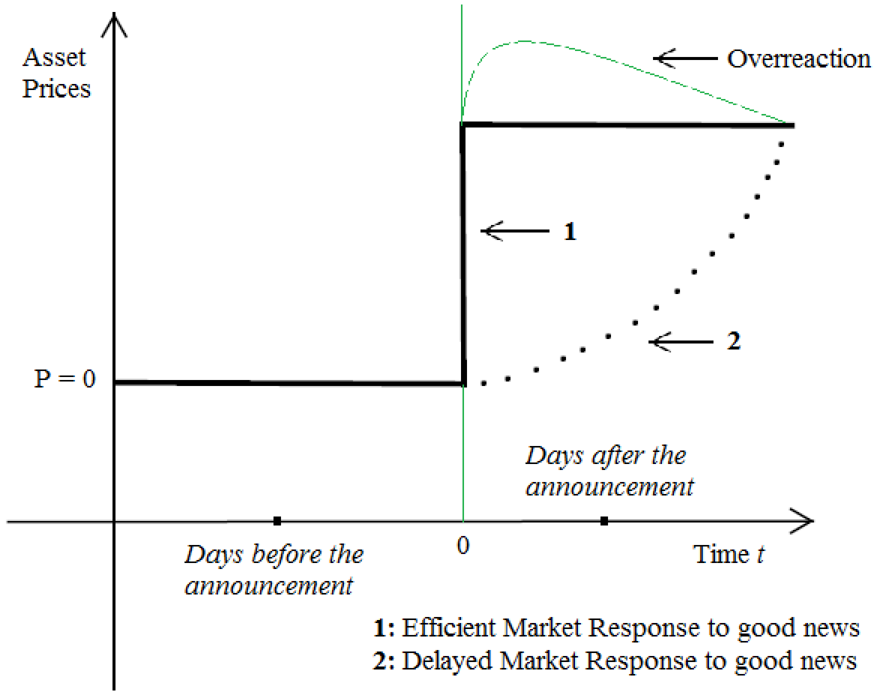

FIGURE 1.2: Efficient Market response to new information

mean a delayed asset price increase. Overreaction to the new set can also be observed, that corresponds to cases when asset prices increase more than it would be expected, and then follow a downward path for a while, until they stabilize.

Below, the concepts of a stochastic process and a martingale are introduced to support the tree different forms of the Efficient Market. A stochastic process is simply a collection of random variables indexed by a parameter t, denoting time. A stochastic process is a martingale with respect to an information set if

$$E(S_{t+1}|I_t) = S_t. \tag{1.1}$$

Letting S_t denote the stock price at time t, the various forms of the Efficient Market Hypothesis may be represented by equation (1.1) with the following information sets,

- **Weak Form** : $I_t = S_t, S_{t-1}$.
- **Semistrong Form** : $I_t =$ all publicly available information.
- **Strong Form** : $I_t =$ all information, public or private.

The expectation E in (1.1) is not taken with respect to the actual probability measure of S_t, P_{S_t}, but with respect to an appropriately standardized measure that accounts for expected return and risk, P_{S_t}. This probability measure is called the **risk neutral measure** and it plays a fundamental role in the theory

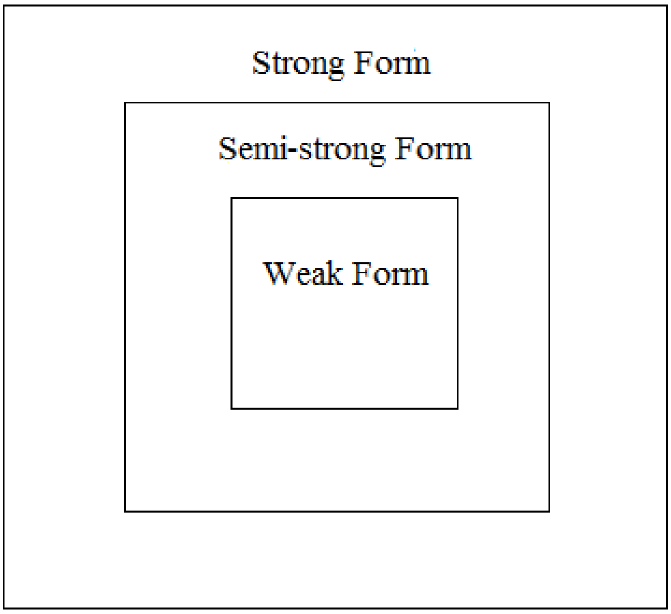

FIGURE 1.3: Forms of Market according to the Efficient Market Hypothesis

of modern Finance. It can be said that the Market in its Semistrong Form includes the Weak Form, while in its Strong Form it includes all the other two forms, as illustrated in Figure (1.3).

Chapter 2

Dealing with Uncertainty

2.1 Decision making under Certainty

Decision making in cases that lack of uncertainty is somekind obvious and expected. For example, the problem set of (A, B) actions is assumed and

If we decide on A we get 10 euros/ hour.
If we decide on B we get 20 euros/ hour.

In this situation option B is preferred. This choice seems quite negligible as people tend to assume that everyone would prefer more money to less. But the same choice problem given in a different way may have a different answer. For example, actions (A^*, B^*) are now assumed:

If we decide on A^* we spend 10 euros/ hour.
If we decide on B^* we spend 20 euros/ hour.

In this case, A^* and B^* are two production schedules and less money is preferred to more, as the problem is one of production cost minimization. In more general situations, there would be a continuum of production schedules described by a production function $f(x)$, where x is a vector of production inputs. Given a vector of matching input prices w and a target production level y_0 , the problem of cost minimization could be written as

$$\min_x w'x \tag{2.1}$$

such that

$$f(x) = y_0.$$

The problem of cost minimization enables us to overcome the difficulty of describing a production procedure just by a production function. The production procedure can now be explained accurately using conceptually simple programming problem, that can, in principle, be solved using known mathematical methods.

2.2 Decision making under Uncertainty

Under uncertainty, decision making is even more difficult. The theory of choice under uncertainty was formally introduced in *Theory of Games and Economic Behaviour* by John von Neumann and Oskar Morgenstern (1944).

A lottery (or gamble) with n possible outcomes can be described by the pair (x, p) , where $x = (x_1, x_2, ..., x_n)$ is a vector of payoffs, one for each of the n possible outcomes, and $p = (p_1, p_2, ..., p_n)$ is a vector of probabilities associated with the payoffs x. For example, a simple lottery that pays x_1 with probability p_1, and x_2 with probability $1 - p_1$, may be denoted by the pair $(x_1, x_2), (p_1, 1 - p_1)$. In the case that the random experiment described by the lottery produces outcomes over a segment of the real line, the payoff x is a continuous random variable and p is a probability density function, usually denoted by f, so in the continuous case we denote the lottery by the pair (x, f).

A major step towards the development of modern probability theory occurred in the 17th century when mathematicians assumed that the value ξ of a gamble (x, p) is given by its **mathematical expectation**

$$\xi = \mu \equiv \sum_{i=1}^{\infty} p_i x_i. \tag{2.2}$$

The meaning of the mathematical expression above is that the expectation of a random variable represents the long-run payoff to a person that takes the gamble many times. The type of the gamble, though, defines the long-run payoff. In a **fair gamble** the person should have a zero long-run gain from playing the gamble many times, while in a **biased gamble** the person could make or loose money in the long-run depending on whether the bias was to his advantage or not. Thus, in a biased in-his-favor gamble, a person should be willing to pay for the chance of taking this gamble exactly as much money as he would be able to win by playing the lottery many times. If the lottery is biased against the player, on the other hand, then he should be paid in order to play.

$$fair\ gamble = \sum_{i=1}^{\infty} p_i x_i = 0$$

Assuming that the value ξ of a gamble equals its mathematical expectation seems quite logical at first sight. But a re-examination of the problem reveals that the mathematical expectation neglects a really important aspect of a person's reaction; it does not take into account the natural aversion that people show towards risk. Thus, it would be reasonable to say that anyone would be willing to forego the gamble by being offered a certain amount less than the gamble's mathematical expectation, the difference between the two being the price we pay to avoid taking a risk. In the same way, one could easily imagine a situation in which a very poor person could be persuaded to forego the gamble by being offered an amount that is much less than its mathematical expectation, while a richer person could only be persuaded to forego the lottery with an amount closer, although, less than its mathematical expectation.

These reactions to gambles were examined in the **St. Petersburg paradox**. The St. Petersburg paradox, which arose by a coin flipping game named *The St. Petersburg Lottery*, was first proposed by the mathematician Nicolas Bernoulli but published by his brother Daniel Bernoulli in the St. Petersburg Academy Proceedings (1738). This application is actually the first to import the concept of the **Expected Utility function**.

2.2.1 The St. Petersburg Lottery

The St. Petersburg lottery game is a fair coin flipping game. A coin is flipped until heads occur. The winning prize is determined by the number of flips i, where $(i = 1, 2, 3...)$, and it is equal to 2^i - ducats according to Bernoulli, let us use euros now. Thus, if heads occur first, the prize is $2^1 = 2$ euros and the game stops. If tails occur first, the coin is flipped again until heads appear. The probability of the heads' appearance after i tosses equals to:

$$\begin{aligned} \text{P(first Heads on trial } i) &= \text{P(Tails on trial 1)} \times \text{P(Tails on trial 2)} \times ... \times \\ &\quad \text{P(Tails on trial } i-1) \times \text{P(Heads of trial } i) \\ &= \frac{1}{2^i}. \end{aligned}$$

The table below (Table (2.1)) represents the sequence of the first 5 tosses. The sum of the expected gains of the outcomes is the expected value of the lottery.

TABLE 2.1: The St. Petersburg Lottery- First 5 tosses

i	Outcome	P(i)= $1/2^i$	Prize= 2^i	Expected gain= $prize * P(i)$
1	H	1/2	2€	1€
2	TH	1/4	4€	1€
3	TTH	1/8	8€	1€
4	TTTH	1/16	16€	1€
5	TTTTH	1/32	32€	1€

This lottery has a unique property: its mathematical expectation is infinite. Since it is theoretically possible that the game could go on for ever, the expected gain in this lottery is

$$\mu = \frac{1}{2} \times 2 + \frac{1}{4} \times 4 + \frac{1}{8} \times 8 + ... = 1 + 1 + 1 + ... = \sum_{i=1}^{\infty} \left(\frac{1}{2}\right)^i 2^i = \infty. \quad (2.3)$$

For a rational player, the game would be meaningful only if the entrance prize is smaller than the expected value of the outcomes. In the St. Petersburg lottery, the finite price of entry is always less than the expected value of the possible infinite outcomes. Thus, we would expect that everyone would enter the game, no matter the entrance prize. Experiments, though, have shown that, a rational player is willing to forego the St. Petersburg lottery for a certain amount of money, trading, in that way, a possible high pay-off with a secure money prize. Experiments have also shown that this amount of money is below 10 €. So there is a **"paradox"** here, i.e., a divergence between the then accepted theory of valuation of gambles and the common sense. Bernoulli set out to resolve this paradox.

Bernoulli supported that a lottery should be worth as much as a person should be willing to accept to forego the lottery. This number is denoted by ξ and is known as the **certainty equivalent of the lottery**. Bernoulli also argued that an amount of cash should be worth more to a poorer person than to a richer one, so the relevant quantity is not wealth per se but the utility for it. In particular, letting $U(.)$ be the individual's utility for wealth and W be his current level of wealth, he argued that if ξ is a fixed amount of money and $W_1 < W_2$ then $U(W_1+\xi)-U(W_1)$ should be greater than $U(W_2+\xi)-U(W_2)$.

According to Bernoulli, an individual with utility for wealth function $U(.)$ and current level of wealth W is connected to a lottery's value equal to the number ξ for which

$$U(W+\xi) = \sum_{i=1}^{\infty} p_i U(W+x_i). \tag{2.4}$$

For $U(.)$ linear, it holds $\xi = \sum_{i=1}^{\infty} p_i x_i$, so that the value of the lottery is equal to its mathematical expectation only under special circumstances. But for $U(.)$ concave, which is the most reasonable specification, the answer differs. For example, taking $U(x) = \ln(x)$ and an initial wealth of $W = 20{,}000$ €:

$$\sum_{i=1}^{\infty} p_i U(W+x_i) = \sum_{n=1}^{\infty} \left(\frac{1}{2}\right)^i \ln(20,000+2^i)$$

$$= \left(\frac{1}{2}\right) \ln(20,001) + \left(\frac{1}{4}\right) \ln(20,002) + \left(\frac{1}{8}\right) \ln(20,004) + ...$$

$$\approx 9.903893,$$

so the certainty equivalent of the St. Petersburg lottery solves

$$\ln(20000+\xi) = 9.903893,$$

from which we get

$$\xi = \exp(9.903893) - 20000 \approx 8.1\, e. \tag{2.5}$$

For an initial wealth of $W = 50{,}000$ € , we obtain an increased certainty equivalent of $\xi \approx 8.8$ €. If, instead, $U(x) = \sqrt{x}$ is used, for $W = 20{,}000$ € we get $\xi \approx 9.1$ € while for $W = 50{,}000$ € we get $\xi \approx 9.8$ €. The result of this mini experiment is that the value of this lottery, for people with various current wealth and different utility for wealth functions, should always be around 10 €.

In 1944, John von Neumann and Oskar Morgenstern in *Theory of Games and Economic Behavior*, using the utility function for wealth of Bernoulli, came up with the **Expected Utility function** over lotteries (or gambles). The definition given for lotteries by von Neumann and Morgenstern is that lotteries are a probability distribution over a specific and finite set of outcomes.

2.3 The Preference Axioms

A utility function over lotteries (or gambles) is constructed according to some hypothesis about an individual's preferences. The binary preference relations are:

- $>$: is strictly preferred to,
- $<$: is strictly dis-preferred to,
- $\geq$: is weakly preferred to,
- $\leq$: is weakly dis-preferred to,
- $\sim$: is indifferent to.

von Neumann and Morgenstern (1947) proposed a set of primitive **axioms** that are reasonable and intuitive and proceeded to prove that any person obeying them should be an expected utility maximizer. Savage (1954) demonstrated the same axioms, providing the fundamentals for Subjective Expected Utility model (SEU). Savage, for technical reasons, introduced three more axioms, *Non-degeneracy, Small Event Continuity* and *Uniform Monotonicity*, which do not have the substantive weight of the first four axioms.

Theorem 1 (**von Neumann-Morgenstern Axioms**). *Assume Ω is the set of alternatives over which a player has preferences and x, y, z the alternatives.*

1. ***Completeness:*** For any preference alternatives x and $y \in \Omega$, either $x \geq y$ or $x \leq y$. If both are true then $x \sim y$. We suppose that an individual always adopts a certain attitude towards 2 alternatives, or is equally attracted to them, therefore indifferent.

2. ***Transitivity:*** For any tree preference alternatives x, y, $z \in \Omega$, if $x \geq y$ and $y \geq z$, then $x \geq z$. An individual's preferences are internally consistent. We suppose that the choice maker is rational so he would not break the "chain" of his own preferences unreasonably.

3. ***Continuity:*** For any x, y, $z \in \Omega$ given $x > y > z$, there is an α,$\beta \in (0, 1)$ such that $\alpha x + (1 - \alpha)y > z$ and $y > \beta x + (1 - \beta)z$.

4. ***Independence:*** For any x, y, $z \in \Omega$ and $\alpha \in (0, 1)$, if $x > y$ then $\alpha x + (1 - \alpha)z > ay + (1 - \alpha)z$.

An individual's behavior satisfying these four axioms indicates that he/she has a certain utility function. Preferences of this individual can be represented on an interval scale and he/she is considered to be a utility maximizer.

Based on the ideas of de Finetti (1937) and von Neumann and Morgenstern (1947), Savage (1954) proposed the first complete axiomatic Subjective Expected Utility Theory. He introduced necessary and sufficient conditions for the utility and probability, as well as the examination of choice under uncertainty as an Expected Utility maximizing behavior. In Savage's approach probability does not appear as a primitive concept, is not an objective property, but the decision maker's **subjective assessment of the likelihood** of various events.

Theorem 2 (**Savage**). *A preference $\geq$ satisfies axioms 1-4 if and only if there is a finitely additive probability measure P and a function $u : C \to R$ such that for every pair of acts F and G*

$$F = \int_{\Omega} u(F(\omega))dP \geq \int_{\Omega} u(G(\omega))dP = G,$$

where P is unique and u is unique up to positive affine transformation.

According to Savage, an individual who satisfies the Axioms reduces the uncertainty to a subjective probability measure P which reflects his beliefs. Outcomes are perceived based on the individual's utility function u and acts based on the **Expected Utility**, where the Expected Utility of an act F/G is the probability-weighted average of the utilities of F's/G's consequences. Through his work, Savage established the fundamentals for Bayesian statistics and its application to Game Theory.

Chapter 3

Expected Utility Maximization and Risk

A milestone in choice under uncertainty is the Expected Utility maximization, as proposed by Bernoulli (1738) and expanded by von Neumann & Morgenstern (1947), as well as other significant scientists like Ramsey (1931), de Finetti (1937) and Savage (1954). When decision making includes large money amounts, as it happens in the Financial Markets, then the efficiency of the Expected Utility Theory in risk assessment and portfolio allocation is quite challenging. At first, Expected Utility maximization seems adequate; deviations from Expected Utility towards better risk assessment will be discussed later (Chapter (6)).

3.1 Expected Utility

The Expected Utility is calculated by taking the weighted average of all possible outcomes, with the weights being assigned by the probability that any particular event could occur.

Theorem 3 (**Expected Utility Theorem**). *Consider a gamble with* $a_1, a_2, ..., a_i$ *possible outcomes and* $p_1, p_2, ..., p_i$ *the possibilities of these outcomes to occur. The Expected Utility of this gamble is:*

$$EU(x) = p_1 U(a_1) + p_2 U(a_2) + ... + p_i U(a_i), \tag{3.1}$$

where $U(.)$ the individual's utility of an outcome a_i.

If the gamble has continuous outcomes then the initial definition of the Expected Utility is transformed as follows:

$$E_f U(x) = \int_{-\infty}^{\infty} U(x) f(x) dx, \tag{3.2}$$

where $U(x)$ is the utility function of an individual. $F(x)$ is the c.d.f. of the probability distribution x and it holds

$$\int_{-\infty}^{\infty} f(x) dx = 1. \tag{3.3}$$

An individual with utility function prefers a gamble (x, f) to another gamble (y, g) if and only if

$$E_f U(x) \equiv \int_{-\infty}^{\infty} U(x) f(x) dx > \int_{-\infty}^{\infty} U(y) g(y) dy \equiv E_g U(y). \tag{3.4}$$

The individual would still prefer (x, f) to (y, g) if $U(.)$ was to change by an affine transformation, i.e., be replaced by $\alpha + \beta U(.)$, so $U(.)$ is determined up to location α and scale β . In this case $U(.)$ is a cardinal utility function.

von Neumann and Morgenstern (1947) provided an axiomatization of the Expected Utility Theory in an appendix of their landmark book *Theory of Games and Economic Behavior* (Section (2.3)). To honor their contribution to the mathematical foundations of the theory, the utility for wealth function is usually referred to as the von Neumann-Morgenstern utility function. Expected Utility Theory is widely used in theoretical and practical analysis. Even though, in many cases it has been proven that people break the behavioral axioms- *see the Appendix (C) for the Allais Paradox and the Ellsberg Paradox.*

Bernoulli, in his logarithm function, which is concave, argued that people with various current wealth and utility for wealth functions have a certain prize for entering the lottery. Bernoulli supported that the former is logical as people tend to be **risk-averters**, preferring certain outcomes over uncertain or less certain ones.

3.1.1 Expected Utility Maximaziation

An individual aims to maximize his Expected Utility

$$E_f U(x) = Maximized.$$

This results from the hypothesis that the utility increases when an increase of wealth takes place. Mathematically, this means that the first partial derivative of the wealth utility function is positive:

$$\frac{\partial U(W)}{\partial W} > 0. \tag{3.5}$$

3.2 Risk

The part of the **return variance of a portfolio** that is non-market-related is the **Non-Systematic Risk**. Non-Systematic Risk can be reduced through portfolio diversification by the individual, but the total risk does not end up there. There is a risk that cannot be avoided, known as **Systematic Risk**. Systematic Risk is irrelevant to the portfolio diversification as it is the return variance of a portfolio caused by internal or external market fluctuations.

As seen in Figure (3.1), Non-Systematic Risk is reduced when an increase of asset number occurs. This means that an increased diversification, which is managed through adding more assets in the portfolio, leads to lower levels of Non-Systematic Risk. But, there is an asymptotic limit of Non-Systematic

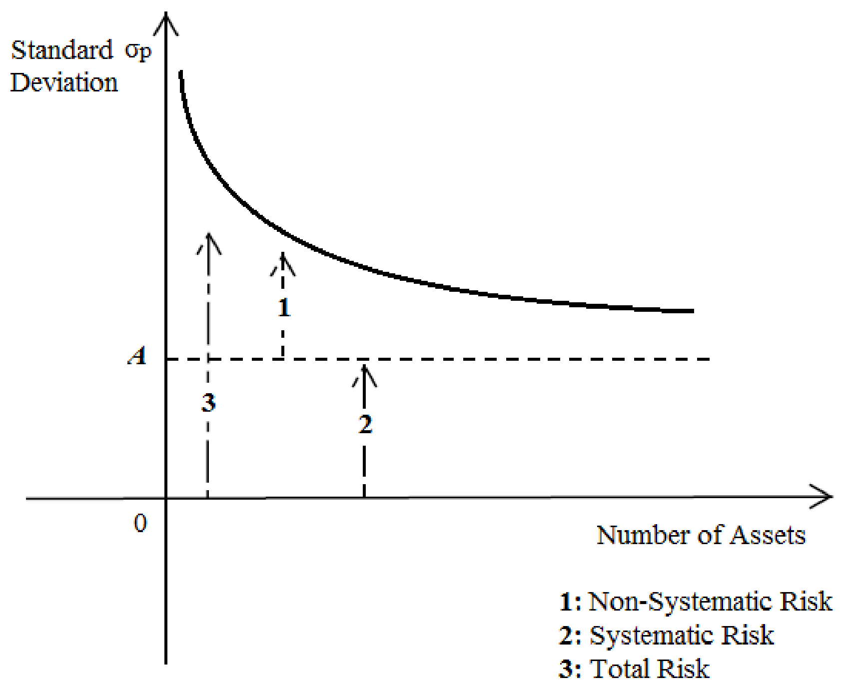

FIGURE 3.1: Systematic, Non-Systematic and Total Risk

Risk (Figure (3.1), line A), beyond which the individual can not further affect it. This is the exact point where Systematic Risk appears. Systematic Risk is the market risk, thus it cannot be reduced through diversification by the individual. It depends on the development level of the financial system. So, it is the financial system which manages an auto-diversification by attempting to reduce every kind of impact, either internal or external.
There are numerous risks to be considered when a portfolio is constructed. Some of these risks are:

- **Investment market risk**: The possibility all investments in a market sector, (such as shares), will be affected by an event.
- **Investment specific risk**: The possibility a particular investment may underperform the market or its competitors.
- **Market timing risk**: The possibility your investment may be sold at a time when the sale price is at a low-point or purchased when the sale price is at a high-point.
- **Inflation risk**: The possibility your investment return is below the inflation rate which reduces the spending power of your money.
- **Credit risk**: The potential failure of a debtor to make payments on amounts they have borrowed.
- **Interest rate risk**: The possibility your investment will be adversely impacted by a fall or rise in interest rates.

- **Legislative risk**: The possibility a change in legislation will impact the appropriateness of certain investments for you.
- **Liquidity risk**: The ease with which you can sell or liquidate your investments. Some investments impose exit fees or have limitations on withdrawals. Other investments may be difficult to sell due to a lack of buyers.
- **Hedging risk**: A technique designed to reduce the risk from part of an investment portfolio often by using derivatives. While hedging can reduce losses, it also has a cost and therefore can reduce profits.
- **Currency risk**: Relates to global investments. It is a form of risk that arises from the change in price of one currency against another. Whenever investors or companies have assets or business operations across national borders, they face currency risk if their positions are not hedged.
- **Derivatives risk**: Where financial derivatives are used as an alternative to directly owning or selling underlying assets in order to manage risk and/or enhance returns. Risks associated with derivatives can include; the value of the derivative declining to zero; the value of the derivative not moving in line with the underlying asset and, the derivative may be difficult or costly to reverse, and
- **Opportunity cost**: The investment return you may forego from an asset as a result of investing in your preferred asset. That is, there is a risk the preferred asset you invest in may not return more than the second-choice (next best alternative) asset you did not invest in. (GWM Adviser Services Limited , 2009)

A portfolio contains a large number of assets, so that the variance remains reduced. Both portfolio return and variance are strongly connected to economic events. Historically, it has been observed that in periods of economic recession, return levels are low while variance levels are high. Vice versa, during periods of economic blossom, variance levels are low whereas return levels are high. Figure (3.2) illustrates a market portfolio and its relation to market risk. It can be observed that, as we move to higher risk levels, expected return levels increase. The risk an investor undertakes is compensated with better payoffs. An individual aims to reduce risk over a certain amount of return or to increase return over a certain amount of risk. Nevertheless, individuals' **attitudes toward risk** may present severe differences.

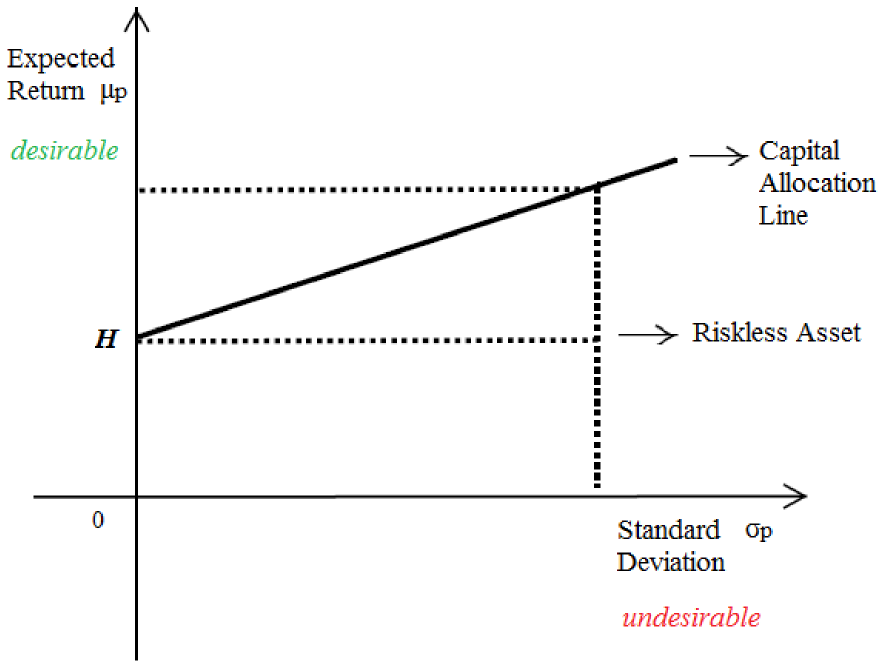

FIGURE 3.2: Portfolio and Risk: Capital Allocation Line

3.3 Attitudes toward Risk

The von Neumann-Morgenstern utility function, unlike the Bernoulli's concave one, indicates that people may have different attitudes toward risk, no matter their former overall behavior. The shape of a person's utility function reveals his attitude toward risk (Figure (3.3)):

- A **risk averter** individual has a concave utility function and it holds that the second derivative equals to $U_a''(W) < 0$. Expected utility increases with a decreasing rhythm when the expected return increases.
- A **risk neutral** individual has a linear utility function and it holds that the second derivative equals to $U_n''(W) = 0$. The individual is indifferent to the extra utility that an increase in expected return offers.
- A **risk seeker** individual has a convex utility and it holds that the second derivative equals to $U_b''(W) > 0$. Expected utility increases with a rising rhythm when the expected return increases.

The different types of investors can also be illustrated as a relation between expected returns and risk (Figure (3.4)). In terms of Figure (3.4), for the risk averter it holds that $\frac{\partial U}{\partial \sigma} < 0$ and $\frac{\partial U}{\partial \mu} > 0$, while for the risk seeker it holds that $\frac{\partial U}{\partial \sigma} > 0$ and $\frac{\partial U}{\partial \mu} < 0$.

The most commonly used function for Expected Utility is the **quadratic utility function**

$$U(W) = aW - bW^2, \tag{3.6}$$

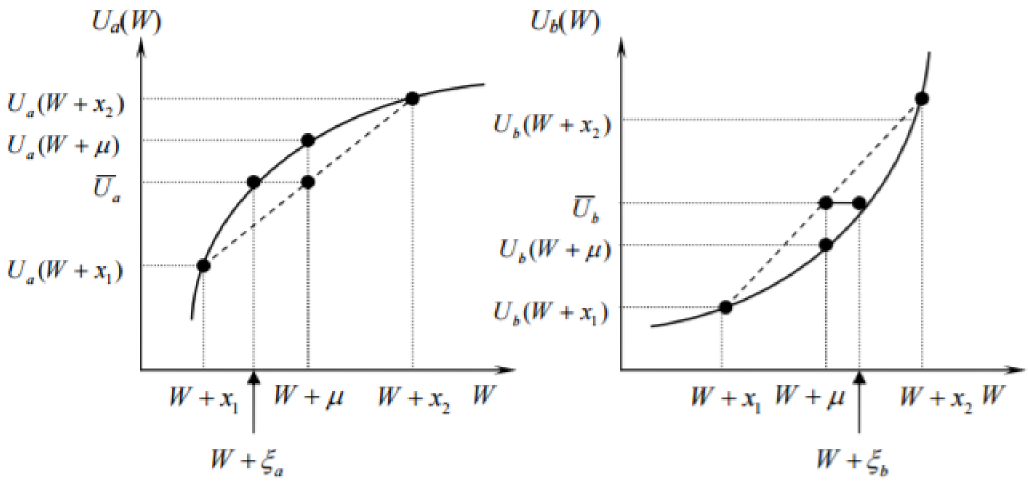

FIGURE 3.3: (a) The Concave Utility of a Risk Averter
(b) The Convex Utility of a Risk Seeker

where $a, b > 0$ and $W < a/2b$. The first and the second derivatives of the function are:

$$U'(W) = a - 2b \qquad \textit{and} \qquad U''(W) = -2b.$$

The individual's function, in order for him to be a risk averter, should have a positive second derivative $U''(W) < 0$. Thus, for a risk averter individual with a quadric utility function, it holds that b>0.

Moreover, it is important to know how much averse to risk an individual is. To this effect, there is a set of tools to measure risk. A **risk measure** is a function that is used to quantify risk. A risk measure is meant to determine the quantity of an asset (or set of assets) to be kept in reserve. The aim of a reserve is to guarantee the presence of capital that can be used as a (partial) cover if the risky event manifests itself, generating a loss. From a mathematical point of view, a measure of risk is a scalar function $\rho : X \to R$ mapping the space of random variables X to the set of real numbers R.

The most frequently used measures of risk aversion are the **Arrow-Pratt measures** of absolute and relative risk-aversion. These risk measures are named after John W. Pratt's paper (1964) and Kenneth J. Arrow's one (1965). Arrow and Pratt showed that the degree of concavity of a utility function reveals the grade of a certain individual's risk aversion. The coefficient of absolute risk aversion is

$$A(W) = -\frac{U''(W)}{U'(W)} \tag{3.7}$$

and it has 3 possible outcomes:

$$A'(W) < 0, \quad \textit{decreasing aversion to the risk}$$

$$A'(W) = 0, \quad \textit{steadfast aversion to the risk}$$

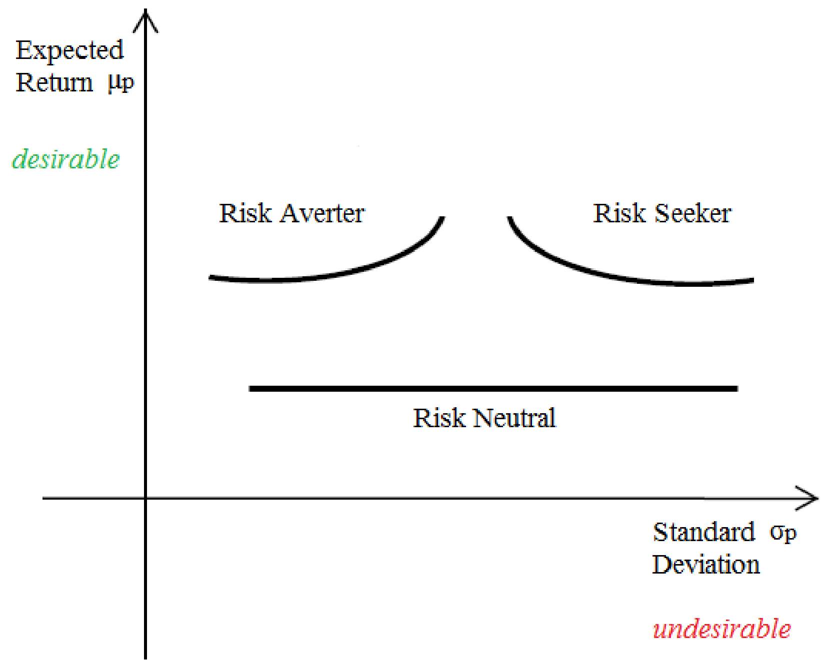

FIGURE 3.4: Attitudes towards Risk: Indifference curves in the mean-s.d. space

$$A'(W) > 0, \quad \textit{increasing aversion to the risk.}$$

The coefficient of relative risk aversion is

$$R(W) = -x\frac{U''(W)}{U'(W)}, \tag{3.8}$$

where x is the payoff of a given lottery and $U(W)$ the utility derived from that payoff. In Finance, there are more statistical methods used to measure and quantify the level of risk within a portfolio that will be analyzed in the next Chapters.

Chapter 4

Mean-Variance Portfolios

Harry Markowitz was the first to develop the **Modern Portfolio Theory** in 1950. Markowitz was a PhD student at the University of Chicago when he published a paper on portfolio allocation (1952). In this paper, Markowitz was analyzing the ways investors should use in order to choose portfolios optimally. More specifically, the theory he proposed concerns how investors should allocate their money on different assets, based on the asset returns and risk levels. In this framework, the risk is defined as the variance of the portfolio. For his contribution in Portfolio Theory, Markowitz later earned the Nobel Prize in Economics, jointly with W. Sharpe and M. Miller.

Since Markowitz proposed his pioneering theory, the variance has been the main risk measure in Economics and Finance. Mean-Variance analysis, though, presents some certain characteristics which automatically make it a weak risk measure. In order for Mean-Variance analysis to be consistent with Expected Utility, returns must be normally distributed and/or the utility function used must be of a quadratic form. It also penalizes upside and downside results and is unable to examine return skewness and kurtosis, failing to describe the risk of low probability events.

4.1 Mean-Variance Portfolios from Risky Assets

The most important contribution of Markowitz (1952) is that it is favorable to diversify a portfolio because this will reduce the portfolio's standard deviation (risk), as long as the correlation between assets is less than 1. This result can be shown by a portfolio of M assets. Assume M risky assets $i = (1, 2, 3, ...M)$ and let $R = [R_1, R_2, ...R_M]$ be the vector of the risk returns. Assets are normally distributed with Mean and Variance-Covariance of returns given by:

$$E[R] = a = \begin{bmatrix} a_1 \\ . \\ . \\ . \\ a_M \end{bmatrix} \qquad and \qquad Cov[R] = \Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} & ... & \sigma_{1M} \\ \sigma_{21} & \sigma_{22} & ... & \sigma_{2M} \\ . & . & ... & . \\ . & . & ... & . \\ \sigma_{M1} & \sigma_{M2} & ... & \sigma_M^2 \end{bmatrix}.$$

The M-vector of weights is

$$w = (w_1, w_2, ...w_M) = \sum_{i=1}^{M} w_i = 1$$

and equals to the sum vector $1 = (1, 1, ...1)$. The M-vector of weights indicates the wealth of each asset.
The portfolio return is given by

$$R_P = w'R = \sum_{i=1}^{M} w_i R_i$$

where

$$\mu_P \equiv w'\mu = \sum_{i=1}^{M} w_i \mu_i,$$

$$\sigma_P^2 = var[R_P]w'\Sigma w.$$

The **minimum variance portfolio** with expected return μ_P is the solution to the problem:

$$\min_w \frac{1}{2}\sigma_P^2 = \frac{1}{2}w'\Sigma w \tag{4.1}$$
$$\text{s.t. } w'1 = 1, \qquad w'\mu = \mu_P.$$

Lemma 4 (Mean-Variance Efficient Portfolio). *A portfolio w^* is Mean-Variance Efficient if it does not exist any other portfolio w that has an equal or higher return and a lower variance:*

$$w'\mu \geq w^{*\prime}\mu \qquad \textit{and} \qquad w'\Sigma w < w^{*\prime}\Sigma w^*$$

Lagrangian:

$$\min_{w,\lambda_1,\lambda_2} L = \frac{1}{2}w'\Sigma w + \lambda_1(1 - w'1) + \lambda_2(\mu_P - w'\mu). \tag{4.2}$$

The first order conditions:

- $$\frac{\partial L}{\partial w} = \sum w - \lambda_1 1 - \lambda_2 \mu = 0$$
- $$\frac{\partial L}{\partial \lambda_1} = 1 - w'1 = 0$$
- $$\frac{\partial L}{\partial \lambda_2} = \mu_P - w'\mu = 0.$$

Solving for w^*:

$$w^*(\mu_P) = \lambda_1 \sum^{-1} 1 + \lambda_2 \sum^{-1} \mu = 1. \tag{4.3}$$

Solving for λ_1, λ_2 we get the system:

$$\lambda_1 1' \sum^{-1} 1 + \lambda_2 \mu' \sum^{-1} 1 = 1$$

$$\lambda_1 1' \sum^{-1} \mu + \lambda_2 \mu' \sum^{-1} \mu = \mu_P,$$

where

$$\lambda_1 = \frac{(\mu' \sum^{-1} \mu) - (1' \sum^{-1} \mu)\mu_P}{(1' \sum^{-1} 1)(\mu \sum^{-1} \mu) - (1' \sum^{-1} \mu)^2} = \frac{\Gamma - B\mu_P}{\Delta}$$

and

$$\lambda_2 = \frac{(1' \sum^{-1} 1)\mu_P - (1' \sum^{-1} \mu)}{(1' \sum^{-1} 1)(\mu \sum^{-1} \mu) - (1' \sum^{-1} \mu)^2} = \frac{A\mu_P - B}{\Delta}.$$

Given the λ_1 and λ_2 values, the (4.3) has the variance:

$$\begin{aligned} \sigma^2(\mu_P)' &= w(\mu_P)\Sigma w(\mu_P) \\ &= w(\mu_P)'\Sigma(\lambda_1 \Sigma^{-1} 1 + \lambda_2 \Sigma^{-1} \mu) \\ &= \lambda_1 w(\mu_P)' 1 + \lambda_2 w(\mu_P)' \mu \\ &= \lambda_1 + \lambda_2 \mu_P \\ &= \frac{A\mu_P^2 - 2B\mu_P + \Gamma}{\Delta}. \end{aligned} \tag{4.4}$$

Equation (4.4) is a **parabola**.

After examining the first and the second derivatives of $\sigma_P^2(\mu_P)$

$$\frac{d\sigma_P^2(\mu_P)}{d\mu_P} = \frac{2(A\mu_P - B)}{\Delta} = 0 \quad \textit{and} \quad \frac{d^2\sigma_P^2(\mu_P)}{d\mu_P} = \frac{2A}{\Delta} > 0 \tag{4.5}$$

it is revealed that $\sigma_P^2(\mu_P)$ is a strictly convex function of μ_P with minimum

$$\frac{d\sigma_P^2(\mu_P)}{d\mu_P} = \frac{2(A\mu_P - B)}{\Delta} = 0 \Leftrightarrow \mu_P = \frac{B}{A}. \tag{4.6}$$

Examining the first and the second derivatives of

$$\sigma_P(\mu_P) = \sqrt{\frac{A\mu_P^2 - 2B\mu_P + C}{\Delta}} \tag{4.7}$$

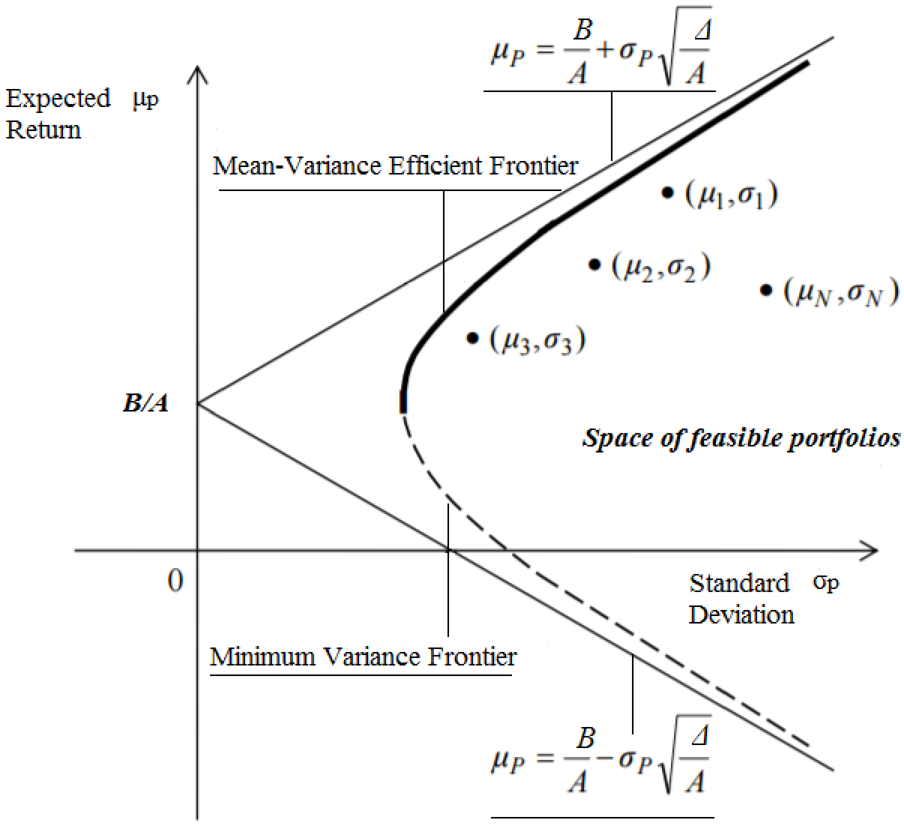

FIGURE 4.1: Portfolio Frontier with Risky assets

we see that σ_P is also a strictly convex function of μ_P:

$$\frac{d\sigma_P(\mu_P)}{d\mu_P} = \frac{A\mu_P - B}{\Delta\sigma_P} \qquad \text{and} \qquad \frac{d^2\sigma_P(\mu_P)}{d\mu_P} = \frac{1}{\Delta\sigma_P^3} > 0. \tag{4.8}$$

Equation (4.7) is a **hyperbola**. Figure (4.1) is the graphic representation of equation (4.7). Every σ_P corresponds to two different feasible portfolios, each of them connected to a different return level. A rational individual would prefer the portfolio with the higher return than the one with the lower return, provided the same risk level. This leads to the fact that the **Efficient Portfolio Frontier** is the upper part of the hyperbola. The asympotes of the Efficient Frontier are given by: $\mu_P = \frac{B}{A} \pm \sigma_P\sqrt{\frac{\Delta}{A}}$. The returns and standard deviation of individual assets $(\mu_{Ri}, \sigma_{Ri}), i = (1, ..., M)$ belong to the **Space of Feasible Portfolios**.

4.2 Mean-Variance Portfolios that include a Riskless Asset

In this section a riskless asset, that is, an asset with guaranteed return R_f and zero variance is included. As seen in Section (4.1), M risky assets are assumed with returns $R \sim Normal_M(\mu, \Sigma)$, and the riskless asset as the 0^{th} asset is added. The minimum variance portfolios in this case solves the optimization problem

$$\min_{w} \frac{1}{2} w' \Sigma w \tag{4.9}$$
$$\text{s.t. } w'(\mu - R_f 1) = \mu_P - R_f.$$

A comparison between this new problem and the problem of the portfolio constructed from Risky Assets only, reveals that the *summing-up* constraint $w'1 = 1$ does not apply in this case. Also, the *required return* constraint $w'\mu = \mu_P$ has been changed. The reason why there is no *summing-up* can be explained based on the idea of the optimization problem. The optimization problem is written only in terms of the N risky assets so once these weights $w = (w_1, w_2, ..., w_M)'$ are chosen we can always choose the weight of the riskless asset to be $w_0 = 1 - w'1$. For similar reasons, the *required return* constraint has been modified to a **required excess-return constraint**, that is return in excess of the risk-free rate R_f.
Lagrangian of the riskless asset problem:

$$\min_{w,\lambda} L = \frac{1}{2} w' \Sigma w + \lambda[\mu_P - R_f - w'(\mu - R_f 1)]. \tag{4.10}$$

The first order conditions:

- $$\frac{\partial L}{\partial w} = \Sigma w - \lambda(\mu - R_f 1) = 0$$
- $$\frac{\partial L}{\partial \lambda} = \mu_P - R_f - w'(\mu - R_f 1) = 0.$$

Solving for w:

$$w^* = \lambda \Sigma^{-1}(\mu - R_f 1). \tag{4.11}$$

Substituting (4.11) into $\frac{\partial L}{\partial \lambda}$ the following is obtained:

$$\begin{aligned}
\mu_P - R_f &= w^*(\mu - R_f 1) \\
&= \lambda(\mu - R_f 1)' \Sigma^{-1} (\mu - R_f 1) \\
&= \lambda[\mu' \Sigma^1 \mu - R_f \mu' \Sigma^{-1} 1 - R_f 1' \Sigma^{-1} \mu + R_f^2 1' \Sigma^{-1} 1] \\
&= \lambda[\Gamma - 2 R_f B + R_f^2 A]
\end{aligned}$$

or

$$\lambda = \frac{\mu_P - R_f}{\Gamma - 2R_f B = R_f^2 A}. \tag{4.12}$$

Plugging this into (4.11) we get the optimal portfolio weights for the risky assets

$$w^*(\mu_P) = \frac{\mu_P - R_f}{\Gamma - 2R_f B + R_f^2 A}\Sigma^{-1}(\mu - R_f 1) \tag{4.13}$$

while, the weight received by the riskless asset is

$$w_0^* = 1 - 1'w.$$

The equation of the minimum-variance set is

$$\begin{aligned} \sigma^{2*} &= w^{*\prime}\Sigma w^* \qquad (4.14)\\ &= w^{*\prime}\lambda\Sigma\Sigma^{-1}(\mu - R_f 1)\\ &= \lambda w^{*\prime}(\mu - R_f 1)\\ &= \lambda(\mu_P - R_f)\\ &= \frac{(\mu_P - R_f)^2}{\Gamma - 2R_f B + R_f^{2A}} \end{aligned}$$

where, in the fourth line of the above display we have used the *required excess return* constraint. Solving (4.14) for μ_P we obtain

$$\mu_P = R_f \pm \sigma_P\sqrt{\Gamma - 2R_f B + R_f^2 A}. \tag{4.15}$$

Figure (4.2) presents the **Minimum-Variance Frontier**.

Define excess returns by $\tilde{R} = R - R_f 1$ with mean $\tilde{\mu} \equiv \mu - R_f 1$ and variance Σ, i.e.,

$$\tilde{R} \sim Normal_M(\tilde{\mu}, \Sigma).$$

In terms of these excess returns, the portfolio optimization problem may be written as

$$\min_w \frac{1}{2} w'\Sigma w$$

$$\text{s.t. } w'\tilde{\mu} = \tilde{\mu_P}$$

where, $\tilde{\mu}$ is the required excess return of the portfolio. The solution of this problem in (4.11) may be rewritten in the new notation as

$$w^* = \lambda\Sigma^{-1}\tilde{\mu}, \tag{4.16}$$

where λ is given in (4.12).

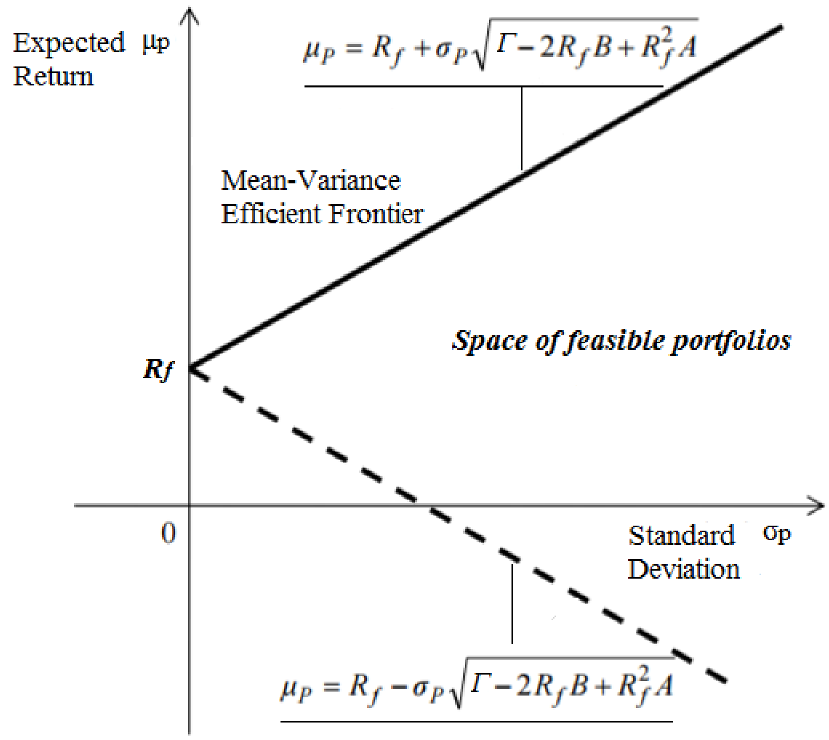

FIGURE 4.2: Minimum Variance Frontier

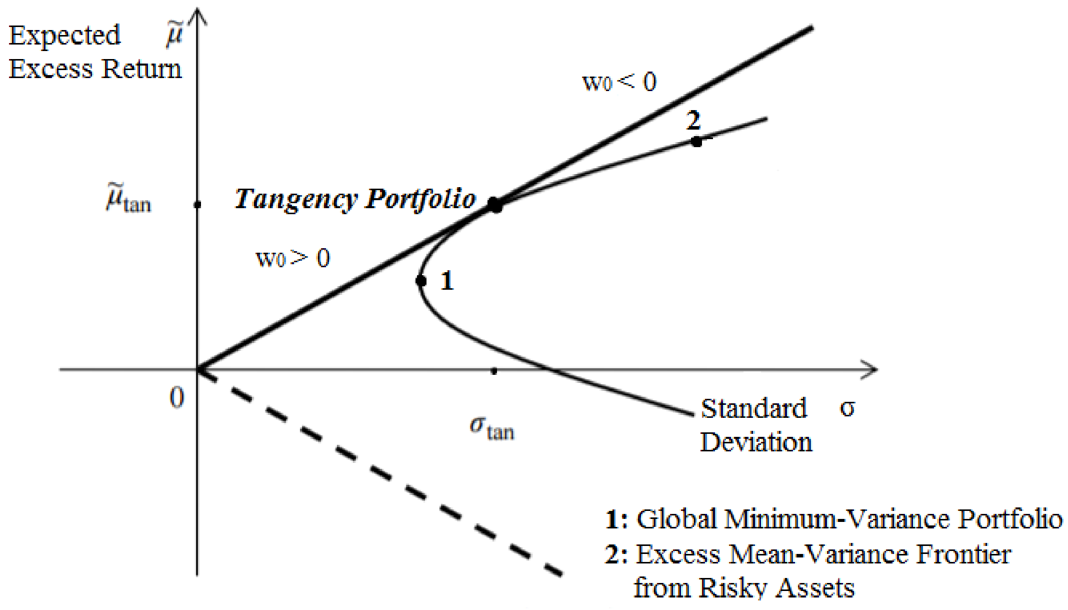

FIGURE 4.3: Excess Mean-Variance Frontier

For $\lambda = 1/(1'\Sigma^{-1}\tilde{\mu})$, a value that λ is bound to take for some required portfolio return μ, we obtain the **tangency portfolio**

$$w_{tan} = \frac{\Sigma^{-1}\tilde{\mu}}{1'\Sigma^{-1}\tilde{\mu}}. \tag{4.17}$$

As shown in Figure (4.3), the point in $(\tilde{\mu}, \sigma)$ space that corresponds to the tangency portfolio is the only point at which the Excess Mean – Variance Frontier from the risky assets only touches the Excess Mean – Variance Frontier that includes a risk-free asset.

By this, we conclude that any optimal portfolio is a linear combination of the risk-free asset and the tangency portfolio. Any investor faced with the problem of choosing an optimal portfolio, first computes the tangency portfolio for the risky-assets-only problem, and then decides how to split his total capital between this portfolio and the risk-free asset. The point $(\tilde{\mu}_{tan}, \sigma_{tan})$ in Figure (4.3) is obtained if the investor decides to invest all his capital on the tangency portfolio, i.e., if $w_0 = 0$ and $w^* = w_{tan}$. The rest of the points on the Excess Mean – Variance Frontier can be obtained either by **lending** at the risk-free rate (depositing the money at a bank account that gives return equal to the risk-free rate), in which case $w_0 > 0$ and $w^* = (1 - w_0)w_{tan}$, or by **borrowing** at the risk free rate, in which case $w_0 < 0$ **and** $w^* = (1 - w_0)w_{tan}$, and investing the borrowed capital at the tangency portfolio. Investors who hold a positive amount of the risk-free asset, $w_0 > 0$, choose points on the Excess Mean – Variance Frontier that lie below $(\tilde{\mu}_{tan}, \sigma_{tan})$. On the other hand, investors who hold a negative amount of the risk-free asset, $w_0, 0$, choose points above $(\tilde{\mu}_{tan}, \sigma_{tan})$.

The investor's utility function provides all the information needed to specify

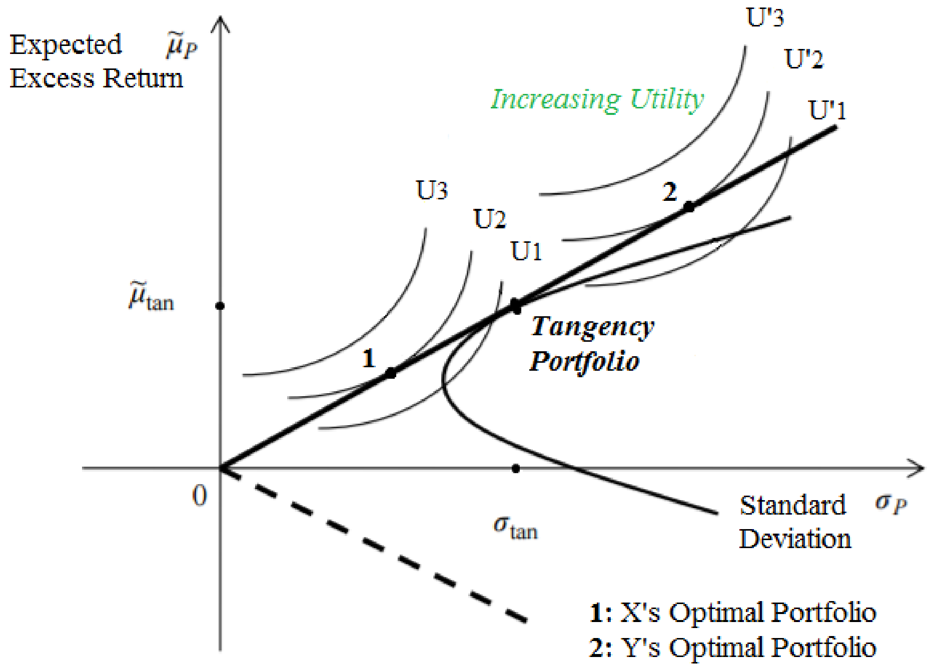

FIGURE 4.4: Excess Mean-Variance Frontier explained

the choice of an optimal portfolio among the set of efficient portfolios. Figure (4.4) depicts the choices of two investors X and Y and their utilities with risk aversion parameters θ_X and θ_Y, respectively, and $\theta_X > \theta_Y$, where investor X is more **risk averse** than investor Y. Investor X picks a positive position on the risk-free asset, i.e.,$w_0^X > 0$, while investor Y picks a riskier portfolio with a negative position in the risk-free asset, i.e., $w_0^Y < 0$.

Figures (4.3) and (4.4) present the tangency portfolio as touching the Excess Mean – Variance Frontier on its upper limb. If the return of the global minimum variance portfolio exceeds the risk-free return, i.e., if $B/A > R_0$, it is observed that the tangency is on the upper limb. If the reverse is true the tangency is on the lower part of the Frontier. The assumption that $B/A > R_0$ is therefore realistic.

4.3 Mean- Variance Analysis consistency with Expected Utility

But, is MV always functional? The answer is rather negative, since MV is consistent with the Expected Utility Theorem only under certain circumstances.

Theorem 5 (**MV Consistency with EU**). *Mean-variance analysis is consistent with Expected Utility maximization, i.e., there exists* $V(.)$ *such that* $V(\mu_R, \sigma_R^2) = E[U(R)]$, *if and only if either of the following conditions are satisfied:*

1. R is normally distributed (or, more generally, elliptically distributed).

2. U(.) takes the quadratic form, i.e.,
$$U(R) = aR - bR^2, a > 0, b > 0, R < a/2b \ .$$

Proof. (a) Expand $U(R)$ around $R = \mu_R$ to obtain

$$U(R) = U(\mu_R) + \tfrac{1}{1!}U^{(1)}(\mu_R)(R-\mu_R) + \tfrac{1}{2!}U^{(2)}\mu_R(R-\mu_R)^2$$

$$+\tfrac{1}{3!}U^{(3)}\mu_R(R-\mu_R)^3 + \tfrac{1}{4!}U^{(4)}\mu_R(R-\mu_R)^4 + ...$$

Taking expectations we obtain

$$E[U(R)] = U(\mu_R) + \tfrac{1}{1!}U^{(1)}(\mu_R)E[(R-\mu_R)] + \tfrac{1}{2!}U^{(2)}(\mu_R)E[(R-\mu_R)^2]$$

$$+\tfrac{1}{3!}U^{(3)}(\mu_R)E[(R-\mu_R)^3] + \tfrac{1}{4!}U^{(4)}(\mu_R)E[(R-\mu_R)^4] + ...$$

All odd central moments greater or equal to three of a normal random variable are zero, as are all even central moments that are greater than the fourth central moment. This means that for a normal random variable the only non-zero moments are the second $E[(R-\mu_R)^2)] = \sigma_R^2$ and the fourth $E[(R-\mu_R)^4] = 3\sigma_R^4$. Substituting these quantities into the equation above and dropping the zero terms we obtain

$$E[U(R)] = U(\mu_R) + \tfrac{1}{2}U^{(2)}(\mu_R)\sigma_R^2 + \tfrac{1}{8}U^{(4)}(\mu_R)\sigma_R^4$$

$$\equiv V(\mu_R, \sigma_R^2), \text{ as desired.}$$

(b) We have
$$E[U(R)] = E(aR - bR^2)$$

$$= aE(R) - bE(R^2)$$

$$= aE(R) - bVar(R) + E(R)^2$$

$$= a\mu_R - b\mu_R^2 - b\sigma_r^2$$

$$\equiv V(\mu_R, \sigma_R^2), \text{ as desired.}$$ □

The assumptions are restrictive as in real markets return distributions may be skewed and fat-tailed. Assume two assets with different return distributions, as seen in Figure (4.5). The Figure reveals that, although the two distributions share the same mean and variance, conceal different risk levels. If an individual could observe the two different distributions would prefer the right-skewed one (A) - which indicates gains - over the left-skewed one (B) - which indicates losses. But from the MV aspect, the two assets are the same, misleading us to believe that the two returns are identical.

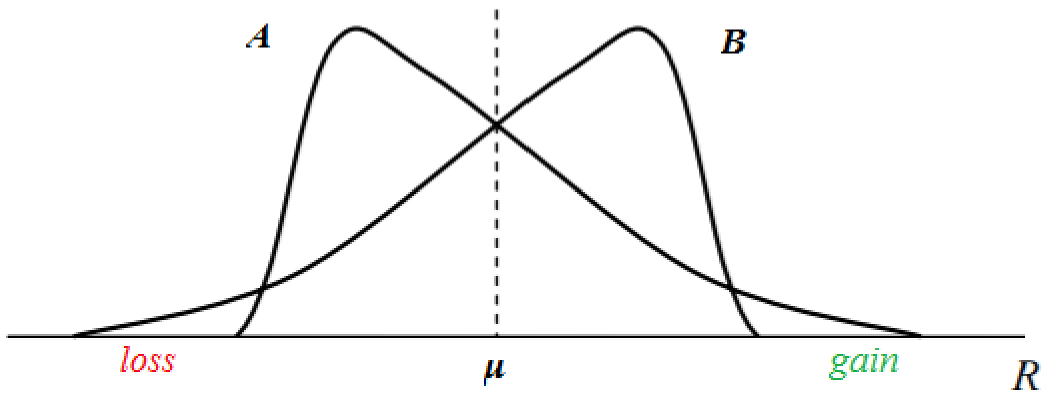

FIGURE 4.5: Return Distributions with identical Mean and Variance

Chapter 5

The Regression Approach to Mean-Variance Analysis

Almost 50 years after Markowitz's fundamentals in Mean-Variance Portfolios, Mark Britten-Jones, in his paper *The Sampling Error in Estimates of Mean-Variance Efficient Portfolio Weights* (1999), managed to express the connection between the finance issue of the Mean-Variance Analysis and the econometric process of the OLS regression. In this paper, Britten-Jones tests the hypotheses of Mean-Variance efficient portfolio weights using 20 years of data on 11 country stock indexes. By applying this approach, Britten-Jones resulted to large sampling errors in estimates of the weights of a global efficient portfolio.

5.1 OLS Regression to Mean-Variance Analysis

Britten-Jones (1999), assumes that a riskless asset is available for both borrowing and lending in each period, and calculates excess returns by subtracting the returns of this riskless asset from the return of each of the K assets in consideration. Let $x_t' = [x_{1t}, x_{2t}, ..., x_{Kt}]$ be the vector of excess returns in period $t = (1, 2, ..., T)$, and form the $T \times K$ matrix as follows:

$$X = \begin{bmatrix} x_1' \\ x_2' \\ . \\ . \\ . \\ x_T' \end{bmatrix} = \begin{bmatrix} x_{11} & x_{21} & .... & x_{K1} \\ x_{21} & x_{22} & ... & x_{K2} \\ . & . & . & . \\ . & . & . & . \\ . & . & . & . \\ x_{1T} & x_{2T} & ... & x_{KT} \end{bmatrix}.$$

Now, let $1 = \begin{bmatrix} 1 & 1 & ... & 1 \end{bmatrix}'$ be a T vector of ones. Viewed as a portfolio excess return, the T vector of ones 1 is highly desirable as it has positive excess return with zero standard deviation (risk). The regression approach to portfolio selection is based on minimizing the squared deviations between the excess returns on the constructed portfolio and the excess returns in 1. This minimization can be performed using an artificial ordinary least squares (OLS) regression.

Theorem 6 (OLS Regression to tangency portfolio). *Consider the artificial OLS regression*

$$1 = Xb + U. \tag{5.1}$$

results to the OLS coefficient vector given by

$$\hat{b} = (X'X)^{-1}X'1. \tag{5.2}$$

that is a set of risky-asset-only portfolio weights for a sample efficient portfolio. The scaled (so that weights sum to one) coefficient vector $\hat{b}/1'\hat{b}$ is thus the familiar tangency portfolio

$$\frac{\bar{\Sigma}^{-1}\bar{x}}{1'\bar{\Sigma}^{-1}\bar{x}} \tag{5.3}$$

derived from quadratic programming, where the sample mean $\bar{x} = X'1/T$, and the (maximum likelihood) sample covariance $(X - 1\bar{x}')'(X - 1\bar{x}')/T$, are used as parameters.

To make the theorem clear, let $\bar{x} = X'1/T$ be the K vector of sample means, and $\bar{\Sigma}(X - 1\bar{x}')'(X - 1\bar{x}')/T$ be the $K \times K$ sample covariance matrix. This later matrix may be written as

$$\begin{aligned}
\bar{\Sigma} &= (X - 1\bar{x}')'(X - 1\bar{x}')/T \qquad (5.4)\\
&= (X' - \bar{x}1')(X - 1\bar{x}')/T\\
&= (X'X - X'1\bar{x}' - \bar{x}1'X + \bar{x}1'1\bar{x}')/T\\
&= X'X/T - X'1\bar{x}'/T - \bar{x}1'X/T + \bar{x}1'1\bar{x}'/T\\
&= X'X/T + \bar{x}\bar{x}' - \bar{x}\bar{x}' + \bar{x}\bar{x}' = X'X/T - \bar{x}\bar{x}'.
\end{aligned}$$

In this way,

$$X'X = T(\bar{\Sigma} + \bar{x}\bar{x}').$$

Therefore,

$$\begin{aligned}
\hat{b} &= (\bar{\Sigma} + \bar{x}\bar{x}')^{-1}X'1/T = (\bar{\Sigma} + \bar{x}\bar{x}'^{-1}\bar{x}) \qquad (5.5)\\
&= \left(\bar{\Sigma}^{-1} - \frac{\bar{\Sigma}^{-1}\bar{x}\bar{x}'\bar{\Sigma}^{-1}}{1 + \bar{x}'\bar{\Sigma}^{-1}\bar{x}}\bar{x}\right)\\
&= \frac{\bar{\Sigma}^{-1}\bar{x}}{1 + \bar{x}'\bar{\Sigma}^{-1}\bar{x}}.
\end{aligned}$$

Scaling $\hat{b}$ so that the coefficients sum to one, we obtain:

$$\hat{w} \equiv \frac{\hat{b}}{1'\hat{b}} = \frac{\bar{\Sigma}^{-1}\bar{x}}{1'\bar{\Sigma}^{-1}\bar{x}}. \tag{5.6}$$

This portfolio is the **tangency portfolio**.

The regression in equation (5.1) is unusual for tree reasons:

1. there is no intercept
2. the dependent variable is non-stochastic, and

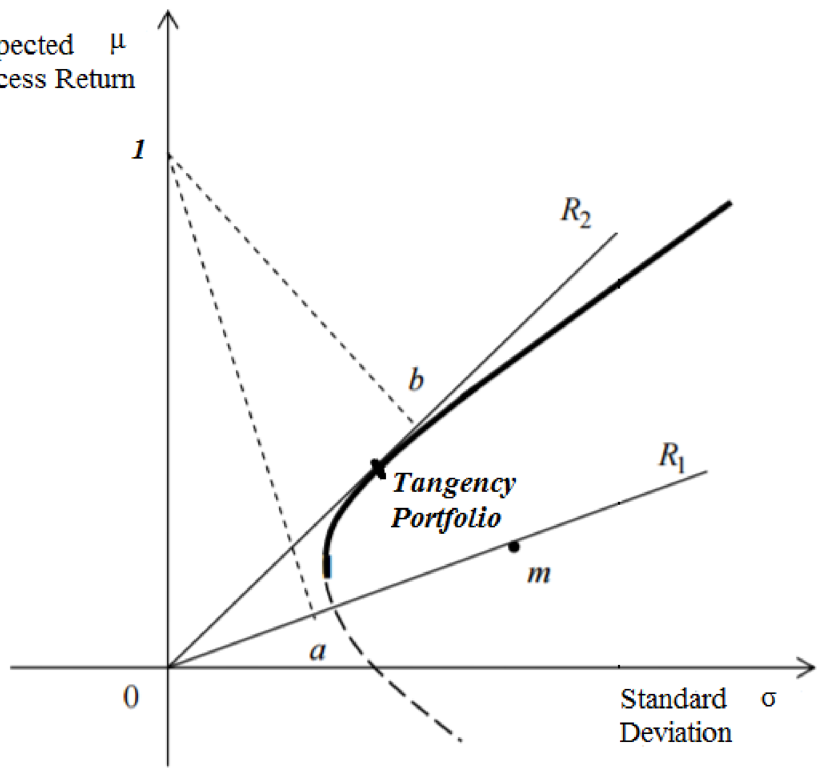

FIGURE 5.1: Tangency Portfolio

3. the residual vector u is correlated with the regressors, which are stochastic.

But it has a simple interpretation:

1. The dependent variable 1 is a sample counterpart to arbitrage profits (i.e., positive excess return with zero risk)

2. The coefficients b represent the weights on risky assets in the portfolio

3. Xb represents excess returns on this portfolio

4. The residual vector u shows deviations in this portfolio's returns from 1.

The least squares distance can be depicted using the mean-std.dev. diagram. OLS finds a portfolio whose returns are located in mean-std.deviation space as closely as possible to the point (0,1). In other words, the least squares problem in (5.1) finds the weight vector $\hat{b}$ that produces returns that mimic as closely as possible an arbitrage return. The arbitrage return 1 is located at the point (0,1). The feasible set, constructed from the sample mean and sample covariance matrix, has an efficient boundary shown by the line $0R_2$, from the origin passing through the tangency portfolio (Figure (5.1)).

5.1.1 Recovering the Entire Mean- Std.Deviation Frontier

The regression in (5.1) only computes the tangency portfolio. While this is certainly a very interesting point on the mean-std.deviation frontier, there is a way to modify the above methods to produce the entire frontier. To achieve this, a **constraint on the portfolio return** is imposed of the form

$$\bar{x}'w \equiv \bar{x}'\frac{b}{1'b} = \mu_P, \tag{5.7}$$

which can be written as

$$\bar{x}'b = \mu_P 1'b \tag{5.8}$$

or

$$(\bar{x} - \mu_P 1)'b = 0. \tag{5.9}$$

The **restricted regression** is

$$\begin{aligned} 1 &= Xb + u \\ s.t. \quad Rb &= 0. \end{aligned} \tag{5.10}$$

where, the restriction matrix (a vector in this case) is $R = (\bar{x} - \mu_P 1)'$. The regression in (5.10) computes a vector b that produces portfolio returns both as close as possible to arbitrage returns, and satisfying a given return requirement. By varying the required portfolio return μ_P, entire frontier can be recovered.

5.1.2 Testing the Efficiency of a Given Portfolio

Consider testing the null hypothesis:

$$H_0 : w = w_0 \tag{5.11}$$

against the alternative

$$H_1 : w \neq w_0 \tag{5.12}$$

In terms of the OLS coefficients b, the null may be written as:

$$H_0 : \frac{b}{1'b} = w_0 \qquad or \qquad H_0 : b = aw_0, \tag{5.13}$$

where α is an arbitrary constant, that can ex post be set equal to the sum of the OLS coefficients. Another way of writing the null hypothesis is to say that b is proportional to a given portfolio weights vector w_0 , i.e. that:

$$H_0 : b \propto w_0 \tag{5.14}$$

against the alternative

$$H_1 : b \propto w_0. \tag{5.15}$$

Under the null hypothesis, the model in (5.1) becomes

$$1 = \alpha(Xw_0) + u^*. \tag{5.16}$$

The null can now be tested by comparing the sum of squared residuals of the unrestricted model in (5.1) and the restricted model in (5.16). The test statistic is

$$F = \frac{(u^{*\prime}u^* - u'u)/(K-1)}{u'u/(T-K)} \sim F_{K-1,T-K}. \quad (5.17)$$

This statistic has a nice geometric interpretation. Noting that, in terms of Figure (5.1)

$$\bar{1a} = \sqrt{u^{*\prime}u^*} \qquad \textit{and} \qquad \bar{1b} = \sqrt{u'u} \quad (5.18)$$

so the test statistic can be rewritten as,

$$F = (\frac{\bar{1a}}{\bar{1b}} - 1)(\frac{T-K}{K-1}). \quad (5.19)$$

We see, therefore, that the statistic is a function of the ratio of the length of the lines $1a$ and $1b$, that measure the distance between the portfolios m and d, respectively, from the arbitrage point $(0,1)$. This ratio is, by definition, greater or equal to 1, and the larger it is, the less likely it becomes for a given portfolio m to be efficient.

5.2 Active Portfolio Management

How can we use this methodology to do active portfolio management? The following simple example demonstrates the process. Consider an investor who wants to form portfolios out of two assets, namely, a stock denoted by s and a bond denoted by b. To this end, he obtains T observations of the assets' excess returns given by,

$$X = \begin{bmatrix} x_1^s & x_1^b \\ x_2^s & x_2^b \\ . & . \\ . & . \\ x_T^s & x_T^b \end{bmatrix}. \quad (5.20)$$

The optimal (tangency) portfolio may be computed by the artificial regression

$$1 = Xb + u$$

which produces the **passive portfolio weights**,

$$\hat{w} \equiv \frac{\hat{b}}{1'\hat{b}} = \begin{bmatrix} \hat{w}^s \\ \hat{w}^b \end{bmatrix}. \quad (5.21)$$

Suppose now that there is one conditioning (or state) variable (such as the dividend yield or the spread between long and short Treasury yields) which affects the conditional distribution of returns. We observe a time series of this state variable

$$z = \begin{bmatrix} z_0 \\ z_1 \\ . \\ . \\ z_{T-1} \end{bmatrix} \tag{5.22}$$

where the dating reflects the fact that z is known at the beginning of each return period and can be used to predict the excess returns of the assets. The investor wishes to take the information in the conditioning variable into account, and compute optimal portfolio weights that depend on it. To do this, we construct the augment returns matrix

$$\tilde{X} = \begin{bmatrix} x_1^s & x_1^b & z_0 x_1^s & z_0 x_1^b \\ x_2^s & x_2^b & z_1 x_2^s & z_1 x_2^b \\ . & . & . & . \\ . & . & . & . \\ x_T^s & x_T^b & z_{T-1} x_T^s & z_{T-1} x_T^b \end{bmatrix}. \tag{5.23}$$

The expanded set of assets can be interpreted as managed portfolios, each of which invests in a single basis asset (i.e., s or b) an amount that is proportional to the value of the state variable z. Running the artificial regression

$$1 = \tilde{X}c + u \tag{5.24}$$

we compute portfolio weights

$$\hat{\omega} \equiv \frac{\hat{c}}{1'\hat{c}} = \begin{bmatrix} \hat{\omega}_1 \\ \hat{\omega}_2 \\ \hat{\omega}_3 \\ \hat{\omega}_4 \end{bmatrix} \tag{5.25}$$

for each of the two basis assets and the two managed portfolios. The **active portfolio weights** may now be computed by

$$\hat{w}_t \equiv \begin{bmatrix} \hat{w}_t^s \\ \hat{w}_t^b \end{bmatrix} = \begin{bmatrix} \hat{\omega}_1 + \hat{\omega}_3 z_{t-1} \\ \hat{\omega}_2 + \hat{\omega}_4 z_{t-1} \end{bmatrix}. \tag{5.26}$$

Comparing the active portfolio weights in (5.26) with the passive ones in (5.21) we see that while the passive weights are fixed for all t, the active weights change at every time period to track the changes in the conditioning variable z_{t-1}. If the conditioning variable z has no predictive ability, $\hat{\omega}_3$ and $\hat{\omega}_4$ will be close to zero (statistically insignificant), and the active weights reduce to the passive ones.

Chapter 6

Choquet Expected Utility and Conditional Value-at-Risk Portfolios

In 2008, the bankruptcy of Lehman Brothers, an event that provoked the beginning of a global financial crisis, made it clear that risk measurement is an issue of great importance. Nowadays, "*Financial institutions have to allocate so-called economic capital in order to guarantee solvency to their clients and counterparties*" (Tasche, 2002). A well known risk measure is **Value-at-Risk (VaR)**.

A significant fact in VaR's history can be traced back in 1994, when J.P.Morgan decided to make public its unique VaR System called *RiskMetrics*. *RiskMetrics* faced significant attention at that time and monopolized as a risk measure. It wasn't until Artzner *et al.* (1999) outlined certain shortcomings of VaR, when other risk measures started been examined. Artzner *et al.* were the first to introduce a more reliable risk measure, **Tail Conditional Expectation (TCE)**, later known as **Conditional Value-at-Risk (CVaR)**.

6.1 Value-at-Risk (VaR)

Value-at-Risk (VaR) is a widely used risk measure because its concept is easily understandable. VaR focuses on the down-side risk and is a function based on two features, time and confidence level.

Theorem 7 (**VaR**). *VaR can be described as the* $\alpha\%$ *certainty that the value of a portfolio will not decline by more than* $VaR_{N,\alpha}$ *euros in the next* N *days, where* N *the time horizon and* α *the confidence level.*

VaR can be used as a risk measure in different kinds of portfolio structures.

Definition 1 (**VaR of a Single Stock Asset**). *Let* S *be the value (price) of a financial asset and let* ΔS_1 *be the daily change of this value. The 1-day Value-at-Risk of the asset at the* $100\alpha\%$ *is*

$$P(\Delta S_1 \geq -VaR_{1,\alpha}) = \alpha. \tag{6.1}$$

$VaR_{N,\alpha}$ the negative of the $(1-a)th$ quantile of the distribution of ΔS_N. ΔS_N is a monetary amount measured in euros/dollars and it is expressed as a "loss"

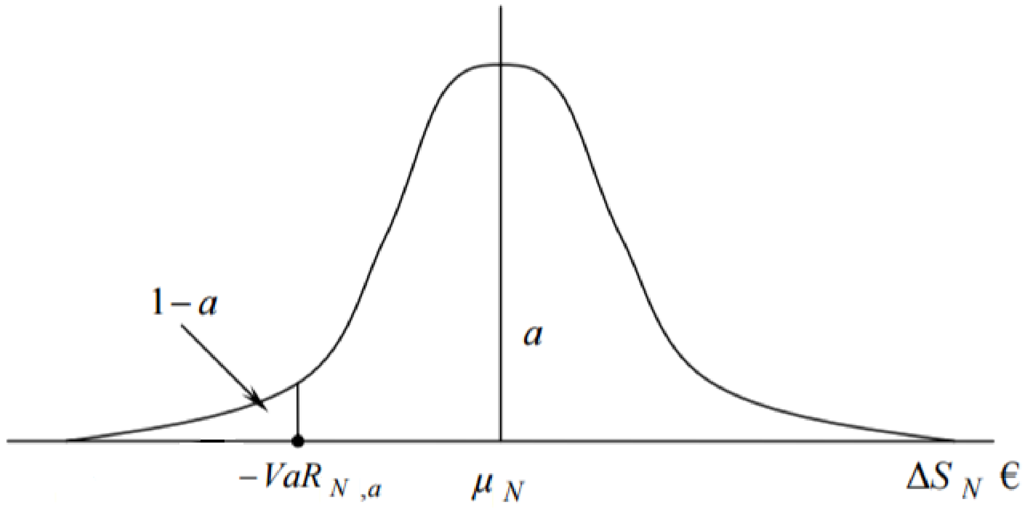

FIGURE 6.1: The distribution of N-day price change of a share and the N-day VaR at the 100α% confidence level

indicator so for the negative sign to be avoided. To compute $VaR_{N,\alpha}$ in applications the distribution of ΔS_N needs to be estimated. The most frequently assumption made is that ΔS_N is approximately normal with $N - day$ mean value change μ_N and N-day variance of value change σ_N^2

$$\Delta S_N \sim N(\mu_N, \sigma_N^2). \tag{6.2}$$

Considering that ΔS_N is approximately normal, the **N-day VaR** at the 100α% confidence level is given by

$$VaR_{N,\alpha} = z_\alpha \sigma_N - \mu_N, \tag{6.3}$$

where z_α is the number for which the probability that a standard normal will be less than z_α is α.

For an $M - day$ estimation:

$$\mu_M = \frac{M}{N}\mu_N \qquad and \qquad \sigma_M = \sqrt{\frac{M}{N}}\sigma_N. \tag{6.4}$$

So the **M-day VaR** equals

$$VaR_{M,\alpha} = z_\alpha \sqrt{\frac{M}{N}}\sigma_N - \frac{M}{N}\mu_N. \tag{6.5}$$

Even though Var is widely used, it is not a **coherent risk measure**. The background of coherent risk measures was firstly introduced by Artzner *et al.* in 1999.

6.1.1 Why not VaR?

When the portfolio return distribution is approximately normal, VaR is a suitable measure of risk. However, if returns are non-normal, problems can arise. The following story gives a first glimpse of these problems.

Barings Bank was one of the oldest investment banks in Britain. Despite its prestige, Barings, in order to survive in the late 20th century, employed young brokers who could work with financial assets of all classes. One of Barings employees was Nick Leeson. Leeson used to work at the Barings' back office and he soon discovered he was talented in the derivatives market. He was later transferred in Singapore, betting on market shifts around the world and became Barings' top employee, as his speculations accounted for 10% of the Bank's profits. That was the time when Leeson decided to create a secret Barings account. In this account, Leeson was adding profits created by small differences in the prices of Nikkei 225 futures occurring in two different markets, the Osaka Securities Exchange and the Singapore International Monetary Exchange. He started gambling on the future direction of the Japanese Market, risking huge amounts of money on the Nikkei, believing that the Japanese Stock Market would rise. Unfortunately for him, Nikkei's stock went the other way round, facing a huge decrease when an earthquake occurred in Kobe in 1995. Leeson's losses ascended that quick that soon shook down not only him but the hole bank. Losses came to more than $1 billion, an amount the bank could not cover. It collapsed in March of 1995 and was bought by the Dutch financial company ING for just one British pound.

But what did go wrong with Barings' internal system to miss out Leeson's malpractices? By that time, Barings Bank used to work with VaR. If the broker mechanically sets the investor's margin equal to the VaR n, σ of his position, the investor can trick him by moving some of the mass of his portfolio value-change distribution to the left and to the right, as seen in Figure (6.2). This new disconnected distribution has exactly the same VaR as the original one, but the investor's potential losses (as well as his potential gains) are obviously much larger. If the investor is a risk taker he might very well prefer distribution (B), especially since he will not have to finance his risk taking. Obviously, such behavior can wreak havoc on the entire financial system by allowing investors to take risky positions that they might not be able to finance if their bet does not materialize. This leads to the fact that it was quite easy for Leeson to mislead the internal system in his favor. This is the reason why Conditional Value-at-Risk is preferred to VaR, as it will be discussed in Section (6.4).

6.2 Choquet Expected Utility and the idea of Pessimism

The Choquet Expected Utility (Schmeidler 1982, 1989) model is a model of decision making under uncertainty generalizing the Expected Utility (EU) model. Under the comonotone independence axiom, an appealing and intuitive axiom requiring that the usual independence axiom holds only when hedging effects are absent, preferences turn out to be represented through a Choquet integral.

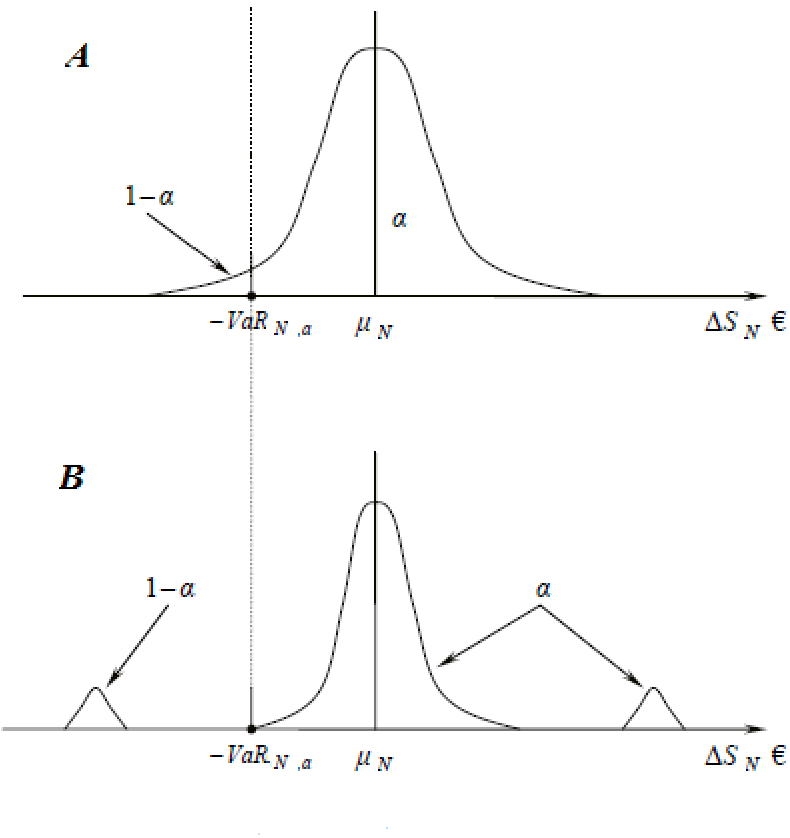

Figure 6.2: Tricking the broker

This model allows for taking into account a fuller array of behaviors under risk and under uncertainty. As a result, the model offers a simple theoretical foundation for explaining

- actual economic phenomena in finance, in insurance, phenomena that cannot be accounted for in the framework of EU theory,
- new indexes for inequality measurement(Chateauneuf & Cohen, 2015).

Consider the problem of choosing between two real-valued random variables, X and Y, with distribution functions, F and G, respectively. According to the Expected Utility Theory (Section (3.1)), X is preferred to Y if

$$E_F u(X) = \int_{-\infty}^{\infty} u(x)dF(x) > \int_{-\infty}^{\infty} u(x)dG(x) = E_G u(Y).$$

The classic Expected Utility expression can be formulated with the use of the Lebesque measure, which leads to:

$$E_F u(X) = \int_0^1 u(F^{-1}(t))dt > \int_0^1 u(G^{-1}(t))dt = E_G u(Y), \tag{6.6}$$

where F^{-1} and G^{-1} are the quantile functions corresponding to F and G.

Definition 2 (**Comonotonicity of two functions**). *The two function $X, Y \in \Omega$ are comonotonic if there exists a third function $Z \in \Omega$ and increasing functions f and g such that $X = f(Z)$ and $Y = g(Z)$.*

Bassett *et al.* (2004) suggested the behavior of the sums of the quantile functions as the property of comonotonic random variables with the greatest importance. For comonotonic random variables X, Y, it holds

$$F^{-1}_{X+Y}(u) = F^{-1}_X(u) + F^{-1}_Y(u).$$

By comonotonicity, $U \sim U[0,1]$ such that $Z = F^{-1}_X(u) + F^{-1}_Y(u)$ where g is left continuous and increasing, so by monotone invariance, $F^{-1}_g(U) = gF^{-1}_U = F^{-1}_X + F^{-1}_Y$.
Bassett *et al.* (2004) proposed one of the most interesting assumptions to extend the classical Expected Utility Theory described in Section (2.3). The variables have undergone some changes so an equivalent formulation occurs, where X is preferred to Y if

$$E_{v,F} u(X) = \int_0^1 u(F^{-1}(t))dv(t) > \int_0^1 u(G^{-1}(t))dv(t) = E_{v,G} u(Y). \tag{6.7}$$

Now the preference functions are (u, v), where u is the utility and v the transformed probability assessment. The **independence axiom** (Section (2.3)) has been transformed to be wider in order to emerge a larger class of preferences representable as Choquet capacities which introduce the idea of **pessimism**. This kind of independence axiom was proposed by Schmeidler (1989) who named it **comonotonic independence**.

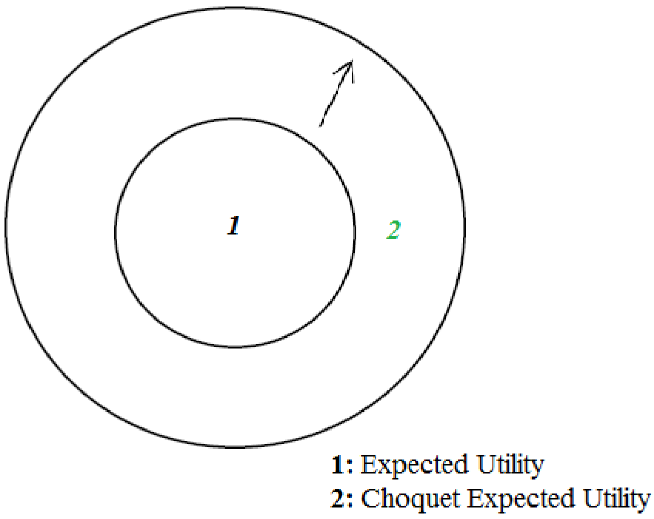

FIGURE 6.3: Choquet Expected Utility: A generalization of Expected Utility

Definition 3 (Comonotonicity of two acts). *Two acts x and $y \in \Omega$ are comonotonic, or similarly ordered if for no s and t in S*

$$x[t] > x[s] \qquad and \qquad y[s] > y[t]$$

which is translated as the assumption that if x is better in state t than state s, then y is also better in t than s.

Definition 4 (Comonotonic Independence). *For all pairwise comonotonic $x, y, z \in \Omega$ and $\alpha \in (0,1)$ $x > y \Rightarrow \alpha x + (1-\alpha)z > \alpha y + (1-\alpha)z$.*

The main characteristic of Choquet Expected Utility is the fact that, unlike the classical Expected Utility Theory, it can overweight a certain category of events, either the optimistic or the pessimistic ones. When the probability distortion function $v_a(t)$ is concave, Choquet Expected Utility allows for pessimism, whereas when $v_a(t)$ is convex it reflects optimism. The simplest form of Choquet Expected Utility introducing pessimism is based on distortion v

$$v_\alpha(t) = \min(t/\alpha, 1) \tag{6.8}$$

which leads to

$$E_{v_\alpha, Fu(X)} = \alpha^{-1} \int_0^\alpha u(F^{-1}(t))dt. \tag{6.9}$$

A **Probability distortion** is a divergence in the way an individual values a probability uniformly between 0 and 1. The distortion equation (6.8) *"acts to inflate or deflate the probabilities according to the rank ordering of the outcomes. The distortion may systematically increase the implicit likelihood of the least favorable events"*(Bassett *et al.*, 2004). Equation (6.9) exaggerates the probability of the proportion α of least favorable events, and ignores the probability of the $1 - \alpha$ most favorable events.

6.2.1 Is Choquet distortion irrational?

Daniel Kahneman and Amos Tversky introduced the **Prospect Theory** (1979) and the **Cumulative Prospect Theory** (1992), to explain violations of the Expected Utility. Choquet distortion, as described in Section (6.2), cannot be identified through Prospect Theory of Kahneman and Tversky (1979). Prospect theory indicates that losses and gains are perceived differently. This leads to biased choices, as an individual tends to prefer perceived gains to perceived losses. Also known as *Loss-Aversion Theory*, the general idea is that in cases of two equal options an individual would choose the option presented in terms of possible gains more likely to the option expressed in terms of possible losses. Given that choices are independent, the probability of gain or loss is assumed as being 0.5 respectively, even though the real probabilities are not necessarily equal. Essentially, the probability of gain is generally perceived as greater and, consequently, it is preferred by the individual. Prospect Theory calculates the Expected Utility of an uncertain prospect, which is equal to the sum of the utilities of the outcomes, each weighted by its probability:

$$V = \sum_{i=1}^{n} \pi(p_i) v(x_i), \tag{6.10}$$

where V is the Expected Utility of the outcomes, $x_1, x_2, \ldots, x_n$ are the potential outcomes and $p_1, p_2, \ldots, p_n$ the probabilities of these outcomes. **"v"** is a function that assigns a value to an outcome.

Graphically, the **value function** presents deviations from a certain reference point and is normally concave for gains (risk aversion), convex for losses (risk seeking) and steeper for losses than for gains (loss aversion), as seen in Figure (6.4).

In 1992, Kahneman and Tversky developed an updated version of Prospect Theory, the **Cumulative Prospect Theory** (1992). Cumulative Prospect Theory incorporates rank-dependent functionals which transform cumulative instead of individual probabilities, satisfying the **stochastic dominance** property that the Prospect Theory ignored, as well as being able to extend to prospects with a large number of outcomes. This leads to the aforementioned overweighting of extreme events which occur with small probability, rather than to an overweighting of all small probability events. According to Cumulative Prospect Theory:

1. Gains and losses, i.e. income, are the carriers of value, not final assets or wealth.

2. The value of each outcome is multiplied by a decision weight, not an additive probability.

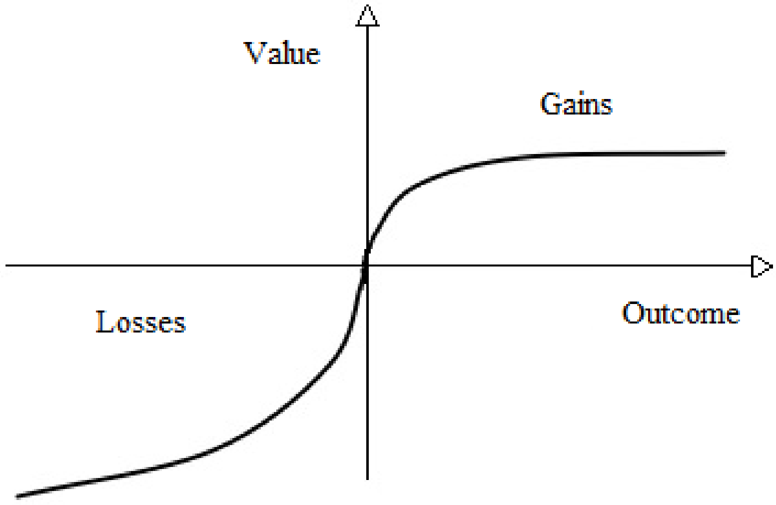

FIGURE 6.4: Prospect Theory: The value function is s-shaped, asymmetrical and steeper for losses than gains indicating that losses outweigh gains. The same Figure is used in Cumulative Prospect Theory.

The rank-dependent or **cumulative functional** that Kahneman and Tversky (1992) used, was first proposed by Quiggin (1982) for decision under risk and by Schmeidler (1989) for decision under uncertainty. This renewed version of Prospect Theory can clearly support the idea of the Choquet distortion, since it considers similar distortions of probabilities in the context of decision under ambiguity.

6.3 Coherent Measures of Risk

Only when the article of Artzner *et al.* was published in 1999, providing definitions and stating axioms on measures of risk, the inability of Value-at-Risk (VaR) as a risk measure became clear. VaR's failure consists in not satisfying one risk measures axiom, the **subadditivity**. According to Artzner *et al.* (1999):

Theorem 8 (Coherent Risk Measures). *For real-valued random variables* $X \in X$ *on* (Ω, A)*, a mapping* $\rho : XR$ *is called a coherent risk measure if it is*

1. ***Monotone:*** $X, Y \in X$, with $X \leq Y \Rightarrow \rho(X) \geq \rho(Y)$.
2. ***Subadditive:*** $X, Y, X + Y \in X, \Rightarrow \rho(X + Y) \leq \rho(X) + \rho(Y)$.
3. ***Linearly homogeneous:*** For all $\lambda \geq 0$ and $X \in X, \rho(\lambda X) = \lambda\rho(X)$.
4. ***Translation invariant:*** For all $\lambda \in R$ and $X \in X, \rho(\lambda + X) = \rho(X) - \lambda$.

6.3.1 Understanding Coherency

Bassett *et al* (2016) explain coherent risk measures through an interesting story of Samuelson (1963). Samuelson describes asking a colleague at lunch whether he would be willing to make a 50 - 50 bet, with the prices of 200 dollars in the case of winning and 100 dollars in case of losing. His colleague (later revealed to be E. Cary Brown) responded *"No, but I would be willing to make 100 such bets"*. Brown's response was considered irrelevant to the problem given, as it seemed that, in his response, there was a confusion about expected utility maximization combined with a fundamental misunderstanding of the law of large numbers. Of course Brown was way clever than that. So we see that,

$$d\phi(t) = \frac{1}{2}\delta_{1/2}(t) + \frac{1}{2}\delta_1(t).$$

For one coin flip it holds:

$$E_{v,F}(X) = \frac{1}{2}(-100) + \frac{1}{2}(50) = -25.$$

But, if the coin gets flipped a hundred times:

$$B = \sum_{i=1}^{100} X_i \sim Bin(0.5, 100)$$

then the outcome changes over Brown

$$\begin{aligned} E_{v,F}(B) &= \frac{1}{2}2\int_0^{1/2} F_B^{-1}(t)dt + \frac{1}{2}(5000) \\ &= 1704.11 + 2500 \\ &= 4204.11. \end{aligned}$$

A solution that indicates that the chance of loosing on the 100 coin tosses is about 1/2300.

6.4 Conditional Value-at-Risk (CVaR)

It can be proved that VaR and σ are not coherent risk measures. VaR does not satisfy the axiom of **subadditivity** and σ the **monotonicity** axiom. Although, there is a risk measure that includes all the above violated axioms by VaR and σ. This measure of risk is the **Conditional Value at Risk (CVaR)**, also known as **Tail Conditional Expectation (TCE)** and **Expected Shortfall (ES)**. The advantage of CVaR over VaR is that it can not be fooled by moving some of the mass of the $\Delta\Sigma_N$ distribution to minus infinity. Conditional Value-at-risk is the risk measure used to reduce the probability that a portfolio will face large losses. The CVaR's operating procedure is to assess the possibility at a specific confidence level that a specific loss will exceed the Value-at-Risk.

CVaR derives by taking the weighted average between the Value-at-Risk and losses exceeding the Value-at-Risk.

Rockafellar and Uryasev, introduced the term Conditional Value-at-Risk in their paper *Optimization of Conditional Value-at-Risk*, in 2000. This risk measure can also be found in other papers with a different name. Acerbi and Tasche (2001) refer to *Expected Shortfall* and , before them, Artzner *et al* (1999) refers to *Tail Conditional Expectation.*

Theorem 9 (**Conditional Value-at-Risk (CVaR)**). *Let X be a continuous random variable representing loss. Given a parameter $0 < \alpha < 1$, the α-CVaR of $X is$*

$$CVaR\alpha(X) = E[X|X \geq VaR\alpha(X)]. \tag{6.11}$$

Definition 5 (**Expected Shortfall**). *Let X be the profit-loss of a portfolio on a specified time horizon T and let $A = A\% \in (0,1)$ some specified probability level. The expected $A\%$ shortfall of the portfolio is then defined as*

$$ES^{\alpha}(X) = -\frac{1}{2}(E[X1_{X,x^{(\alpha)}}] - x^{\alpha}(P[X,x^{\alpha}] - \alpha)). \tag{6.12}$$

Definition (5) was first introduced by Acerbi *et al.* (2001) and it was proved to be coherent by Pflug (2000).

According to Bassett *et al.*, the leading example of a coherent risk measure is

$$\rho_{v\alpha}(X) = -\alpha^{-1}\int_0^{\alpha} F^{-1}(t)dt \tag{6.13}$$

where

$$\rho_{v\alpha}(X) = -E_{v\alpha,F(X)} = -\int_0^1 F^{-1}(t)dv(t)$$

which means that CVaR is the negative Choquet expected Utility with a distortion v_a.

Figure (6.6) reveals the difference between VaR and CVaR performance. We can observe that VaR weights follow a non-regular form, not being globally convex, while CVaR weights are convex which, consequently, leads to safer results.

The R code for Figure (6.6) is given by:

```
#set.seed(1)
n <- 10000

x1 <- rnorm(n)
x1 <- 0.05+0.03*((x1-mean(x1))/sqrt(var(x1)))
x2 <- rchisq(n,4)
x2 <- 0.10+0.07*((x2-mean(x2))/sqrt(var(x2)))

w <- seq(0.2,0.6,0.001)
VaR <- w
```

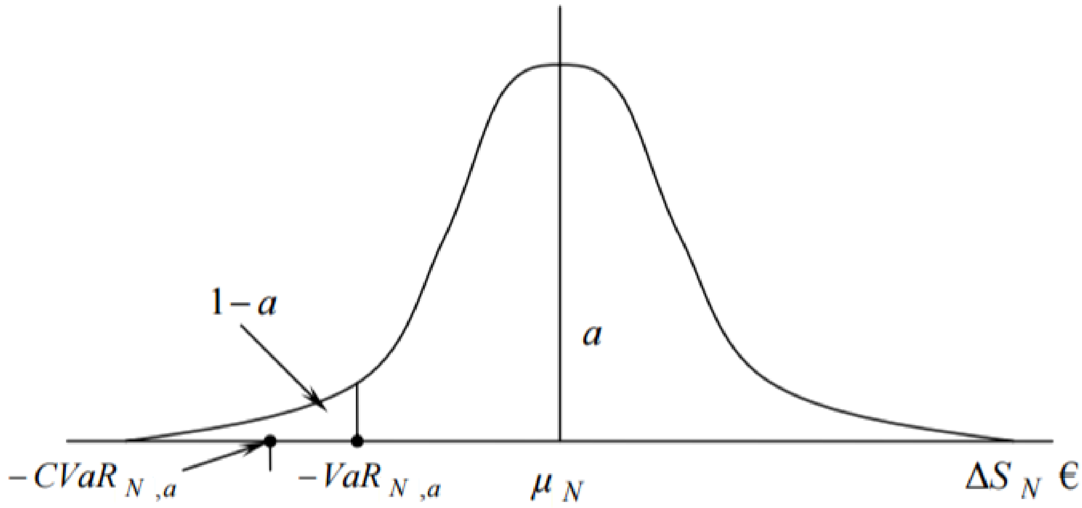

FIGURE 6.5: The distribution of N-price change of share and the N-day VaR and CVaR at the $100\alpha\%$ confidence level

```
CVaR <-w

for(i in 1:length(w)){
r <- w[i]*x1+(1-w[i])*x2
VaR[i] <- -quantile(r,0.1)
      CVaR[i] <- -mean(r[r<quantile(r,0.1)])
}

plot(w,VaR,type="l",col=2,ylab=c("VaR, CVaR"))
lines(w,(min(VaR)/min(CVaR))*CVaR,lty=5,col=3)
legend(0.35,-0.032,c("VaR","CVaR"),lty=c(1,5),col=c(2,3))
```

6.5 Pessimistic Risk Measures

Definition 6. *A risk measure ρ will be called pessimistic if, for some probability measure ϕ on $[0,1]$*

$$\rho(X) = \int_0^1 \rho_{v_\alpha}(X)d\phi(\alpha) \tag{6.14}$$

By **Fubini**,

$$\begin{aligned}\rho(X) &= -\int_0^1 \alpha^{-1}\int_0^\alpha F^{-1}(t)dtd\phi(\alpha)\\ &= -\int_0^1 F^{-1}(t)\int_t^1 \alpha^{-1}d\phi(\alpha)dt\end{aligned}$$

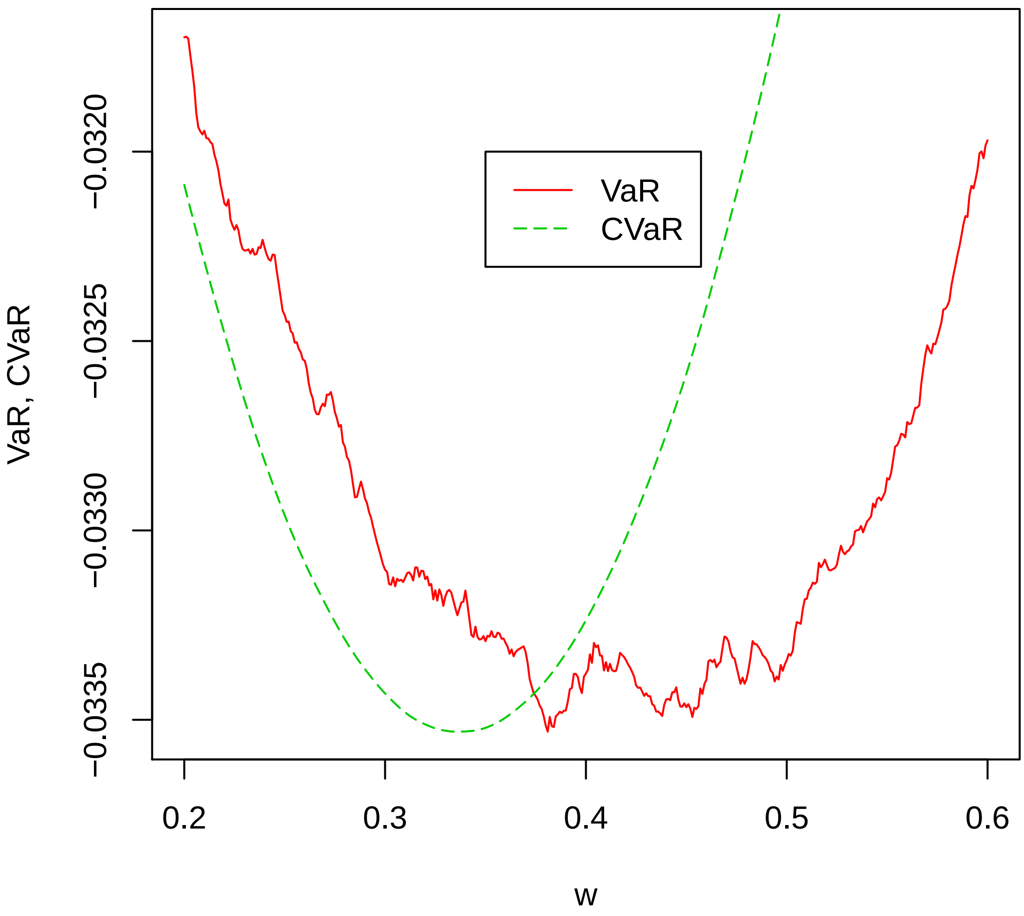

FIGURE 6.6: VaR vs CVaR performance

$$= -\int_0^1 F^{-1}(t)dv(t).$$

The Fubini theorem is used in order to make it clear why a risk measure like CVaR should be considered pessimistic.

Any pessimistic risk measure can be approximated by taking

$$d\phi(t) = \sum \phi_i \delta_{\tau_i}(t) \tag{6.15}$$

where δ_τ denotes (Dirac) point mass 1 at τ.
Then,

$$\rho(X) = -\phi_0 F^{-1}(0) - \int_0^1 F^{-1}(t)\gamma(t)dt \tag{6.16}$$

where $\gamma(t) = \sum \phi_i T_i^{-1} I(t < \tau_i)$ and $\phi_i > 0$ with $\sum \phi_i = 1$. The positivity of the point masses ϕ_i inform us that density weights are decreasing, which leads to accentuation of the likelihood of the least-favorable outcomes and reveals the pessimistic character of these risk measures.

According to Kusuoka (2001):

Theorem 10 (**Kusuoka's Theorem of Coherent Risk Measures**). *A regular risk measure is coherent in the sense of Artzner et al. if and only if it is pessimistic.*

1. *Pessimistic Choquet risk measures correspond to concave v , i.e.,* ***monotone decreasing dv****.*

2. *Probability assessments are distorted to accentuate* ***the probability of the least favorable events****.*

3. *The crucial coherence requirement is* ***subadditivity****, or submodularity, or 2-alternatingness in the terminology of Choquet capacities.*

Kusuoka gave the context of coherent risk measures through a Theorem containing the tree basic features that describe Pessimistic risk measures. It can be said, that Kusuoka's Theorem can conclude this Chapter smoothly, as it briefly contains all the mathematical concepts explained. Pessimistic risk measures have a concave v, a fact that indicates risk aversion. A distortion in their probability assessment has been undertaken resulting to an increase in the implicit likelihood of the least favorable events. Finally, coherent risk measures, in the sense of Artzner *et al.* (1999), can be considered only when the 4 Axioms take place (Section (6.3)). VaR's characteristics do not include this last prerequisite, which is, though, satisfied by CVaR. So, CVaR is not only a coherent risk measure, but it can also be considered as a pessimistic one. CVaR, besides being potentially pessimistic, presents another distinctive feature; its minimization leads to **Quantile regression**.

Chapter 7

The Regression Approach to CVaR Analysis

Quantile Regression was a breakthrough in regression analysis, proposed by Koenker and Bassett (1978). This approach permits estimating various quantile functions of a conditional distribution, like the median, the 0.25 quantile, or the 0.90 quantile. Quantile regression is really useful when the distribution is asymmetric or fat-tailed. Due to its usefulness, Quantile regression maintains its position in the latest literature as well, as seen in works of Koenker (2000), Koenker & Hallock (2001), Koenker (2005).

7.1 Quantile Regression to CVaR Analysis

In 1978, Koenker *et al.* introduced a new econometric method, the **Quantile Regression**. In the risk case, the minimization of CVaR leads to Quantile regression.

Theorem 11. *Let X be a real-valued random variable with $EX = \mu < \infty$ and $\rho_\alpha(U) = U(\alpha - I(U < 0))$. Then,*

$$\min_{\xi \in R} E_{\rho\alpha}(X - \xi) = \alpha\mu + \rho_{v\alpha}(X).$$

So a risk can be estimated by the equivalent form of:

$$\hat{\rho}_{v\alpha}(X) = (na)^{-1} \min_{\xi \in R} \sum_{i=1}^{n} \rho_\alpha(x_i - \xi) - \hat{\mu}_n \tag{7.1}$$

where $\hat{\mu}_n$ denotes an estimator of $EX = \mu$.

Proof.

$$E_{\rho\alpha}(X - \xi) = \alpha(\mu - \xi) - \int_{-\infty}^{\xi} (x - \xi) dF_X(x), \quad \xi_\alpha = F_x^{-1}(\alpha)$$

$$= \alpha\mu - \int_0^\alpha F^{-1}(t)dt$$

$$= \alpha\mu + \rho_{v\alpha}(X).$$

□

Imposing the extra condition of *regularity* for risk measure (Kusuoka (2001), Theorem (10)), Bassett *et. al.* (2004) show that a regular risk measure is *coherent if and only if* it is *pessimistic*, i.e., if the distortion function ν_α is concave in t. The concavity of ν_α corresponds to overweighting the probabilities of large losses and underweighting the probability of large gains, which is pessimistic. This is also related to the *cumulative prospect theory* of Kahneman and Tversky (1992) that consider similar distortions of probabilities in the context of decision under *ambiguity* (Subsection (6.2.1)).

In econometrics, empirical strategies for minimizing α-risk are related to *Quantile regression* as introduced by Koenker and Bassett (1978). Bassett *et. al.* (2004), Theorem 2, show that minimizing α-risk (CVaR risk) is equivalent to estimating an α-quantile regression of the type described below.

Let $x_j, j = 1, ..., p$ be the returns of m financial assets and consider the artificial regression

$$1 = \sum_{j=1}^{p} x_j \beta_j + u \qquad s.t. \quad \sum_{j=1}^{p} \beta_j = 1. \tag{7.2}$$

This regression was proposed by Britten-Jones (1999), who showed that the OLS estimates, normalized to sum to 1, correspond to Markowitz Mean-Variance (MV) optimal portfolio weights (Chapter (5)). The intuition behind this regression is that since 1 is a positive constant it can be thought off as an *arbitrage profit*, while since $\sum_{j=1}^{p} x_j \beta_j$ is the return of a portfolio with weights $\beta_1, ..., \beta_m$, minimizing

$$E\left(1 - \sum_{j=1}^{p} x_j \beta_j\right)^2 \quad s.t. \quad \sum_{j=1}^{p} \beta_j = 1. \tag{7.3}$$

produces a portfolio that resembles an arbitrage profit as nearly as possible. As it turns out, this is the classical MV optimal portfolio.

Bassett *et. al.* (2004) replace the **square loss** function of least-squares estimation with the **check loss** function $\rho_\alpha(u) = u(\alpha - I(u < 0))$ of quantile regression estimation, and estimate α-risk optimal portfolios that correspondent to the Choquet/pessimistic framework discussed above. They consider portfolios that minimize

$$E\rho_\alpha\left(1 - \sum_{j=1}^{p} x_j \beta_j\right) \quad s.t. \quad \sum_{j=1}^{p} \beta_j = 1 \tag{7.4}$$

for $\alpha < 0.5$, i.e., overweighting bad outcomes and underweighting good ones.

The advantage of the optimization formulation becomes clear when used in portfolios. Let $Y = X^T \pi$ denote a portfolio of assets comprised of $X = (X_1, ..., X_p)^T$ with portfolio weights π. Suppose now that we observe a random sample $x_i = (x_{i1}, ..., x_{ip}) : i = 1, ..., n$ from the joint distribution of asset returns, and we wish to consider portfolios minimizing the Lagrangian:

$$\min_{\pi} \rho_{v\alpha}(Y) - \lambda\mu(Y). \tag{7.5}$$

The additional constraint that the portfolio weights π sum to one is introduced:

$$\min_{\pi} \rho_{v\alpha}(X_\pi^T) \tag{7.6}$$

$$\text{s.t. } \mu(X_\pi^T) = \mu_0 \quad \text{where} \quad 1^T\pi = 1.$$

Taking the first asset as **numeraire** we can write the sample analogue of this problem:

$$\min_{\beta,\xi \in R^P} \sum_{i=1}^{n} \rho_\alpha(x_{i1} - \sum_{j=2}^{p}(x_{i1} - x_{ij})\beta_j - \xi) \tag{7.7}$$

$$\text{s.t. } \bar{x}^T\pi(\beta) = \mu_0 \quad \text{where} \quad \pi(\beta) = (1 - \textstyle\sum_{j=2}^{p} \beta_j, \beta^T)^T.$$

This is a linear quantile regression as introduced by Koenker and Bassett (1978). Instead of solving for a scalar quantity representing the α sample quantile, we are solving for p coefficients of a linear function estimating the α conditional quantile function of the numeraire return given the other returns.

The problem in (7.7) can be formulated as a conventional quantile regression problem if a single pseudo observation to the sample consisting of response $\kappa(\bar{x_1} - \mu_0)$ and a row $\kappa(0, \bar{x_1} - \bar{x_2}, ..., \bar{x_1} - \bar{x_p})^T$ is added, in order to implement the mean return constraint. The zero element corresponds to the intercept column of the design matrix, and now, the problem structured is a completely standard linear quantile regression. When a sufficiently large κ is used, the results arising satisfy the constraint given. Changing μ_0 values, an empirical CVaR-risk frontier analogous to the Mean-Variance frontier occurs. The Experiment of Chapter (8) is based on these results, as seen in the paper of Bassett et al. (2004).

Chapter 8

MV Portfolio versus CVaR Portfolio

Monte Carlo simulation is a computational- mathematical approach based on the assumption that a randomly chosen sample tends to perform identically to the population from which it was picked, presenting the same characteristics and properties. Monte Carlo is also used to check theoretical or analytical results which are based on "asymptotics". This means that they hold when the sample size goes to infinity. In Finance, Monte Carlo simulation has a leading role in risk assessment. It provides the opportunity to evaluate different risk scenarios, even the most extreme ones. By its means, all possible risk scenarios are generated and different kind of risk measures are tested. This master thesis experiment consists of a Monte Carlo examination of the properties of simple single-quantile portfolios when returns follow skewed and fat-tailed distributions. This Monte Carlo approach is elaborated with the use of the programming language `R`. Computer code in `R` has been assessed in order to implement more elaborate portfolios that minimize a weighted average of CVaR risk. The experiment's procedure is described below.

8.1 An Experiment

The design of the experiment is as follows:

There are 4 assets with independent returns. Assets 1 and 2 have identical means (0.05) and standard deviation (0.02) and, similarly, assets 3 and 4 also have the same mean (0.09) and standard deviation (0.07). Although the pair of asset 1 and asset 2 are the same from the aspect of mean and variance, asset 1 is normally distributed, while asset 2 has a reversed χ^2_3 density, thus skewed to the left. As for the second pair of assets, asset 3 is normally distributed while asset 4 is χ^2_3 ans, thus skewed distributed to the right.

The R code generating returns from these assets is given by:

```
library(KernSmooth)
library(quantreg)
n <- 200000
x1 <- rnorm(n)
x1 <- 0.05+0.02*((x1-mean(x1))/sqrt(var(x1)))
x2 <- -rchisq(n,3)
x2 <- 0.05+0.02*((x2-mean(x2))/sqrt(var(x2)))
x3 <- rnorm(n)
x3 <- 0.09+0.07*((x3-mean(x3))/sqrt(var(x3)))
x4 <- rchisq(n,3)
x4 <- 0.09+0.07*((x4-mean(x4))/sqrt(var(x4)))
x <- cbind(x1,x2,x3,x4)
mu = apply(x, 2, mean)
V = var(x)}
```

According to a *Folk Theorem* in Finance, odd moments are desirable by investors while even moments are undesirable. This means that investors prefer assets with more mean and skewness and avoid assets with more variance and kurtosis. MV portfolio -as its name indicates- takes into account only the two first moments of the returns distributions, the mean and the variance, while CVaR portfolio accounts for moments, including skewness and kurtosis. For the first set of assets, we expect that CVaR will tend to prefer the normally distributed asset (asset 1) to its twin that is skewed to the left (asset 2) but, for the second pair of assets will prefer the asset skewed to the right (asset 4) rather than the normally distributed one (asset 3). MV tends to be symmetrical in its preferences between the assets of the two pairs, being unable to distinguish upside from downside risk, and thus its desire to avoid both. We could say that MV is like a person with color blindness, being able to distinguish only two colors from the whole color set, while CVaR can see the entire spectrum.

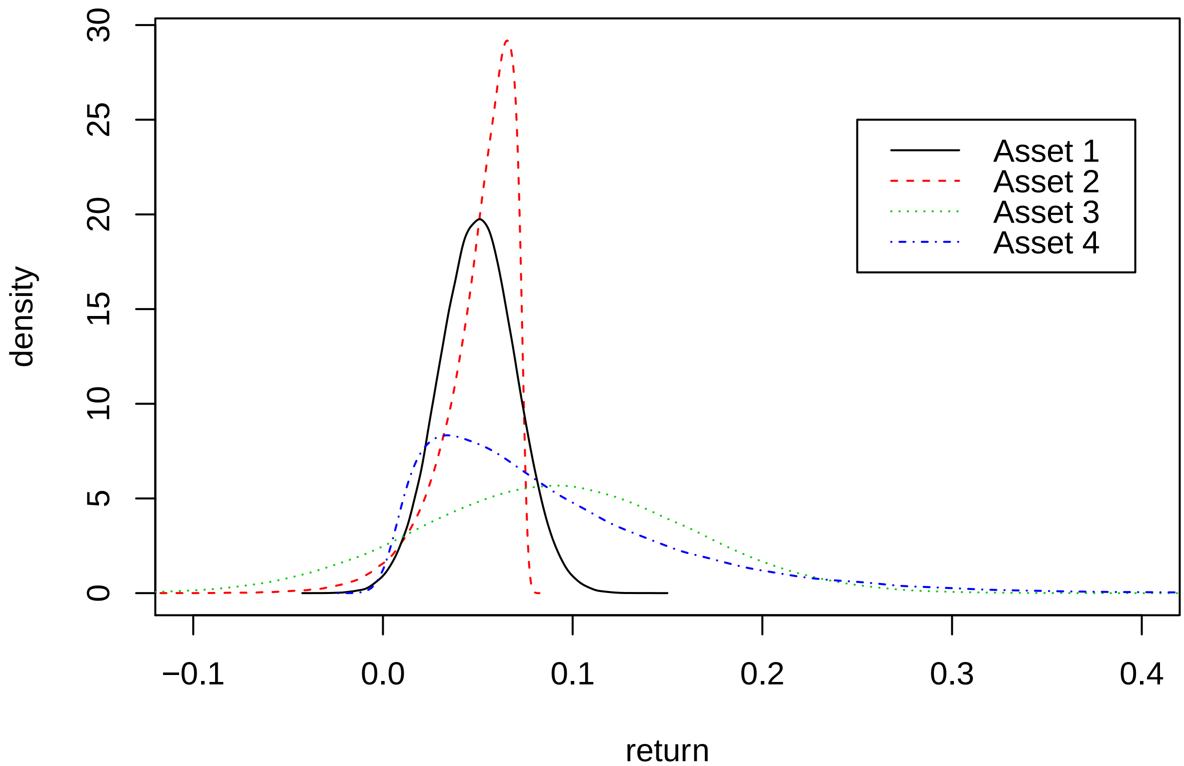

FIGURE 8.1: Four asset densities. The pair of assets 1 and 2, as well as the pair of assets 3 and 4, have the same mean and variance, even though their return distribution differs. Assets 1 and 4 are those with the better performance in both tails, compared to their identical assets, 2 and 3 respectively.

The return distributions are characterized as follows:

```
est1 <- bkde(x1)
est2 <- bkde(x2)
est3 <- bkde(x3)
est4 <- bkde(x4)

plot(est2,type="l",xlim=c(-0.1,0.4),lty=2,xlab="return",ylab="density",col=2)
lines(est1,col=1)
lines(est3,lty=3,col=3)
lines(est4,lty=4,col=4)
legend(0.25,25,c("Asset 1","Asset 2","Asset 3","Asset 4"),lty=c(1,2,3,4),
col=c(1,2,3,4))


ones <- rep(1,n)
coef <- lm(ones~x1+x2+x3+x4-1)$coef
wtan <- coef/sum(coef)
wtan
wtan%*%c(mean(x1),mean(x2),mean(x3),mean(x4))

x2e <- x1-x2
x3e <- x1-x3
x4e <- x1-x4

coef <- lm(x1~x2e+x3e+x4e)$coef
coef
wmin <- c(1-sum(coef[2:4]),coef[2:4])
round(wmin,5)
wmin%*%c(mean(x1),mean(x2),mean(x3),mean(x4))

Xe <- cbind(1,x2e,x3e,x4e)
#coef <- rq(x1~Xe-1, tau=0.1)$coef
#coef <- rq(x1~x2e+x3e+x4e, tau=0.1)$coef
coef
```

In order to perceive the differences between the MV and the CVaR portfolio's preferences, we generate a sample of $n = 100,000$ observations.
With a required portfolio return $\mu = 0.07$, we obtain

$$\pi(MV) = (0.279, 0.221, 0.252, 0.248). \tag{8.1}$$

For the CVaR, with $\mu = 0.07$ again and $\alpha = 0.1$, we obtain

$$\pi(CVaR) = (0.299, 0.201, 0.151, 0.349). \tag{8.2}$$

We now generate a sample of $n = 1,000,000$ observations from the returns distribution of these four assets, in order to compare portfolio choices. Let $X = (X_1, ..., X_p)$ a vector of potential asset returns and $Y = X^T\pi$ the returns on the portfolio with π as the weights. To minimize CVaR-risk subject to a constraint on mean return, consider $min_\pi \rho_{va}(Y) - \lambda\mu(Y)$. This leads to a linear quantile regression problem

$$min_{(\beta,\gamma)} \sum_{i=1}^{n}(x_{i1} - \sum_{j=2}^{\pi}(x_{i1} - x_{ij})\beta_j - \gamma)^2 \qquad s.t.\bar{x}^T\pi(\beta) = \mu_0 \tag{8.3}$$

where

$$\pi(\beta) = (1 - \sum_{j=2}^{\pi}\beta_j, \beta^T)^T.$$

Thus, the evaluation of the MV and the CVaR portfolio allocations resulted from the reproduction of the $\pi(MV)$ and the $\pi(CVaR)$ 100,000 times. Figure (8.2) reveals the outcome of this reproduction, giving the returns of the MV and CVaR portfolios. It is obvious that CVaR portfolio has a better performance than MV portfolio, as it out performs MV portfolio both in the left and right tail. CVaR returns, which are distributed with right skewness, reveal less possible losses in the lower tail but, also, more possible gains in the upper tail.

As seen in Table (8.1), with a required portfolio return of $\mu = 0.05$, MV distributes the portfolio between assets 1 and 2, given that their mean is the same as the required return (0.05), avoiding completely assets 2 and 3. CVaR also distributes the portfolio between assets 1 and 2 only, but attributes different weights for each of these two assets. Given $\mu = 0.05$ and $\alpha = 0.1$, we observe that CVaR attributes weights of 0.59 for asset 1, 0.38 for asset 2 and 0 for assets 3 and 4, preferring the normally distributed asset to its twin which is skewed to the left, as expected.

For $\mu = 0.09$, the mean of assets which satisfies this required return is the mean of the second pair, assets 3, 4. As we expected, both MV and CVaR now place the weights on assets 3 and 4. But, while MV attributes equal weights between pair of assets 3 and 4 (0.05), CVaR distinguishes the two assets again, preferring asset 4 (0.795) which is skewed to the right to asset 3 (0.205) which is normally distributed. For $0.06 \leq \mu \leq 0.08$, MV tends to have a symmetrical preference between the assets of each pair, while CVaR still prefers asset 1 over asset 2 and asset 4 over asset 3.
The R code generating Table (8.1):

```
kappa <- 1000000
x1a <- c(x1,kappa*(0.05-0.05))
Xea <- rbind(Xe,kappa*c(0,0,-0.04,-0.04))
#Mean-Variance
coef <- lm(x1a~Xea-1)$coef
coef
wMV <- c(1-sum(coef[2:4]),coef[2:4])
```

```
round(wMV,5)
wMV %*% mu
#CVaR
coef <- rq(x1a~Xea-1, tau=0.10)$coef
coef
wCV <- c(1-sum(coef[2:4]),coef[2:4])
round(wCV,5)
wCV %*% mu
```

Table 8.1: Portfolio weights- Skewness

$\mu = 0.05$	w_1	w_2	w_3	w_4
wMV	0.500	0.500	0.000	0.000
wCV, $\alpha = 0.10$	0.608	0.392	0.000	0.000
wCV, $\alpha = 0.15$	0.584	0.416	0.000	0.000
wCV, $\alpha = 0.20$	0.560	0.440	0.000	0.000
$\mu = 0.06$				
wMV	0.375	0.375	0.125	0.125
wCV, $\alpha = 0.10$	0.440	0.310	0.082	0.168
wCV, $\alpha = 0.15$	0.422	0.328	0.088	0.162
wCV, $\alpha = 0.20$	0.409	0.341	0.092	0.158
$\mu = 0.07$				
wMV	0.250	0.250	0.250	0.250
wCV, $\alpha = 0.10$	0.275	0.225	0.119	0.381
wCV, $\alpha = 0.15$	0.266	0.234	0.137	0.363
wCV, $\alpha = 0.20$	0.261	0.239	0.153	0.347
$\mu = 0.08$				
wMV	0.125	0.125	0.375	0.375
wCV, $\alpha = 0.10$	0.128	0.122	0.155	0.595
wCV, $\alpha = 0.15$	0.128	0.122	0.189	0.561
wCV, $\alpha = 0.20$	0.128	0.122	0.218	0.532
$\mu = 0.09$				
wMV	0.000	0.000	0.500	0.500
wCV, $\alpha = 0.10$	0.000	0.000	0.204	0.796
wCV, $\alpha = 0.15$	0.000	0.000	0.251	0.749
wCV, $\alpha = 0.20$	0.000	0.000	0.287	0.713

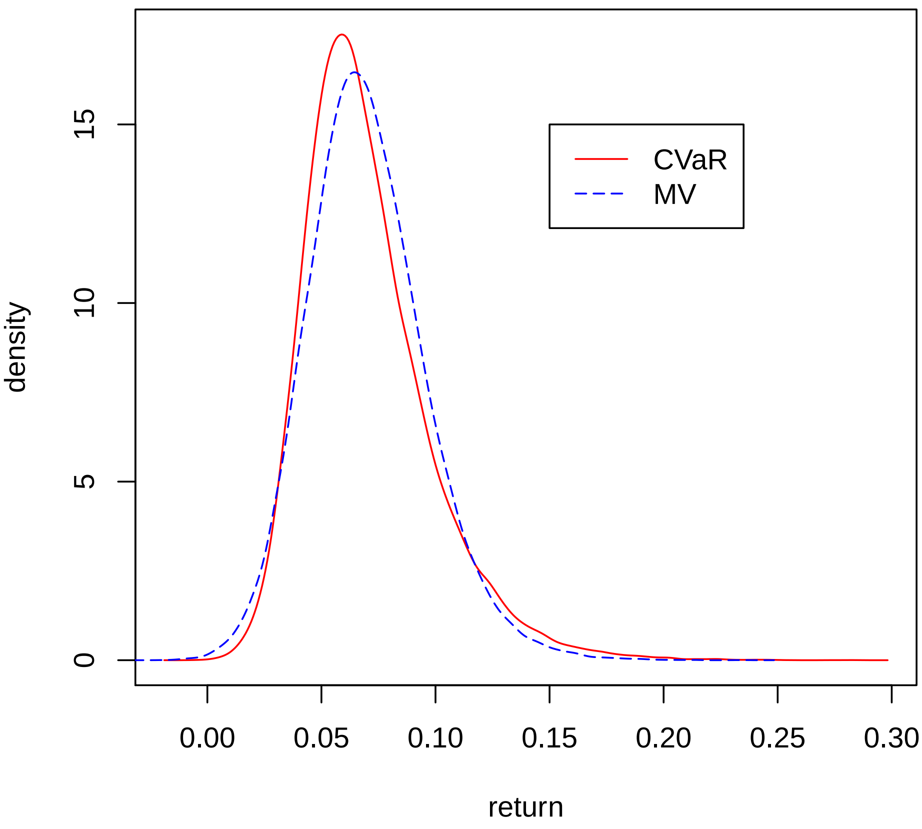

FIGURE 8.2: MV vs CVaR

In the case of kurtosis, the power of the *Folk Theorem* appears once again. Kurtosis, an even moment, is undesirable as shown in the CVaR results of Table (8.2). Fat tails are less appealing than thin tails. So, CVaR prefers assets 1 and 4, which appear having less kurtosis than their identical assets 2 and 3 respectively. MV, on the contrary, presents symmetrical preferences between the assets of the two pairs, as its results depend only on the mean and the variance. As in the former experiment of skewness, MV is unable to identify the kurtosis of the assets.

In Table (8.2), with a return equal to 0.05 MV distributes the portfolio between the assets 1 and 2, which have the same mean as the required return. CVaR also distributes the portfolio on these assets, given their mean, but also examines their kurtosis which results to larger weights on asset 1. CVaR, given $\mu = 0.05$ and $\alpha = 0.005$, attributes weights of 0.561 for asset 1 and 0.418 for asset 2, preferring the asset with the less kurtosis. For $\mu = 0.09$ both CVaR and MV distribute the portfolio between assets 3 and 4, which is now the pair of assets with the same mean as the required return. CVaR, though, prefers more asset 4 tending to avoid asset 3 which has fatter tails. Given $\mu = 0.09$ and $\alpha = 0.005$, CVaR attributes weights of 0.526 for asset 4 and 0.474 for asset 3, while MV attributes symmetrical weights to these two assets (0.50). Even for the intermediate return price $\mu = 0.07$, CVaR distinguishes the assets, setting larger weights on assets 1 (0.260) and 4 (0.263) over their twin assets 2 (0.240) and 3 (0.237) respectively.

The results of this experiment are remarkable. While MV takes into account only the mean and the variance of an asset distribution, CVaR investigates whether the asset in skewed or fat-tailed as well. In cases of non-normally distributed assets MV cannot allocate their differences based on their skewness or kurtosis. On the contrary, CVaR can examine these two moments as well and establish weights into the assets that contain the best characteristics, i.e. right skewness and thin tails.

TABLE 8.2: Portfolio weights- Kurtosis

$\mu = 0.05$	w_1	w_2	w_3	w_4
wMV	0.500	0.500	0.000	0.000
wCV, $\alpha = 0.005$	0.561	0.418	0.011	0.011
wCV, $\alpha = 0.010$	0.551	0.441	0.004	0.004
wCV, $\alpha = 0.050$	0.518	0.482	0.001	-0.001
$\mu = 0.06$				
wMV	0.375	0.375	0.125	0.125
wCV, $\alpha = 0.005$	0.420	0.330	0.121	0.129
wCV, $\alpha = 0.010$	0.409	0.341	0.121	0.129
wCV, $\alpha = 0.050$	0.387	0.363	0.123	0.127
$\mu = 0.07$				
wMV	0.250	0.250	0.250	0.250
wCV, $\alpha = 0.005$	0.260	0.240	0.237	0.263
wCV, $\alpha = 0.010$	0.263	0.237	0.241	0.259
wCV, $\alpha = 0.050$	0.252	0.248	0.247	0.253
$\mu = 0.08$				
wMV	0.125	0.125	0.375	0.375
wCV, $\alpha = 0.005$	0.121	0.129	0.357	0.393
wCV, $\alpha = 0.010$	0.125	0.125	0.358	0.392
wCV, $\alpha = 0.050$	0.122	0.128	0.370	0.380
$\mu = 0.09$				
wMV	0.000	0.000	0.500	0.500
wCV, $\alpha = 0.005$	0.005	-0.005	0.474	0.526
wCV, $\alpha = 0.010$	0.000	0.000	0.482	0.518
wCV, $\alpha = 0.050$	0.000	0.000	0.492	0.508

Conclusion

Financial Markets are composed of marketplaces where different asset classes, corresponding to different value levels, are negotiated. Due to this fact, risk measurement in the financial field constitutes a process of great importance. In the context of this thesis, two risk measures in Finance, Mean-Variance (MV) and Conditional Value-at-Risk (CVaR) have been examined comparatively and counterbalanced through a Monte Carlo experiment in R.

This thesis follows two *paths*, that both lead to a regression analysis. It all begun with Bernoulli who, in the 17th century, expanded the idea of the expected value and introduced the Utility function for Wealth. Later, von Neumann and Morgenstern (1947), using the Utility function for Wealth of Bernoulli, came up with the Expected Utility function. Savage (1954), based on the ideas of von Neumann-Morgenstern, as well as de Finetti (1937), proposed a complete set of axiomatics, the Axioms of Choice. Mean-Variance Analysis of Markowitz (1952) is connected to Expected Utility Theory but it is consistent with this Theory only under certain circumstances, since it demands returns to be normally distributed and/or U(.) to follow a quadratic form. 50 years after Markowitz's fundamentals in Mean-Variance Portfolios, Britten-Jones (1999) presented the connection between the Mean-Variance Analysis and the OLS regression. So, Mean-Variance Analysis is an alternative process of the OLS regression.

The second path begins with the Choquet capacity and the Comonotonic Preferences of Schmeidler (1989), so the Choquet Expected Utility is formulated. The idea of Choquet Expected Utility is to introduce a function v that distorts probabilities of individual events, and v of a certain form incorporates pessimism. This is equivalent to the loss function corresponding to coherent measures of risk, like Conditional Value-at-Risk (CVaR) of Rockafeller and Uryasev (2000), but unlike Value-at-Risk (VaR) which is not a coherent risk measure. Bassett *et al.* (2004), imposing the condition of regularity for risk measures, argued that a regular risk measure is coherent if and only if it is pessimistic. Based on Koenker *et al.* (1978), Bassett *et al.* (2004) showed that minimizing CVaR risk is equivalent to estimating a Quantile regression. So, CVaR/ Pessimistic Analysis is an alternative process of the Quantile regression, hence a more sufficient method compared to the MV Analysis.

Expected Utility may be criticized both as a positive, as well as a normative guide to economic behavior. Mean-Variance Portfolio Allocation is also unsatisfactory since it relies on unpalatable assumptions of Gaussian returns, or quadratic utility. On the contrary, Choquet Expected Utility provides a simple and tractable generalization of Expected Utility that allows for pessimism.

Pessimistic Portfolio Allocation can be formulated as a Quantile regression problem, thus providing an attractive practical alternative to the dominant Mean-Variance approach of Markowitz (1952).

These results are supported by the outcome of the Monte Carlo experiment in `R`. The Experiment illustrates two pairs of assets, the pair of assets 1 and 2, and the pair of assets 3 and 4. The assets of each pair have the same mean and variance, even though their return distributions differ. Assets 1 and 4 are those with a better performance in both tails, compared to their identical assets, 2 and 3 respectively. CVaR recognizes the assets with the best characteristics and attributes more weights to asset 1 and 4, both in the Skewness as well as in the Kurtosis experiment. In contrast, MV examines only the mean and the variance of the assets, thus it is not able to attribute different weights between asset 1 and 2, 3 and 4. While MV takes into account only the mean and the variance of an asset distribution, CVaR investigates whether the asset is skewed or fat-tailed as well. This means that CVaR works functionally in cases where MV is appropriate but MV is not functional in cases where CVaR is appropriate to use.

But, what does it hold in reality? Derivatives and portfolios containing derivative based strategies like bull-spread and bear-spread strategies, introduce into portfolios non-linearities that make the normal assumptions impalatable. Therefore, MV -and VaR- may be very misleading, while CVaR copes satisfactorily. This leads to the fact that Pessimistic Portfolio Allocation is not always compulsory, but its use is necessary in these cases.

In the paper of Bassett *et. al.* (2004), which has been elaborated and extended in the context of this thesis, more general distortions are considered that cannot be expressed as quantile/CVaR-risk problems but are still Pessimistic, Choquet-expectation optimal choices. One very interesting distortion function is given in the Example at the end of Section 2, where the investor picks portfolios according to the *minimum* of several (say 2 or 3) realizations of the underling asset distribution. This leads to potentially very pessimistic behavior, that can be examined in future research.

Bibliography

[1] Acerbi, C., and D. Tasche (2002). Expected Shortfall: A Natural coherent alternative to Value at Risk. *Economic Notes*, 31, 379–388.

[2] Arrow, K. (1965). Aspects of the Theory of Risk Bearing, *Yrjo Jahnsson Saatio*, Helsinki.

[3] Artzner, P., F. Delbaen, J.-M. Eber, and D. Heath (1999). Coherent Measures of Risk. *Math. Finance*, 9, 203–228.

[4] Bernoulli, D. (1738). Exposition of a New Theory of the Measurement of Risk. An English translation of the Latin original appears in *Econometrica*, 1954, pp.23-36.

[5] Bassett, G., R. Koenker, and G. Kordas (2004). Pessimistic Portfolio Allocation and Choquet Expected Utility. *Journal of Financial Econometrics*, 2, 477–492.

[6] Bassett, G., R. Koenker, and G. Kordas (2016, May). *Risk, Choquet Portfolios and Quantile Regression.* University of Copenhagen, 18–20 May 2016.

[7] Britten-Jones, M. (1999). The Sampling Error in Estimates of Mean-Variance Efficient Portfolio Weights. *Journal of Finance*, 54, 655–671.

[8] Chateauneuf, Al. and Cohen, M. (2015). *Choquet expected utility model: A new approach to individual behavior under uncertainty and to social welfare*, 8–15. Retrieved from: www.researchgate.net/publication/265548323

[9] Chambers, J. M. (1998). *Programming with Data: A Guide to the S Language.* New York: Springer-Vergal.

[10] Ellsberg, D. (1961). Risk, Ambiguity, and the Savage Axioms. *The Quarterly Journal Of Economics*, 75, 643–669.

[11] Everitt, B. S. (1998). *The Cambridge Dictionary Of Statistics.* Cambridge University Press, 2002.

[12] Fama, E. (1965). Random Walk in Stock Market Prices. *Financial Analysts Journal*, 21, 55–59.

[13] Fama, E., L. Fisher, M. Jensen and R. Roll (1969). The adjustment of Stock Prices to New Information. *International Economic Review*, 10.

[14] GWM Adviser Services Limited AFSL 230692 (2009). *Understanding investment concepts* (Version 4, 4–5). Sydney.

[15] Kahneman, D., and A. Tversky (1979). Prospect Theory: An Analysis of Decision Under Risk. *Econometrica*, 47, 263-291.

[16] Koenker, R., and G. Bassett (1978). Regression Quantiles. *Econometrica*, 46, 33–50.

[17] Kordas, G. (2010). *Portfolio Analysis*. Lecture Notes.

[18] Kusuoka, S. (2001). On Law Invariant Coherent Risk Measures. *Advances in Mathematical Economics*, 3, 83–95.

[19] Allais, M. (1953). Le Comportement de l'homme rationnel devant le risque: critique des postulats et axiomes de l'école américaine. *Econometrica: Journal of the Econometric Society*, 21, 503–546.

[20] Neumann, J. von, and O. Morgenstern (1944). *Theory of Games and Economic Behavior*. Princeton University Press, 1947.

[21] Pratt, W. (1964). Risk Aversion in the Small and in the Large. *Econometrica*, 32, 122–136.

[22] Rockafellar, R., and S. Uryasev (2000). Optimization of conditional VaR. *Journal of Risk*, 2, 21–41.

[23] Savage, L.J. (1954). *The Foundations of Statistics*. Dover, 1972.

[24] Schmeidler, D. (1989). Subjective Probability and Expected Utility Without Additivity. *Econometrica*, 57, 571-587.

[25] Tasche, D. (2002). Expected Shortfall and Beyond. *Journal of Banking and Finance*, 26, 1519–1533.

[26] Tversky, A., and D. Kahneman (1992). Advances in Prospect Theory: Cumulative Representation of Uncertainty. *Journal Risk and Uncertainty*, 5, 297–323.

Appendix A

Statistical background

This Appendix part consists of some basic statistic terms used in this Master thesis. It conducts a replication of the *THE CAMBRIDGE DICTIONARY OF STATISTICS, Second Edition* by B.S. Everitt (2002).

"**Covariance**
The expected value of the product of the deviations of two random variables, x, and y, from their respective means, μ_x and μ_y, i.e.,

$$cov(x, y) = E(x - \mu_x)(y - \mu_y)$$

The corresponding sample statistic is

$$c_{xy} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})$$

where (x_i, y_i), $i = 1, ..., n$ are the sample values on the two variables and $\bar{x}$ and $\bar{y}$ their respective means.

Expected Value
The mean of a random variable, X, generally denoted as $E(X)$. If the variable is discrete with probability distribution, $Pr(X = x)$, the $E(X) = \sum_x xPr(X = x)$. If the variable is continuous the summation is replaced by an integral. The expected value of a function of a random variable, $f(x)$, is defined similarly, i.e.

$$E(f(x)) = f(u)g(u)du$$

where g(x) is the probability distribution of x.

Hyperbolic distribution
Probability distribution, $f(x)$, for which the graph of $logf(x)$ is a hyperbola. [*Statistical Distribution in Scientific Work*, Volume 4, 1981, edited by C. Taille, G.P. Palit and B. Baldessari, Reidel, Dordrecht.]

Kurtosis
The extent to which the peak of a unimodal probability distribution or frequency distribution departs from the shape of a normal distribution, by either being more pointed (leptokurtic) or flatter (platykurtic). Usually measured for a

probability distribution as

$$\mu_4/\mu_2^2$$

where μ_4 is the fourth central moment of the distribution, and μ_2 is ts the variance. For a normal distribution this index takes the value three and often the index is redefined as the value above minus three so that the normal distribution would have a value zero.(Other distributions with zero kurtosis are called mesokurtic). For a distribution which is leptokurtic the index is positive and for a platykurtic curve it is negative.

Lagrange multipliers
A method of evaluating maxima or minima of a function of possibly several variables, subject to one or more . [*Optimization, Theory and Applications*, 1979, S.S. Rao, Wiley Eastern, New Dehli.]

Mean
A measure of location or central value for continuous variables. For a sample of observations $x_1, x_2, ..., x_n$ the measure is calculated as

$$\bar{x} = \frac{\sum^n i = 1x_i}{n}$$

Monte Carlo methods
Methods for finding solutions to mathematical and statistical problems by simulation. Used when the analytic solution of the problem is either intractable or time consuming. [*Simulation and the Monte Carlo Method*, 1981, R.Y. Rubenstein, Wiley, New York.]

Normal Distribution
A probability distribution, $f(x)$, of a random variable, X, that is assumed by many statistical methods. Specifically given by

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} exp[-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}$$

where μ and σ^2 are, respectively, the mean and the variance of x. This distribution is bell-shaped (...).

Outlier
An observation that appears to deviate markedly from the other members of the sample which it occurs. (...)the term refers to an observation which appears to be incosistent with the rest of the data, relative to an assumed model. Such extreme observations may be reflecting some abnormality in the measured characteristic of a subject, or they may result from an error in the measurement or recording. (...)

Probability distribution

For a discrete random variable, a mathematical formula that gives the probability of each value of the variable. (...) For a continuous random variable, a curve described by a mathematical formula ehich specifies, by the way of areas under the curve, the probability that the variable falls within a particular interval. (...)

Quantiles
The set of four variate values that divide a frequency distribution or a probability distribution into five equal parts.

R
A fast clone of S-PLUS which has the distinct advantage of being frely available.

Random variable
A variable, the values of which occur according to some specified probability distribution.

Regression modelling
A frequently applied statistical technique that serves as a basis for studying and characterizing a system of interest, by formulating a reasonable mathematical model of the relationship between a response variable, y and a set of q explanatory variables, $x_1, x_2, ..., x_q$. The choice of the explicit form of the model may be based on previous knowledge of the system or on considerations such as 'smoothness' and continuity of y as a function of the x variable. In very general terms all such models can be considered to be of the form

$$y = f(x_1, ..., x_q) + \epsilon$$

where the function f reflects the true but unknown relationship between y and the explanatory variables. The random additive error ϵ which is assumed to have mean zero and variance σ_ϵ^2 reflects the dependence of y on quantities other than $x_1, ..., x_q$. The goal is to formulate a function $\hat{f}(x_1, x_2, ..., x_p)$ that is a reasonable approximation of f. If the correct parametric form of f is known, then methods such as least squares estimation or maximum likelihood estimation can be used to estimate the set of the unknown coefficients. If f is linear in the parameters, for example, then the model is that of multiple regression. If the experimenter is unwilling to assume a particular parametric form of f then nonparametric regression modelling can be used, for example, kernel regression smoothing, recursive partitioning regression or multivariate adaptive regression splines.

Skewness
The lack of symmetry in a probability distribution. Usually quantified by the index, s, given by

$$\sigma = \frac{\mu_3}{\mu_2^{3/2}}$$

where μ_2 and μ_3 are the second and the third moments about the mean. the index takes the value zero for a symmetrical distribution. A distribution is said to have positive skewness when it has a long thin tail to the right, and to have negative skewness when it has a long thin tail to the left.

Standard deviation
The most commonly used measure of the spread of a set of observations. Equal to the square root of the variance.

Variance
In a population, the second moment about the mean. An unbiased estimator of the population value is provided by σ^2 given by

$$\sigma^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar(x))^2$$

where $x_1, x_2, ..., x_n$ are the n sample observations and $\bar{x}$ is the sample mean."

Appendix B

R Code

B.1 MV and CVaR portfolio weights for Skewness

The R code generating the results presented in Table (8.1):

```
library(KernSmooth)
library(quantreg)

#set.seed(1)
n <- 200000

x1 <- rnorm(n)
x1 <- 0.05+0.02*((x1-mean(x1))/sqrt(var(x1)))
x2 <- -rchisq(n,3)
x2 <- 0.05+0.02*((x2-mean(x2))/sqrt(var(x2)))
x3 <- rnorm(n)
x3 <- 0.09+0.07*((x3-mean(x3))/sqrt(var(x3)))
x4 <- rchisq(n,3)
x4 <- 0.09+0.07*((x4-mean(x4))/sqrt(var(x4)))

x <- cbind(x1,x2,x3,x4)
mu = apply(x, 2, mean)
V = var(x)
#_________________________________________________________________________________

#par(mfrow=c(2,2))

est1 <- bkde(x1)
est2 <- bkde(x2)
est3 <- bkde(x3)
est4 <- bkde(x4)

plot(est2,type="l",xlim=c(-0.1,0.4),lty=2,xlab="return",ylab="density",col=2)
lines(est1,col=1)
lines(est3,lty=3,col=3)
lines(est4,lty=4,col=4)
legend(0.25,25,c("Asset 1","Asset 2","Asset 3","Asset 4"),
```

```
lty=c(1,2,3,4),col=c(1,2,3,4))
#________________________________________________________________________________

ones <- rep(1,n)
coef <- lm(ones~x1+x2+x3+x4-1)$coef
wtan <- coef/sum(coef)
wtan
wtan%*%c(mean(x1),mean(x2),mean(x3),mean(x4))
#________________________________________________________________________________

x2e <- x1-x2
x3e <- x1-x3
x4e <- x1-x4

coef <- lm(x1~x2e+x3e+x4e)$coef
coef
wmin <- c(1-sum(coef[2:4]),coef[2:4])
round(wmin,5)
wmin%*%c(mean(x1),mean(x2),mean(x3),mean(x4))

Xe <- cbind(1,x2e,x3e,x4e)
#coef <- rq(x1~Xe-1, tau=0.1)$coef
#coef <- rq(x1~x2e+x3e+x4e, tau=0.1)$coef
coef
#________________________________________________________________________________

#Add a pseudo obs to impose the return constraint
kappa <- 1000000
x1a <- c(x1,kappa*(0.05-0.05))
Xea <- rbind(Xe,kappa*c(0,0,-0.04,-0.04))

#Mean-Variance
coef <- lm(x1a~Xea-1)$coef
coef
wMV <- c(1-sum(coef[2:4]),coef[2:4])
round(wMV,5)
wMV %*% mu

#CVaR
coef <- rq(x1a~Xea-1, tau=0.10)$coef
coef
wCV <- c(1-sum(coef[2:4]),coef[2:4])
round(wCV,5)
wCV %*% mu
```

B.2 MV and CVaR portfolio weights for Kurtosis

The R code generating the results presented in Table (8.2):

```
library(KernSmooth)
library(quantreg)

#set.seed(1)
n <- 500000

x1 <- rt(n,10)
x1 <- 0.05+0.02*((x1-mean(x1))/sqrt(var(x1)))
x2 <- rt(n,5)
x2 <- 0.05+0.02*((x2-mean(x2))/sqrt(var(x2)))
x3 <- rt(n,10)
x3 <- 0.09+0.07*((x3-mean(x3))/sqrt(var(x3)))
x4 <- rt(n,15)
x4 <- 0.09+0.07*((x4-mean(x4))/sqrt(var(x4)))

x <- cbind(x1,x2,x3,x4)
mu = apply(x, 2, mean)
V = var(x)

library(e1071)                    # load e1071
c(mean(x1),sqrt(var(x1)),kurtosis(x1))
c(mean(x2),sqrt(var(x2)),kurtosis(x2))
c(mean(x3),sqrt(var(x3)),kurtosis(x3))
c(mean(x4),sqrt(var(x4)),kurtosis(x4))
#___________________________________________________________________________________

#par(mfrow=c(2,2))

est1 <- bkde(x1)
est2 <- bkde(x2)
est3 <- bkde(x3)
est4 <- bkde(x4)

plot(est2,type="l",xlim=c(-0.1,0.4),lty=2,xlab="return",ylab="density",col=2)
lines(est1,col=1)
lines(est3,lty=3,col=3)
lines(est4,lty=4,col=4)
legend(0.25,25,c("Asset 1","Asset 2","Asset 3","Asset 4"),
lty=c(1,2,3,4),col=c(1,2,3,4))
#___________________________________________________________________________________

ones <- rep(1,n)
coef <- lm(ones~x1+x2+x3+x4-1)$coef
wtan <- coef/sum(coef)
```

```
wtan
wtan%*%c(mean(x1),mean(x2),mean(x3),mean(x4))
#____________________________________________________________________________________________

x2e <- x1-x2
x3e <- x1-x3
x4e <- x1-x4

#coef <- lm(x1~x2e+x3e+x4e)$coef
#coef
#wmin <- c(1-sum(coef[2:4]),coef[2:4])
#round(wmin,5)
#wmin%*%c(mean(x1),mean(x2),mean(x3),mean(x4))

Xe <- cbind(1,x2e,x3e,x4e)
#coef <- rq(x1~Xe-1, tau=0.1)$coef
#coef <- rq(x1~x2e+x3e+x4e, tau=0.1)$coef
#coef
#____________________________________________________________________________________________

#Add a pseudo obs to impose the return constraint
kappa <- 500000
x1a <- c(x1,kappa*(0.05-0.05))
Xea <- rbind(Xe,kappa*c(0,0,-0.04,-0.04))

#Mean-Variance
coef <- lm(x1a~Xea-1)$coef
coef
wMV <- c(1-sum(coef[2:4]),coef[2:4])
round(wMV,5)
wMV %*% mu

#CVaR
coef <- rq(x1a~Xea-1, tau=0.005)$coef
coef
wCV <- c(1-sum(coef[2:4]),coef[2:4])
round(wCV,5)
wCV %*% mu
```

Appendix C

Interesting divergences from Expected Utility

C.1 The Allais Paradox

The Allais Paradox was introduced by Maurice Allais in his paper *Le Comportement de l'homme rationnel devant le risque: critique des postulats et axiomes de l'école américaine* (1953). The Allais Paradox describes the fact that everyday decisions are not always consistent with the Expected Utility Theory. The main concept of the Allais Paradox is as follows:
An individual is asked to choose one between the following pair gambles:

- Gamble A: 100% chance of receiving 100 millions

- Gamble B: 10% chance of receiving 500 millions, 89% chance of receiving 100 millions, 1% chance of receiving nothing.

The same individual is asked to choose one of the following gambles as well:

- Gamble C: 11% chance of receiving 100 millions, 89% chance of receiving nothing

- Gamble D: 10% chance of receiving 500 millions, 90% chance of receiving nothing.

The expected value of each gamble is A=100, B=139, C=11 and D=50. According to the Expected Utility, the preference $A > B$(A over B) should imply that $C > D$(C over D) is preferred. This experiment indicates that most individuals would choose $A > B$ (A over B) indeed, but would choose $C < D$(D over C). In the first gamble the more certain choice is preferred over a higher expected utility, while in the second gamble a higher expected utility is preferred over a more certain choice.

C.2 The Ellsberg Paradox

The Ellsberg Paradox was developed by Daniel Ellsberg in his paper *Risk, Ambiguity, and the Savage Axioms* (1961). The Ellsberg Paradox concerns Subjective Probability Theory of Savage which is contradictory to the Expected Utility Theory. The main concept of the Ellsberg Paradox is as follows:
There is an urn containing 90 balls from which 30 are red and the remaining 60 are either black or yellow. An individual is asked to choose between the gambles:

- Gamble A: 100 € if the ball is red
- Gamble B: 100 € if the ball is black

And one between the following gambles:

- Gamble C: 100 € if the ball is not black
- Gamble D: 100 € if the ball is not red

In most cases people will choose $A > B$ (A over B) and $D > C$ (D over C). The known information (red balls) is perceived as a more certain choice against the unknown information (black/yellow balls). These choices constitute a violation of the preferences principle that require the ordering of A to B to be preserved in C to D.

In certain gambles, even though people generally prefer certainty over uncertainty, different preferences occur. People sometimes overvalue risky choices and depreciate safer ones. This leads to the result that Expected Utility Theory does not always apply in the real world.