COMPARATIVE STUDY OF RISK MEASURES

ZEHRA EKŞİ

AUGUST 2005

COMPARATIVE STUDY OF RISK MEASURES

A THESIS SUBMITTED TO

THE GRADUATE SCHOOL OF APPLIED MATHEMATICS

OF

THE MIDDLE EAST TECHNICAL UNIVERSITY

BY

ZEHRA EKŞİ

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE

IN

THE DEPARTMENT OF FINANCIAL MATHEMATICS

AUGUST 2005

Approval of the Graduate School of Applied Mathematics

Prof. Dr. Ersan AKYILDIZ
Director

I certify that this thesis satisfies all the requirements as a thesis for the degree of Master of Science.

Prof. Dr. Hayri KÖREZLİOĞLU
Head of Department

This is to certify that we have read this thesis and that in our opinion it is fully adequate, in scope and quality, as a thesis for the degree of Master of Science.

Assist. Prof. Dr. Kasırga YILDIRAK Co-Adviser	Prof. Dr. Hayri KÖREZLİOĞLU Supervisor

Examining Committee Members

Prof. Dr. Hayri KÖREZLİOĞLU

Prof. Dr. Ömer GEBİZLİOĞLU

Assoc. Prof. Dr. Azize HAYFAVİ

Assist. Prof. Dr. Kasırga YILDIRAK

Dr. Coşkun KÜÇÜKÖZMEN

Abstract

COMPARATIVE STUDY OF RISK MEASURES

Zehra Ekşi

M.Sc., Department of Financial Mathematics

Supervisor: Prof. Dr. Hayri Körezlioğlu

August 2005, 109 pages

There is a little doubt that, for a decade, risk measurement has become one of the most important topics in finance. Indeed, it is natural to observe such a development, since in the last ten years, huge amounts of financial transactions ended with severe losses due to severe convulsions in financial markets. Value at risk, as the most widely used risk measure, fails to quantify the risk of a position accurately in many situations. For this reason a number of consistent risk measures have been introduced in the literature. The main aim of this study is to present and compare coherent, convex, conditional convex and some other risk measures both in theoretical and practical settings.

Keywords: coherent risk measures, conditional convex risk measures, convex risk measures, generalized coherent risk measures, value at risk.

Öz

KARŞILAŞTIRMALI RİSK ÖLÇÜMLERİ

Zehra Ekşi
Yüksek Lisans, Finansal Matematik Bölümü
Tez Yöneticisi: Prof. Dr. Hayri Körezlioğlu

Ağustos 2005, 109 sayfa

Şüphesiz ki risk ölçümü finans sektöründe son yılların en tartışılan konusu haline gelmiştir. Finans piyasalarındaki artan işlem hacmi ve son on yılda meydana gelen finansal krizlerin risk konusuna olan ilginin artmasındaki rolü büyüktür. Risk hesaplarında en sık kullanılan ölçüm olmasına rağmen Riske Maruz Değer'in (RMD), birçok durumda, tatmin edici sonuçlar vermediği gözlemlenmiştir. Bu durum literatürde yeni risk ölçümlerinin tanımlanmasına sebep olmuştur. Bu çalışmanın temel amacı, literatürde RMD ölçümüne alternatif olarak ortaya çıkmış olan tutarlı, konveks ve koşullu konveks risk ölçümlerini incelemek ve karşılaştırmaktır.

Anahtar Kelimeler: konveks risk ölçümleri, koşullu konveks risk ölçümleri, riske maruz değer, tutarlı risk ölçümleri.

To my peerless grandpa Mehmet EKŞİ

and

his admirable son.

Acknowledgments

It is a real pleasure to thank those who have contributed in one way or another to this study.

To begin with I should like to thank Prof. Dr. Hayri Körezlioğlu, Assoc. Prof. Dr. Azize Hayfavi and Assist. Prof. Dr. Kasırga Yıldırak for their irreplaceable guidance, suggestions, comments and for giving moral support. I should also like to thank Dr. Coṣkun Küçüközmen for motivating and encouraging me throughout my research. Besides, I want to thank Prof. Dr. Ferit Öktem and Jale Körezlioğlu for their valuable help.

I acknowledge my indebtedness and express my thanks to İrem Yıldırım for her endless support and for being with me all the way.

I am also grateful to Hatice Bahadır, Özcan Baydar, Seval Çevik and Selcen Korkmaz for patiently motivating and encouraging me all the time. In addition, I want to thank Yeliz Yolcu, Ayşegül İşcanoğlu, Serkan Zeytun and Oktay Sürücü for their valuable assistance.

Also, I want to thank to all other members of the Institute of Applied Mathematics for their genialness and sincerety.

Finally, my greatest thanks are to my family-to my admirable father İsmail, my lovely mother Ayṣe, my sisters Zühal and Esra, and to my brother Mehmet-for their love, unfailing support and patience.

Table of Contents

List of Figures

List of Tables

Chapter 1

INTRODUCTION

There is a little doubt that, for a decade, risk measurement has become one of the most important topics in the finance area. Indeed, it is natural to observe such a development, since in the last ten years, huge amounts of financial transactions ended with severe losses due to severe convulsions in financial markets. This stimulated investors to seek better risk quantification methods to maximize the portfolio return and prevent underestimation of risk. Besides, day after day, regulators want to make sure that financial institutions are able to quantify and manage their risk properly and effectively.

When dealing with a problem, the first thing to do should be defining it adequately. However, there is not a unique definition of risk, making the risk measurement problem more complicated to struggle with. Therefore, for better solutions, one should understand the intuition behind this concept at least. There are various types of financial risks such as market risk, credit risk, liquidity risk, operational risk, model risk, etc. This study will mainly focus on the market risk which might be defined as the change in the value of a portfolio due to unexpected changes in market rates such as interest rates, equity prices, exchange rates, etc.

When we concentrate on the market risk the immediate figure that we encounter is Value at Risk (VaR). This is a widely used risk measure in financial markets due to its simplicity in application and interpretation. However it has some deficiencies which creates contradictions with the traditional portfolio theory. This has provided a motivation for researchers to develop new models that are consistent with the portfolio theory. The premise of the models which is coherent risk measures was introduced by Artznear, Delbaen, Eber and Heath in 1997. In their context, coherence of a risk measure symbolizes its consistency with economic intuition. With this innovation, there arose an increasing trend to develop new risk measures. Convex risk measures is one of them and a recently developed one is conditional convex risk measures. These are all academically acquiescenced measures of risk in terms of their consistency with the finance theory.

The aim of this study is to present different types of risk measures and compare them in terms of their underlying theory and practical performance. Second section will be preliminaries including the main mathematical definitions and theorems. In the third chapter, after reviewing the VaR measure, coherent, generalized coherent, convex and conditional convex risk measures are introduced by reviewing their whole underlying theory. The fourth chapter includes more of coherent risk measures that are created mostly for application purposes. In the following chapter an implementation to the Turkish Stock Market data is given. Then, the conclusion follows.

CHAPTER 2

PRELIMINARIES

First of all, it is appropriate to introduce the definitions and theorems which are commonly used throughout the study. Definitions and theorems given in this part are mainly taken from [DuSc58], [KHa01], [Ro70], [Dud89], [FS02b].

Definition 2.1. Metric
Given a set X, a metric for X is a function d from $X \times X$ into $\{x \in R : x \geq 0\}$ such that

1. for all x, $d(x, x) = 0$,
2. for all x and y, $d(x, y) = d(y, x)$
3. for all x, y and z, $d(x, z) \leq d(x, y) + d(y, z)$
4. $d(x, y) = 0$ implies $x = y$,

then d is called a **metric**. (X, d) is called a **metric space**.

Definition 2.2. Cauchy Sequence
A sequence $\{X_n\}$ in a space S with a metric d is called a **Cauchy Sequence** if

$$\lim_{n \to \infty} \sup_{m \geq n} d(x_m, x_n) = 0.$$

Definition 2.3. Complete Metric Space
The metric space (S, d) is **complete** iff every Cauchy sequence in it converges.

Definition 2.4. Compactness

Let X be a set and A a subset of X. A collection of sets whose union includes A is called a **cover** of A. Then the topological space is **compact** iff every open cover has a finite subcover.

Theorem 2.1. If $(K, \mathcal{T})$ is a **compact** topological space and F is a closed subset of K, then F is **compact**.

Theorem 2.2. If (K, T) is **compact** and f is **continuous** from K into another space L, then L is **compact**.

Definition 2.5. Continuity

Given a topological spaces $(X, \mathcal{T})$ and $(Y,\mathcal{U})$, a function f from X into Y is called **continuous** iff for all $U \in \mathcal{U}, f^{-1}(U) \in \mathcal{T}$.

Theorem 2.3. A convex function finite on all of R^n is necessarily continuous. (To learn more about convex analysis see, [Ro70].)

Definition 2.6. Finitely Additive Set Functions

A **set function** is a function defined on a family of sets, and having values either in a Banach Space, which may be the set of real or complex numbers, or in the extended real number system, in which case its range contains one of the improper values $+\infty$ and $-\infty$. A set function μ defined on a family τ of sets is said to be **finitely additive** if τ contains the void set $\varnothing$, if $\mu(\varnothing) = 0$ and if $\mu(A_1 \cup A_2 \cup ... \cup A_n) = \mu(A_1) + \mu(A_2) + ...\mu(A_n)$ for every finite family $\{A_1, A_2, ...A_n\}$ of disjoint subsets of τ whose union is in τ. Thus all such sums must be defined, so that there can not be both $\mu(A_i) = -\infty$ and $\mu(A_j) = +\infty$ for some i and j.

Definition 2.7. Countably Additive Functions

Let μ be a finitely additive, real valued function on an algebra $\mathcal{A}$. Then μ is **countably additive** iff μ is continuous at $\emptyset$, that is $\mu(A_n) \to 0$ whenever $A_n \downarrow \emptyset$ and $A_n \in \mathcal{A}$.

Definition 2.8. Sigma Algebra

Given a set X, a collection $\mathcal{A} \subset 2^X$ is called a **ring** iff $\emptyset \in \mathcal{A}$ and for all A and

B in $\mathcal{A}$, we have $A \cup B \in \mathcal{A}$ and $B \setminus A \in \mathcal{A}$. A ring $\mathcal{A}$ is called an **algebra** iff $X \in \mathcal{A}$. An algebra called a **σ-algebra** if for any sequence $\{A_n\}$ of sets in $\mathcal{A}, \bigcup_{n\geq 1} A_n \in \mathcal{A}$.

Definition 2.9. Measurable Space

A **measurable space** is a pair $(X, \mathcal{Y})$ where X is a set and $\mathcal{Y}$ is a σ-algebra of subsets of X.

Definition 2.10. Measure, Measure Space

A countably additive function μ from a σ-algebra $\mathcal{Y}$ of subsets of X into $[0, \infty]$ is called a **measure**. Then $(X, \mathcal{Y}, \mu)$ is called a **measure space.**

Definition 2.11. Atom

If $(X, \mathcal{Y}, \mu)$ is a measure space, a set $A \in \mathcal{Y}$ is called an **atom** of μ iff $0 < \mu(A) < \infty$ and for every $C \subset A$ with $C \in \mathcal{Y}$, either $\mu(C) = 0$ or $\mu(C) = \mu(A)$.

Definition 2.12. Total Variation

The **total variation** of a finitely additive set function μ is defined as:

$$sup\{\sum_{i=1}^{n} |\mu(A_i)| \mid A_1, ..., A_n \ \ disjoint \ \ sets \ \ in \ \ \mathcal{F}, n \in N\}$$

Theorem 2.4. Let $\mathcal{X}$ be the space of all bounded measurable functions on $(\Omega, \mathcal{F})$ and $ba(\Omega, \mathcal{F})$ denote the space of all finitely additive functions with finite total variation. The integral

$$l(F) = \int F d\mu \quad F \in \mathcal{X}$$

defines a one to one correspondence between continuous linear functional l on $\mathcal{X}$ and finitely additive set functions $\mu \in ba(\Omega, \mathcal{F})$.

(For the proof of theorem, see [FS02b, p:395].)

Definition 2.13. Probability Measure, Probability Space

A measurable space $(\Omega, \mathcal{Y})$ is a set Ω with a σ-algebra $\mathcal{Y}$ of subsets of Ω. A **probability measure** P is a measure on $\mathcal{Y}$ with $P(\Omega) = 1$. Then $(\Omega, \mathcal{Y}, P)$ is called a **probability space**. Members of $\mathcal{Y}$ are called **events** in a probability space.

Definition 2.14. Random Variable

If $(\Omega, \mathcal{A}, P)$ is a probability space and $(S, \mathcal{B})$ is any measurable space, a measurable function X from Ω into S is called a **random variable**. Then the image measure $P \circ X^{-1}$ defined on $\mathcal{B}$ is the probability measure which is called the **law** of X.

Definition 2.15. Bayes Formula

Let A and B two events on $\mathcal{A}$, then the conditional probability of A given B is equal to

$$P(A|B) : \frac{P(A \cap B)}{P(B)}$$

and this is a specific representation of Bayes formula for the case of two events.

Definition 2.16. Absolutely Continuous Probability Measures

Let P and Q be two probability measures on measurable space $(\Omega, \mathcal{F})$. Q is said to be **absolutely continuous** with respect to P, if for $A \in \mathcal{F}$,

$$P(A) = 0 \Longrightarrow Q(A) = 0.$$

Theorem 2.5. (Radon-Nikodym)

Q is absolutely continuous with respect to P on $\mathcal{F}$ if and only if there exists an $\mathcal{F}$ measurable function $\varphi \geq 0$ such that

$$\int F dQ = \int F \varphi dP$$

for all $\mathcal{F}$-measurable functions $F \geq 0$.

The function φ is called the Radon-Nikodym derivative of Q with respect to P and we write $\frac{dQ}{dP} := \varphi$.

Definition 2.17. Seminorm, Norm

Let X be a real vector space. A **seminorm** on X is a function $\| \, . \, \|$ from X into $[0, \infty[$ such that

1. $\parallel cx \parallel = |c| \parallel x \parallel$ forr all $c \in R$ and $x \in X$

2. $\parallel x + y \parallel \leq \parallel x \parallel + \parallel y \parallel$ for all $x, y \in X$

A seinorm is called **norm** iff $\parallel x \parallel = 0$ only for $x = 0$.

Definition 2.18. $\mathcal{L}^p$ Spaces
Let X be a random variable defined on $(\Omega, \mathcal{F}, P)$, then the following norms are defined as

$$\|X\|_p = (E_P|X|^p)^{\frac{1}{p}} = [\int_\Omega |X|^p dP]^{\frac{1}{p}} \quad \text{for} \quad 0 \leq p < \infty$$

and for p=∞

$$\|X\|_\infty = ess.sup|X|.$$

where

$$ess.sup := \sup\{x \in R^+ \mid P\{|X| > x\} > 0\}$$

$\mathcal{L}^p(\Omega, \mathcal{F}, P)$, or just $\mathcal{L}^p$, denotes the real vector space of equivalence classes of all random variables X on (Ω,$\mathcal{F}$,P), which fulfill $\|X\|_p < \infty$.
For p$\geq$ 1, $\|.\|$ is the so called **$\mathcal{L}^p$-norm** of the vector space $\mathcal{L}^p$. We call the elements of $\mathcal{L}^p$ *p-integrable* random variables or functions on $(\Omega, \mathcal{F}, P)$.

Definition 2.19. Banach Space
If X is a vector space and $\parallel . \parallel$ is a norm on it, then $(X, \parallel . \parallel)$ is called a **normed linear space**. A **Banach space** is a normed linear space which is complete, for the metric defined by the norm.

Theorem 2.6. Completeness of $\mathcal{L}^p$
For any measure space $(\Omega, \mathcal{F}, P)$ and $1 \leq p \leq \infty$, $(\mathcal{L}^p(\Omega, \mathcal{F}, P), \parallel . \parallel_p)$ is a Banach space.

Theorem 2.7. Dominated Convergence Theorem
Let $\{X_n, n \in N\}$ be a sequence of extended real random variables. If $\{X_n, n \in N\}$ converges and if there is an integrable positive extended real random variable

T such that for all n ,$|X_n| \leq T$, then

$$E(\lim_{n\to\infty} X_n) = \lim_{n\to\infty} E(X_n)$$

(For the proof of the theorem see, [KHa01, p:110])

Definition 2.20. Topology, Base for a Topology
Given a set X, a **topology** on X is a collection $\mathcal{F}$ of subsets of X (in other words, $\mathcal{F} \subset 2^X$) such that

1. $\emptyset \in \mathcal{F}$ and $X \in \mathcal{F}$.

2. For every $U \in \mathcal{F}$ and $V \in \mathcal{F}$ we have $U \cap V \in \mathcal{F}$

3. For any $\mathcal{U} \subset \mathcal{F}$, we have $\bigcup \mathcal{U} \in \mathcal{F}$.

A **base** for a topology $\mathcal{F}$ is any collection $\mathcal{U} \subset \mathcal{F}$ such that for every $V \in \mathcal{F}, V = \bigcup\{U \in \mathcal{U} : U \subset V\}$.

Definition 2.21. Topological Vector Space
A linear space E which carries a topology is called a **topological vector space** if every singleton $\{x\}$ for $x \in E$ is a closed set, and if addition is continuous from $E \times E$ and multiplication by scalars is continuous from $R \times E$.

Definition 2.22. Convex Hull
If A is a subset of the linear space X, the set $Co(A)$, called the **convex hull** of A, is the intersection of all convex sets containing A; If X is a linear topological space, the set $\bar{C}o(A)$ called closed convex hull of A, is the intersection of all closed convex sets containing A. Hence $Co(A)$ is the set of all convex combination of points of A. We have $Co(\alpha A) = \alpha Co(A)$ and $Co(A + B) = Co(A) + Co(B)$.

Definition 2.23. Locally Convex Spaces
A topological vector space is called a **locally convex space** if its topology has a base consisting of convex sets.

Remark 2.1. Any Banach space is locally convex.

Theorem 2.8. Separating Hyperplane Theorem

Suppose A and B are non empty convex subsets of R^n, that A is open and that the intersection of A with the interior is empty. Then there exists a non-zero vector $v \in R^n$ and a constant c such that:

$$A \subseteq \{x \in R^n \mid v'x > c\}$$

$$B \subseteq \{x \in R^n \mid v'x \leq c\}$$

Theorem 2.9. Separation Theorem

Being any two distinct convex sets in a topological vector space $\mathcal{E}$, B and C, one of which has an interior point, can be separated by a nonzero, continuous linear functional l on $\mathcal{E}$ i.e.

$$\sup_{y \in C} l(y) \leq \inf_{z \in B} l(z)$$

In fact this theorem is a special case of the separating hyperplane theorem.

(For the details of theorem, see [Dud89, p:150].)

Theorem 2.10. Hahn-Banach

Suppose that B and C are two non empty, disjoint, and convex subsets of a locally convex space E. Then, if B is compact and C is closed, there exists a continuous linear functional l on E such that

$$\sup_{x \in C} l(x) < \inf_{y \in B} l(y)$$

This theorem is also one variant of the separating hyperplane theorem.

Definition 2.24. Dual of a Locally Convex Space

On a locally convex space E, the collection

$$E' := \{l : E \to R \mid l \;\; is \;\; continuous \;\; and \;\; linear\}$$

separates the points of E, i.e for any two distinct points $x, y \in E$ there exist some $l \in E'$ such that $l(x) \neq l(y)$. The space E' is called the **dual space** of E

Theorem 2.11. Bipolar Theorem

For E and F being duals and $X \subset E$, then $X^{\circ\circ}$ is the $\sigma(E, F)$-closure of absolutely convex hull of X. In other words, the bipolar theorem states that the bipolar of a subset of locally convex space equals its closed convex hull.

Chapter 3

RISK MEASURES

In finance we are often exposed to risk whenever we take financial positions. Therefore, we have to measure the riskiness of our position to decide if it is acceptable or not. So, what is a risk measure? A **risk measure** is a measure that assigns a value to the risk of a position generally with a pre-determined probability. To understand this definition, first of all one should understand the concepts of *risk* and *measure*.

Although in the finance literature there is not a unique definition for risk, it is commonly accepted that risk has two components.

- exposure, and
- uncertainty

For the better understanding of the risk concept, let us examine the following example which was given in [Hol03, p:21]:

> If we swim in shark-infested waters, we are *exposed* to bodily injury or death from a shark attack. We are *uncertain* because we do not know if we will be attacked. Being both exposed and uncertain, we face risk.

Measure is an operation that assigns value to something and the interpretation of this value is a **metric**. After understanding the intuitive meaning, we will continue with the mathematical definition of a risk measure.

Definition 3.1. Financial Position
Let Ω be a fixed set of scenarios. A financial position is described by a mapping

$$X : \Omega \rightarrow \mathbf{R}$$

where X(w) is the discounted net worth of the position at the end of the trading period if the scenario w $\in \Omega$ is realized.

Definition 3.2. Risk Measure
Let $\mathcal{G}$ represents the set of all positions, that is a set of all real valued functions on Ω, then a risk measure ρ is any mapping from the set of all random variables in to real line, ie.

$$\rho : \mathcal{G} \rightarrow \mathcal{R}$$

For a decade, risk measurement has become a very popular concept in the finance area. Hence a vast number of risk measures were proposed in literature. In this chapter, the static risk measures which are Value-at-Risk, coherent risk measures, generalized coherent risk measures and convex conditional risk measures will mainly be introduced and compared. VaR is the most popular risk measure in practice. Besides, it is one of the oldest risk measures. Thus, it is appropriate for the structure of this study to begin with VaR measure. After this, coherent, generalized coherent, convex and conditional convex risk measures will be considered which were introduced as alternatives to VaR since it has some deficiencies which can not be accepted by the finance theory. Before starting, it is beneficial to remind the reader that financial activities consist of a variety of risks such as market risk, credit risk, liquidity risk, operational risk, model risk, etc. This research will mainly focus on market risk.

3.1 Value-at-Risk

In the late 1970's and 1980's, many major financial institutions started to work on models to measure risk. They did so for their internal risk measurement purposes since as firms became more complex, it was becoming increasingly difficult and important to measure the total financial risk they are exposed to. The term Value-at-Risk(VaR) did not enter finance literature until the early 1990s because firms developed their risk models for internal purposes and they did not publish their models to outsiders. For this reason, tracing the historical development of VaR measures is difficult. Whilst, most firms kept their models secret, in October 1994, JP Morgan decided to make its Risk Metrics system available on internet so outside users could access the model and use it for their risk management purposes. As the VaR was introduced to the market, software suppliers who had received advanced notice started promoting compatible software. Timing for the release of this document and software was excellent, because the last few years and the following of couple years witnessed the declaration of huge financial losses of some rooted firms. The following part is taken just to give an idea and longer list can be found in Holton's book [Hol03].

> In February 1993 Japan's Showa Shell Sekiyu oil company declared a USD 1050 MM loss from speculating on exchange rates. In 1994 the most considerable loss was reported by California's Orange County.

Finally, the dominance of VaR in the market became inevitable by the approval of the limited use of VaR measures for calculating bank capital requirements in 1996 by the Basle Committee. Thus VaR has become the most widely used financial risk measure.

3.1.1 Definition and Properties of VaR

In financial terms VaR is the maximum potential loss in the value of a portfolio given the specification of normal market conditions, the time horizon and statistical confidence level α.The basic concept was nicely described in [Do02,p:10] as:

> Value at risk is a single, summary, statistical measure of possible portfolio losses. Specifically, value at risk is a measure of losses due to normal market movements. Losses greater than the value at risk are suffered only with a specified small probability. Subject to the simplifying assumptions used in its calculation, value at risk aggregates all of the risks in a portfolio into a single number suitable for use in the boardroom, reporting to regulators, or disclosure in an annual report. Once one crosses the hurdle of using a statistical measure, the concept of value at risk is straightforward to understand. It is simply a way to describe the magnitude of the likely losses on the portfolio.

After giving the brief history and financial definition of VaR, it is appropriate to focus on the mathematical representation and properties of a VaR measure.

Definition 3.3. Quantiles
Let X be a random variable on (Ω,F,P) and $\alpha \in$ (0,1) then the α-quantile of the random variable X is any real number q with the property

$$P[X \leq q] \geq \alpha \;\; and \;\; P[X < q] \leq \alpha \tag{3.1.1}$$

The set of all α-quantiles of X is an interval [q_α^-(X) , q_α^+(X)], where

$$q_\alpha^-(X) := \inf\{x \mid P[X \leq x] \geq \alpha\} \tag{3.1.2}$$

and

$$q^+_\alpha(X) := \inf\{x \mid P[X \leq x] > \alpha\} = sup\{x \mid P[X < x] \leq \alpha\} \qquad (3.1.3)$$

Definition 3.4. Value at Risk
For a fixed confidence level $\alpha \in$ (0,1) and a financial position X, the Value at Risk at level α is defined as:

$$VaR_\alpha(X) := -q^+_\alpha(X) = q^-_{1-\alpha}(-X) = inf\{m \mid P[X + m < 0] \leq \alpha\} \qquad (3.1.4)$$

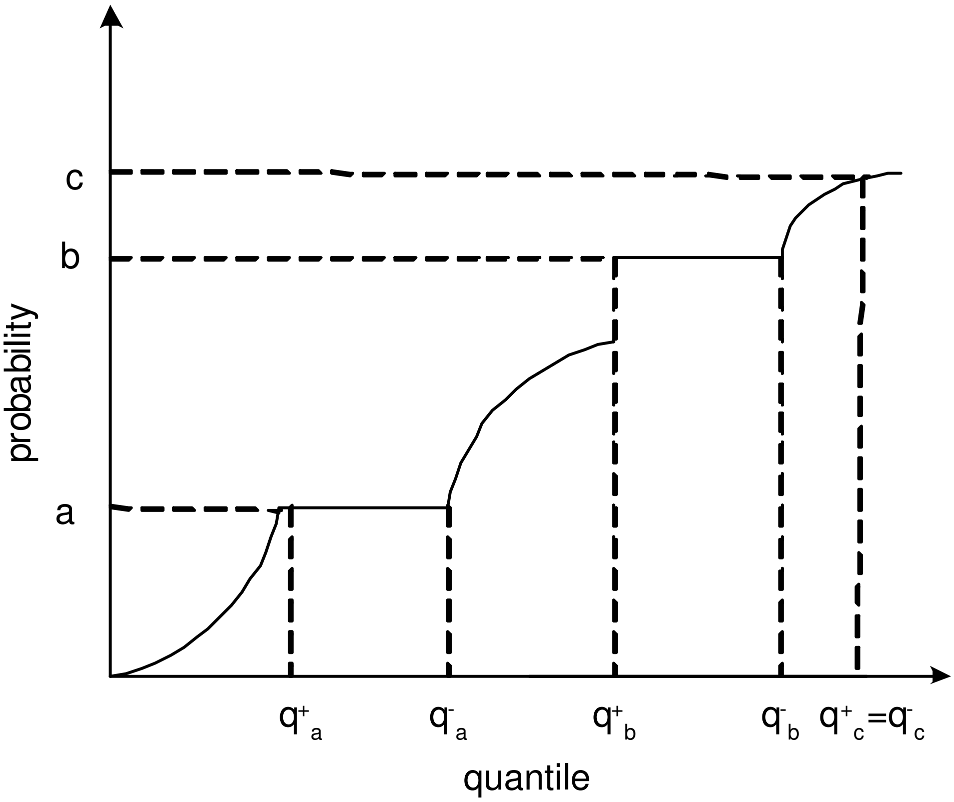

Figure 3.1: VaR estimation in the case of discontinuous distribution

For the case where the distribution of returns are continuous, we have $q^-_\alpha(X) = q^+_\alpha(X)$ and realizing the related VaR figure in the distribution plot is easier. However, in the discrete case which implies the existence of jumps in the distribution, it is a little more complicated to determine the VaR figure directly. VaR estimates for such cases seem like the ones in *Figure 3.1.* From the equation (3.1.4) it can be interpreted that $VaR_\alpha(X)$ is the smallest amount of capital which, if added to the position X, keeps the probability of a negative outcome below the level α. Although VaR is capable of measuring the worst loss which can be expected with probability (1-α), it fails to address how large this loss can be if the α probability events occur.

Proposition 3.1. VaR has the following properties;

1. VaR_α is translation-invariant, i.e

$$VaR_\alpha(X + c) = VaR_\alpha(X) - c \quad c \in R$$

2. VaR_α is positively homogenous, i.e

$$VaR_\alpha(cX) = cVaR_\alpha(X)$$

whenever c>0.

3. VaR_α is monotonic, i.e if

$$X_1 \geq X_2$$

then

$$VaR_\alpha(X_1) \leq VaR_\alpha(X_2)$$

4. VaR_α is comonotone additive, i.e, if X_1 and X_2 are comonotone then

$$VaR_\alpha(X_1 + X_2) = VaR_\alpha(X_1) + VaR_\alpha(X_2)$$

Proof: 1. It is obvious that q_α^+(X+c) = q_α^+(X)+c and VaR is defined as -q_α^+(X), from these observations, condition 1 follows.

2. Due to the statistical properties of quantiles, as in (1), the result is obvious.

3. If we have two positions X_1 and X_2 such that $X_1 > X_2$ under all set of scenarios then;

$VaR_\alpha(X_1) = inf\{m \mid P[X_1 + m < 0] \leq \alpha\} < inf\{n \mid P[X_2 + n < 0] \leq \alpha\}$
$= VaR_\alpha(X_2)$

4. Let $X_1 = f(U)$ and $X_2 = g(U)$ with U uniform on$[0,1]$ and f monotonically increasing,then $VaR_\alpha(X_1) = f(\alpha)$. Similarly $VaR_\alpha(X_2) = g(\alpha)$ and therefore $VaR_\alpha(X_1 + X_2) = f(\alpha) + g(\alpha) = VaR_\alpha(X_1) + VaR_\alpha(X_2)$ □

Remark 3.1. Let V be a real vector space. A real-valued function f on V is called **subadditive** if

$$f(X+Y) \leq f(X) + f(Y) \quad \forall \quad X, Y \in V \tag{3.1.5}$$

The fourth property in Proposition 3.1 implies that VaR is not sub-additive because the total risk of two portfolios is not less than the summation of the risks of the two portfolios. The financial interpretation of this result is that VaR does not encourage diversification. This is a big contradiction between VaR and portfolio theory. Föllmer and Schied in [FS02b, p:180] gave the following numerical example to show how VaR penalizes diversification.

> Consider an investment into two defaultable corporate bonds, each with return $\tilde{r} > r$, where $r \geq 0$ is the return on a riskless investment. The discounted net gain of an investment w>0 in the i_{th} bond is given by
>
> $$X_i = \begin{cases} w & \text{in case of default} \\ \frac{w(\tilde{r}-r)}{1+r} & \text{otherwise} \end{cases}$$
>
> If a default of the first bond occurs with probability $p \leq \lambda$, then
>
> $$P[X_1 - \frac{w(\tilde{r}-r)}{1+r} < 0] = P[default] = p \leq \lambda$$
>
> Since in Definition 3.3 we defined the VaR of a portfolio at confidence level α as
>
> $$VaR_\alpha(X) = inf\{m \mid P[X + m < 0] \leq \alpha\}$$

$VaR_\lambda(X_1)$ becomes

$$VaR_\lambda(X_1) = -\frac{w(\tilde{r} - r)}{1 + r}$$

which is always less than zero.
Since the VaR at the level λ is a negative number we can say that the position X_1 is acceptable in the sense that it does not carry a positive risk.

Now let us see what will happen to the VaR of the portfolio in the case of diversification. Assume that the initial portfolio diversified by investing the amount w/2 into each of the two bonds. Also assume that the two bonds default independently of each other, each of them with probability p. Now the new position we take equals to $Y := (X_1 + X_2)/2$. In this example we have the following sample space

$$\Omega = (\{d, d\}, \{d, nd\}, \{nd, d\}, \{nd, nd\})$$

where d indicates the occurrence and nd indicates the non-occurrence of default. With this set of information we can write the cumulative distribution function of the random variable Y as:

$$P[Y \leq y] = \begin{cases} -w & p^2 \\ \\ \frac{w(\tilde{r}-r)}{2(r+1)} - \frac{w}{2} & p(2-p) \\ \\ \frac{w(\tilde{r}-r)}{r+1} & 1 \end{cases}$$

Here, the probability that Y is negative is equal to the probability that at least one of the two bonds defaults: $p(1-p)+p(1-p)+p^2 = p(2-p)$. If for instance we take p=0.009 and λ=0.01 and use these values in the distribution function above we find;

$$VaR_\lambda(Y) = -\frac{w(\tilde{r}-r)}{2(r+1)} + \frac{w}{2}$$

This is definitely a positive number which shows that VaR may strongly discourage diversification: it penalizes quite drastically the increase of the probability that something goes wrong, without rewarding the significant reduction of the expected loss conditional on the event of default. To show this, let us calculate the expected loss of both diversified and non-diversified portfolios conditional on the event of default. The expected loss from the first portfolio on the event of default is equal to:

$$E[X_1|default] = -w$$

To find the conditional expectation of the second portfolio under the condition of default, firstly one should calculate the conditional probabilities. By using the famous Bayes Formula (see Preliminaries, Def: 2.15) one can find the following conditional probabilities easily;

$$P[\{d,n\} \ | \ d] = P[\{n,d\} \ | \ d] = \frac{(1-p)}{(2-p)}$$

$$P[\{d,d\} \ | \ d] = \frac{p}{(2-p)}$$

After finding the probabilities, conditional expectation of the random variable Y can be calculated as:

$$E[Y \ | \ d] = 2.\frac{(1-p)}{(2-p)}.\frac{w(\tilde{r}-r)}{2(1+r)} + \frac{p}{(2-p)}.(-w) > -w = E[X_1 \ | \ d]$$

As we expect, in the case of diversification there is a reduction in the expected loss conditional on the event of default. Thus, one can conclude that optimizing a portfolio with respect to VaR may lead to a concentration of the portfolio in one single asset with a sufficiently small default probability, but with an exposure to large loss.

Remark 3.2. Let V be a real vector space and C a convex set in V. A real-valued function f on C is called **convex** if

$$f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda)f(y) \tag{3.1.6}$$

for every x and y in C and $0 \leq \lambda \leq 1$.

In the risk minimization process we need the risk surface to be convex, since only if the surface is convex we will have a unique minimum. Hence the risk minimization process will always pick up a unique optimal solution [AcT02a]. For the risk surface to be convex one necessary condition is that the risk measure should satisfy the convexity property. Positive homogeneity and sub-additivity together ensure the convexity of a function. Although VaR is positively homogenous, it is not subadditive in general so it fails to satisfy the convexity property. Thus, the risk surface generated by the VaR measure is not convex. This is another problem arising from the fact that VaR is not sub-additive.

Proposition 3.2. Consider Value at Risk as a measure of risk on the linear space $\mathcal{X} := L^2(\Omega, F, P)$. For a random variable X $\in \mathcal{X}$ with normal distribution N(m,σ^2), its VaR at level α is given by

$$VaR_{\alpha}(X) = -m + \sigma\Phi^{-1}(1-\alpha) \tag{3.1.7}$$

where Φ^{-1} indicates the inverse of the standard normal distribution function Φ; ($\Phi \sim N(0,1)$). Now let us consider a Gaussian subspace $\mathcal{X}_0$, i.e., a linear space $\mathcal{X}_0 \subset \mathcal{X}$ consisting of normally distributed random variables. For $X_1, X_2 \in \mathcal{X}_0$ *and* $\gamma \epsilon [0,1]$, the convex combination $X := \gamma X_1 + (1-\gamma)X_2$ belongs to $\mathcal{X}_0$, and so it is normally distributed with mean

$$m = \gamma m_1 + (1-\gamma)m_2 \tag{3.1.8}$$

and standard deviation

$$\sigma(X) \le \gamma\sigma(X_1) + (1-\gamma)\sigma(X_2) \tag{3.1.9}$$

and its Value at Risk satisfies

$$\begin{aligned} VaR_\alpha(X) &= -m + \sigma(X)\Phi^{-1}(1-\alpha) \\ &\le \gamma VaR_\alpha(X_1) + (1-\gamma)VaR_\alpha(X_2) \end{aligned} \tag{3.1.10}$$

Note that the equation (3.1.10) is the description of usual Risk Metrics VaR. The inequality sign indicates that VaR_α satisfies the convexity property if it is restricted on the Gaussian subspace $\mathcal{X}_0$ and if α belongs to $(0, \frac{1}{2}]$.

Proof: Let us prove (3.1.8), (3.1.9), (3.1.10) in turn. If X_1 and X_2 are two normally distributed random variables with $X_1 \sim N(m_1, \sigma_1)$ and $X_2 \sim N(m_2, \sigma_2)$ and $X := \gamma X_1 + (1-\gamma)X_2$ then;

$$\begin{aligned} E[X] = m &= E[\gamma X_1 + (1-\gamma)X_2] \\ &= \gamma E[X_1] + (1-\gamma)E[X_2] \\ &= \gamma m_1 + (1-\gamma)m_2 \end{aligned} \tag{3.1.11}$$

which proves (3.1.8)

Let ρ be the correlation coefficient of X_1 and X_2,

$$\rho = \frac{\sigma(X_1, X_2)}{[\sigma(X_1)\sigma(X_2)]}$$

then,

$$\begin{aligned}\sigma^2(X) &= \sigma^2(\gamma X_1 + (1-\gamma)X_2) \\ &= \gamma^2\sigma^2(X_1) + (1-\gamma)^2\sigma^2(X_2) + 2\gamma(1-\gamma)\sigma(X_1)\sigma(X_2)\rho \quad (3.1.12) \\ &\leq \gamma^2\sigma^2(X_1) + (1-\gamma)^2\sigma^2(X_2) + 2\gamma(1-\gamma)\sigma(X_1)\sigma(X_2) \\ &= (\gamma\sigma(X_1) + (1-\gamma)\sigma(X_2))^2\end{aligned}$$

Thus the inequality in (3.1.9) is proved.

Now we are ready to prove (3.1.10).

$$\begin{aligned}VaR_\alpha(X) &= VaR_\alpha(\gamma X_1 + (1-\gamma)X_2) \\ &= -m + \sigma(X)\Phi^{-1}(1-\alpha)\end{aligned}$$

where m and $\sigma^2(X)$ are defined on equations (3.1.11), (3.1.12) respectively. We have;

$$VaR_\alpha(X_i) = -m_i + \sigma(X_i)\Phi^{-1}(1-\alpha)$$

and since VaR is positively homogenous we can write;

$$VaR_\alpha(\gamma X_1) + VaR_\alpha((1-\gamma)X_2) = \gamma VaR_\alpha(X_1) + (1-\gamma)VaR_\alpha(X_2)$$

so the equation becomes;

$$\begin{aligned}&= \gamma(-m_1 + \sigma(X_1)\Phi^{-1}(1-\alpha)) + (1-\gamma)(-m_2 + \sigma(X_2)\Phi^{-1}(1-\alpha)) \\ &= -m + \Phi^{-1}(1-\alpha)[\gamma\sigma(X_1) + (1-\gamma)\sigma(X_2)]\end{aligned}$$

by using the information on equation (3.1.9) and making the assumption that $\Phi^{-1}(1-\alpha) \geq 0$, which means that $\alpha \in (0, \frac{1}{2}]$, we can complete the last part of the proof.

$$VaR_\alpha(X) = \gamma VaR_\alpha(X_1) + (1-\gamma)VaR_\alpha(X_2)$$
$$= -m + \Phi^{-1}(1-\alpha)[\gamma\sigma(X_1) + (1-\gamma)\sigma(X_2)]$$

$$\geq -m + \sigma(X)\Phi^{-1}(1-\alpha)$$

$$= VaR_\alpha(\gamma X_1 + (1-\gamma)X_2)$$

Therefore VaR_α does satisfy the axiom of convexity if restricted to the Gaussian space an if α belongs to $(0, \frac{1}{2}]$. □ Embrechts et. al. extended this property and stated that VaR is subadditive in the case in which the joint distribution of returns is elliptic. They also added that in such cases a VaR minimizing portfolio coincides with the Markowitz variance-minimizing portfolio [EMS99, p:72]. So it can be concluded that VaR gives accurate results only when the computationally simpler variance can also do.

After examining the mathematical properties of VaR and realizing its failures in the sequel it is appropriate to summarize these weaknesses one by one. In general VaR is not an adequate risk measure, since:

1. By definition, VaR depends on the reference probability. Thus, an explicit probability on sample space Ω is used to construct it. This makes VaR a model dependent measure of risk. Therefore, it fails to assess the risk of a position in a situation of uncertainty, in the case where no probability measure is known a priori.

2. It can not measure the losses beyond pre-specified probability.

3. It is not subadditive in terms of position, thus portfolio diversification may lead to an increase of risk and this problem prevent us to allocate economic capital to different business units effectively. Moreover, its nonadditivity by risk variable prevents to add up the VaR of the different risk sources (for instance, for a convertible bond, VaR is not simply the sum of its interest rate and equity VaR).

4. It is not convex, so it is impossible to reach one unique global minimum in optimization problems.

5. It is not consistent because it may give conflicting results at different confidence levels.

6. There may arise some computational difficulties in VaR estimations.

7. Since it is a quantile-based variable it is affected by large statistical error, in particular when the chosen probability level is very small [AcNSi01].

In spite of these serious problems, VaR is a widely used risk measure due to the fact that it measures downside risk, it is a single number that represents risk level and it is easy to interpret.

As stated above, although it is not an adequate measure, VaR is widely used for quantifying financial risk. For this reason, in the following subsection different types of VaR computation methods will be introduced briefly for the readers who are more interested in the application.

3.1.2 VaR Estimation Methods

This section deals with some of the most popular methods that are used in the estimation of VaR. To begin with, there are 3 main approaches to VaR estimation which are parametric, non-parametric and simulation approach.

i. Parametric Models

In this approach, firstly it is required to specify explicitly the statistical distribution from which our data observations are drawn. After specifying the distribution the second task is estimating the parameters that fit the data to the pre-specified distribution. Then, fitting curves through the data, we can read the VaR estimate from the fitted curve. In this procedure the crucial points are analyzing the data carefully and determining the best statistical distribution which applies to our data set.

When fitting theoretical distributions to our data set, we can consider two kinds of fitting procedures, conditional and unconditional. If we fit a distribution conditionally this means we are thinking that the chosen distribution is able to explain the behavior of our whole data set. Whereas in the conditional fit, we first adjust our data set and then fit the distribution to the adjusted data set. By standardizing the data set we are able to eliminate the seasonal, holiday and volatility effects from the data set so that the data might fit the chosen distribution better.

Normal, Student-t, Lognormal and Extreme Value distributions are the mostly used statistical distributions in the parametric approach. All these procedures are widely used since they supply clear formula for VaR estimates. The estimation procedure for all cases are the same: first fit the data to the distribution then use the estimated parameters in the given VaR formula. For interested readers detailed information about these procedures is given in the very precious study of Dowd [Do02].

The parametric approach has both advantages and disadvantages. One of the advantages of this approach is that it is easy to use since it gives rise to straightforward VaR formulas. The second important advantage is that the model enables us to use the additional information coming from the assumed distribution function, so it is not too data hungry. The main weakness of this approach is that it depends on parametric assumptions, therefore setting unrealistic assumptions may lead to serious problems.

ii. Non-Parametric Models

Contrary to the parametric approach, non-parametric models do not assume any underlying theoretical distribution and use an empirical distribution of prices. They are only based on the assumption that the near future will look sufficiently like the past so one can use the historical data to forecast the risk over the near future. The first and the most popular non-parametric approach is *historical simulation*(HS) which is a histogram-based approach. In HS, the procedure is quite easy: if you have N number of profit/loss data, rank them in ascending order, then the VaR value for the confidence level α is equal to the$(N.\alpha)^{th}$ highest loss value. Two important properties of this method is that it gives equal weight to all past observations and it assumes i.i.d distributed observations. These may cause some serious problems if we are faced with some seasonality, volatility effects and extreme observations in our data set. To show this problem, Dowd in [Do02,p:61] gave the following example:

> It is well known that natural gas prices are usually more volatile in winter than in the summer so a raw HS approach that incorporates both summer and winter observations will tend to average the summer and winter observations together. As a result, treating all observations as having equal weight will tend to underestimate true risk in the winter, and overestimates them in the summer. What we should do is to give more weight to winter observations if we are estimating a VaR in winter.

As a response to this problem, Boudoukh et. al.introduced the *Age-weighted Historical Simulation* in which the i^{th} observation is weighted with the factor λ^{i-1} where λ is the exponential rate of decay and lies on [0,1] [BRW98]. After weighting all data set one can estimate the VaR value by applying the HS method to the adjusted data set. Although this model is better than the traditional HS model, it is not able to capture the recent changes in volatility.

Hull and White suggested the *Volatility-weighted Historical Simulation* (FHS) in [HW98]. They argued that weighting the observations with their volatility estimates is a better way for producing risk estimates that are sensitive to current volatility estimates. The procedure they suggested is first to estimate the volatility for each observation by using GARCH or EWMA models, then calculate a weighting factor for the i_{th} day by dividing the last days' volatility estimate to the i_{th} days'. After this, multiply each observation with its weighting factor. Finally apply HS to this new volatility -adjusted data set. Empirical results indicates that this approach gives better results than HS.

Another alternative solution to the volatility problem is the *Filtered Historical Simulation* first proposed by Barone-Adessi in [BaGV99]. This model is different from the ones mentioned up to here. The models mentioned above necessitate sacrificing the benefits of non-parametric models to cope with the volatility problem, because for performing these models one has to deal with GARCH specification procedure which does not coincide with the logic of the non-parametric approach. Dowd described the properties of the FHS model in [Do02, p:65] very clearly:

> FHS combines the benefit of the HS with the power and flexibility of conditional volatility models such as GARCH. It does so by bootstrapping returns within a conditional volatility framework, where the bootstrap preserves the non-parametric nature of HS, and the volatility model gives us a sophisticated treatment of volatility.

From the theoretical aspect this model is a little more complicated than the others but in practice it is easy to apply. For interested readers, [Do02, p:65] would be a guide for understanding the theory and application steps of FHS.

In general the majority of the historical simulation methods are conceptually simple and easy to apply. Besides they have a wide range of application area since they can be applied to all kinds of data, e.g., fat-tailed,leptokurtic, etc. These are all advantages of the approach. Of course, it has some disadvantages too. First of all, these methods are data hungry implying that they necessitate large amount of observed data. Thus, they may not be applicable for new market instruments. However, using large numbers of data may also create problems since current market observations may be drawn out by the observation of older examples. This may create lags in the adjustment process of the model to current market conditions. In short, when we are dealing with these models we face the problem of appropriate sample size. Finally, this approach may underestimate the VaR since it makes no allowance for the possible events that might occur but did not occur in the past.

iii. Simulation Models

In VaR estimation *Monte Carlo Simulation Methods* are very popular since they are applicable to any position with any degree of complexity. The first and the crucial task in the Monte Carlo risk estimation process is choosing a suitable model to describe the behavior of return data. The second step is simulating this model and finding a value at the end of each trial. After repeating this process sufficiently, we obtain a simulated distribution and this distribution is assumed to converge to the true but unknown distribution of returns. We can then use this distribution to estimate VaR. When dealing with highly correlated multiple risk factor positions, accompanying MCS with variance reduction techniques, such as Principle Component Analysis, might be a more powerful way to estimate risk. Readers who are interested in variance reduction techniques may consult [GHSh00].

MCS models have also some advantages and disadvantages. The most serious disadvantage is that the estimation process might be too time consuming. One of the advantages of these models is the ability of the models to handle complications like fat-tails, non-linearity, path-dependency, multiple risk factors, etc. Another advantage is that there is plenty of software available for this method and once the appropriate procedures have been set up it is both very easy to use and modify.

In the next section, Coherent Risk Measures, one of the strongest alternatives to VaR, will be introduced.

3.2 Coherent Risk Measures

The deficiencies of VaR which were mentioned in the previous section led many researchers to seek alternative risk measures to quantify financial risk.

In 1997, with the publishing of the article of [ADEH97] finance literature met with the concept of coherent risk measures (CRM). In this context the coherence of a risk measure symbolizes its consistency with economic intuition. After this improvement in the risk measurement area, in 1999, with the article entitled 'Coherent Measures of Risk' [ADEH99], four scholars introduced the entire theory of coherent risk measures for finite probability spaces without complete market assumption to the finance literature. Later on, the theory of CRM was extended to general probability spaces by Delbaen in [D00]. This section includes the review of coherent risk measures both for finite and general probability spaces by taking the studies of the four authors as a main building block.

The theory of coherent risk measures relies on the idea that an appropriate measure of risk is consistent with the finance theory and it inherits the regulator's perspective. Previous studies defined financial risk in terms of changes in the value of a position between two dates but [ADEH99] argues that because risk is related to the variability of the future value of a position, it is better to consider future values only. Thus, risk is a random variable and there is no need to take the initial costs into consideration. They state that for an unacceptable risk two things could be done. The first alternative is changing the position, the other is looking for some commonly accepted instruments when added to the current position makes the future value of the initial position acceptable. Thus, they implicitly define the risk of a portfolio as the amount of the reference instrument such that when we add this amount to our portfolio the position we hold becomes acceptable to a regulator.After this brief introduction let us continue with the assumptions and mathematical background of CRM.

Let Ω be the set of states of nature and assume for now that it is **finite** and all possible states of the world at the end of the period is known. Thus this assumption implies that we know the number of possible events that may occur and there is a finite number of events. However, be careful that the probabilities of the various states may be unknown or not subject to common agreement. Now consider a one period economy starting at time 0 and ending in date T, in which the net-worth of any portfolio is denoted as a random variable $\mathbf{X}$ which takes the value $\mathbf{X}(\omega)$ as the state of nature ω occurs. Also assume that markets at date T are liquid. Let $\mathcal{G}$ represents the set of all risks, that is, the set of all real valued functions on Ω. Since Ω is finite, $\mathcal{G}$ can be identified with $\mathcal{R}^n$, where n=$card(\Omega)$. $\mathcal{L}_+$, denotes the cone of non-negative elements in $\mathcal{G}$ and $\mathcal{L}_-$ denotes its negative.

Definition 3.5. Coherent Risk Measure
A risk measure ρ is coherent if it satisfies the following four axioms:

1. **Translation Invariance:** $\rho(X + \alpha r) = \rho(X) - \alpha$ for all X $\in \mathcal{G}$, $\alpha \in \mathcal{R}$.

2. **Monotonicity:** $\rho(X) \leq \rho(Y)$ if $X \geq Y$ a.s.

3. **Positive Homogeneity:** $\rho(\lambda X) = \lambda \rho(X)$ for $\lambda \geq 0$.

4. **Subadditivity:** $\rho(X + Y) \leq \rho(X) + \rho(Y)$ for all X,Y $\in \mathcal{G}$.

In addition, in the article [ADEH99], Artzner et. al. introduced the following property although it is not a determinant of coherency.

5. **Relevance:** $\rho(X) > 0$ if $X \leq 0$ and $X \neq 0$.

Translation invariance axiom implies that by adding a fixed amount α to initial position and investing it in reference instrument, the risk $\rho(X)$ decreases by α. Also, this condition ensures that the risk measure and returns are in the same unit, namely, currency. The second axiom, *monotonicity* postulates that if $X(\omega) \geq Y(\omega)$ for every state of nature ω, Y is riskier because it has higher risk potential. From the point of view of an investor, the risk assessment of a financial position appears as a numerical representation of preferences. However from the point of view of a regulator, risk measure is viewed as a capital requirement, thus a specific monetary purpose comes into play. Monotonicity property with translation invariance assures that ρ is a monetary measure of risk.

Positive homogeneity axiom means that risk is linearly increasing with the size of the position. This condition may not be satisfied in the real world since markets may not be liquid. Illiquidity of markets implies that increasing the amount of position may create extra risk. However the liquidity assumption stated above ensures the validity of this axiom for this model. The *subadditivity* axiom implies that the risk of a portfolio is always less than the sum of the risks of its subparts. This condition ensures that diversification decreases the risk. Subadditivity together with positive homogeneity implies the *convexity* of the risk measure which is an important property in the portfolio optimization process. *Relevance*, although a necessary, but not sufficient condition for coherency, tells us that a position having zero or negative (at least for some state of nature ω) future net worth has a positive risk. This axiom ensures that the risk measure identifies a random portfolio as risky [JP02].

Definition 3.6. Acceptance set
An acceptance set basically represents the set of acceptable future net worths. ADEH argue that all sensible risk measures should be associated with an acceptance set that satisfies the following conditions:

1. The acceptance set $\mathcal{A}$ contains $\mathcal{L}_+$

2. The acceptance set $\mathcal{A}$ does not intersect the set $\mathcal{L}_{--}$ where

$$\mathcal{L}_{--} = \{X \mid for \;\; each \;\; \omega \in \Omega \;\; , X(\omega) < 0\}$$

2'. The acceptance set $\mathcal{A}$ satisfies $\mathcal{A} \cap \mathcal{L}_- = 0$

3. The acceptance set $\mathcal{A}$ is convex.

4. The acceptance set $\mathcal{A}$ is a positively homogenous cone.

The properties stated above indicate that a reasonable acceptance set accepts any portfolio which always has positive return ($\mathcal{L}_+$), whereas it does not contain a portfolio with sure loss ($\mathcal{L}_{--}$). More stronger axiom 2' states that the intersection of the non positive orthant and acceptance set contains only the origin. Convexity of the acceptance set indicates that linear combination of acceptable portfolios is again acceptable so contained in acceptance set. The last property indicates that an acceptable position can be scaled up or down in size without losing its acceptability.

Definition 3.7. Acceptance set associated to a risk measure
The acceptance set associated with a risk measure ρ is the set $\mathcal{A}_\rho$ defined by

$$\mathcal{A}_\rho = \{X \ \in \ \mathcal{G} : \rho(X) \leq 0\}$$

Definition 3.8. Risk measure associated to an acceptance set
Given the total rate of return r on a reference instrument, the risk measure associated to the acceptance set $\mathcal{A}$ is the mapping from $\mathcal{G}$ to $\mathcal{R}$ denoted by $\rho_{\mathcal{A},r}$ and defined by

$$\rho_{\mathcal{A},r} = \inf\{m \mid m.r + X \ \in \ \mathcal{A}\}$$

Definition 3.8 implies that the risk measure composed by an acceptance set is the minimum amount of capital such that when added to a position, it makes the position acceptable. Notice that if the amount $\rho(X)$ is negative this means the cash amount $-\rho(X)$ can be withdrawn from the position with preserving the acceptability. By the way, risk measurement process pursue the following route: portfolio X is given, the regulator defines the acceptance set $\mathcal{A}$, then the risk measure ρ determines the minimum amount of capital that must be added to X to satisfy the regulator. Thus they define a measure of risk by describing how close or far a position is from acceptability.

In their study [ADEH99], Artzner et. al. stated and proved two propositions to show that a coherent risk measure is consistent with finance theory and the regulator's perspective. Now let us concentrate on these propositions and related proofs.

Proposition 3.3. If the set $\mathcal{B}$ satisfies the four properties given in Definition 3.6, then the risk measure associated to $\mathcal{B}$, $\rho_{\mathcal{B},r}$, is coherent. Moreover the acceptance set induced by a risk measure $\rho_{\mathcal{B}}$ is the closure of $\mathcal{B}$. ($\mathcal{A}_{\rho_{\mathcal{B},r}} = \overline{\mathcal{B}}$)

Proof: Let us first show that $\rho_{\mathcal{B},r}$ is coherent. To show this, one has to prove that $\rho_{\mathcal{B},r}$ satisfies the coherency axioms in Definition (3.5)

1. As a first task, we have to show that $\rho_{\mathcal{B},r}$ is a finite number. Let $\|X\|$ indicates the supremum norm of the random variable X, then,

$-\|X\| \leq X \leq \|X\|$ holds for all X $\in \mathcal{G}$. For $\|X\| + X \geq 0$, we have

$\frac{\|X\|}{r}.r + X \geq 0$. Suppose there exists a scalar m $\in \mathcal{R}$ such that $m > \frac{\|X\|}{r}$, then the strict inequality $m.r + X > 0$ holds. This means that $m.r + X \in \mathcal{L}_+$. Therefore it is contained in $\mathcal{B}$. Also, $m.r + X > 0$ implies that $\rho(X) < m$. Knowing that $m > \frac{\|X\|}{r}$ and $\rho(X) < m$, one can conclude that $\rho(X) \leq \frac{\|X\|}{r}$. Similarly for $X - \|X\| \leq 0$ we have $X - \frac{\|X\|}{r}.r \leq 0$. Now suppose $m < -\frac{\|X\|}{r}$ then, $m.r + X < 0$ and contained in $\mathcal{L}_{--}$. This means $m.r + X \notin \mathcal{B}$, implying that it is not acceptable. Unacceptability of $m.r + X$ demonstrates that $\rho(X) > m$. Knowing this with $m < -\frac{\|X\|}{r}$ indicates that $\rho(X) \geq -\frac{\|X\|}{r}$. Therefore we can conclude that $\rho_{\mathcal{B},r}$ *is a finite number.*

2. Secondly we have to show that $\rho_{\mathcal{B},r}(X)$ satisfies translation invariance. The inequality $\inf\{p \mid X + (\alpha + p).r \in \mathcal{B}\} = \inf\{q \mid X + q.r \in \mathcal{B}\} - \alpha$ proves that $\rho_{\mathcal{B},r}(X + r.\alpha) = \rho(X) - \alpha$, and *translation invariance* axiom is satisfied.

3. The next task is to show that the risk measure satisfies the monotonicity axiom. Assume we have two position X and Y such that $X \leq Y$ and X+m.r $\in \mathcal{B}$ for some m $\in \mathcal{R}$. Let us write $Y + m.r = X + m.r + (Y - X)$ and keep in mind that $Y \geq X$. The first axiom of Definition 3.6 states that an acceptance set contains the nonnegative elements. Thus $Y - X \in \mathcal{B}$. Therefore $Y + m.r = X + m.r + (Y - X) \in \mathcal{B}$. Let $q_x = \{m | mr + X \in \mathcal{B}\}$ and $q_y = \{m | mr + Y \in \mathcal{B}\}$ be two sets. Then,

$$q_x = \{m | mr + X \in \mathcal{B}\} \subset q_y = \{m | mr + Y \in \mathcal{B}\}$$

holds since position Y is closer to the acceptance set than X. Therefore

$$\inf q_x = \rho_{\mathcal{B},r}(X) \ \geq \inf q_y = \rho_{\mathcal{B},r}(Y)$$

holds. Thus the *monotonicity* axiom has been proved.

4. Let m be any number such that m $\in \mathcal{R}$. If $m > \rho_{\mathcal{B},r}(X)$, then for each $\lambda > 0$ we have $\lambda.X + \lambda.m.r \in \mathcal{B}$. This is because of the fact that $m > \rho_{\mathcal{B},r}(X)$ so $X + m.r \in \mathcal{B}$. Besides, $\mathcal{B}$ is a convex set and a positively homogenous cone so it contains the position $\lambda.X + \lambda.m.r$. Since $\lambda.X + \lambda.m.r \in \mathcal{B}$, by using the Definition 3.7 we can conclude that $\rho_{\mathcal{B},r}(\lambda.X) \leq \lambda.m$. On the other hand, if $m < \rho_{\mathcal{B},r}(X)$, then for each $\lambda > 0$ we have $\lambda.X+\lambda.m.r \notin \mathcal{B}$, and this proves that $\rho_{\mathcal{B},r}(\lambda.X) \geq \lambda.m$. Therefore, we conclude with $\rho_{\mathcal{B},r}(\lambda.X) = \lambda.\rho_{\mathcal{B},r}(X)$ implying the *positive homogeneity* of the risk measure $\rho_{\mathcal{B},r}$.

5. The last axiom that should be satisfied is subadditivity. Let X and Y be two positions, m, r two real numbers and $\alpha \in [0, 1]$. Then, the following arguments are true due to the fact that $\mathcal{B}$ is a positively homogenous cone.

$$If \ \ X + m.r \in \mathcal{B} \longrightarrow \alpha X + \alpha.m.r \in \mathcal{B}$$

$$If \ \ Y + n.r \in \mathcal{B} \longrightarrow (1-\alpha)Y + (1-\alpha)m.r \in (B)$$

Besides, the convexity of $\mathcal{B}$ ensures that

$$\alpha.X + (1-\alpha).Y + \alpha.m.r + (1-\alpha).m.r \in \mathcal{B}$$

Now take $\alpha = \frac{1}{2}$, then $\frac{1}{2}[(X + Y) + (m + n).r] \in \mathcal{B}$ then by the positive homogeneity property of $\mathcal{B}$, $[(X + Y) + (m + n).r] \in \mathcal{B}$. Now define $\{s|(X + Y) + s.r \in \mathcal{B}\}$ which indicates the set of numbers that make the position (X+Y) acceptable when added to it. It is obvious that $\{s|(X + Y) + s.r \in \mathcal{B}\} \supset \{m|X + m.r \in \mathcal{B}\} + \{n|Y + n.r \in \mathcal{B}\}$. When we take infimum on the sets above we can conclude that $\rho_{\mathcal{B},r}(X+Y) \leq \rho_{\mathcal{B},r}(X) + \rho_{\mathcal{B},r}(Y)$ which proves the *subadditivity* property for the risk measure.

6. For each $X \in \mathcal{B}$, $\rho_{\mathcal{B},r}(X) \leq 0$ hence X $\in \mathcal{A}_{\rho_{\mathcal{B},r}}$. Proposition 3.4 and points

(1) through (5) ensures that $\mathcal{A}_{\rho_{\mathcal{B},r}}$ is closed, which proves $\mathcal{A}_{\rho_{\mathcal{B},r}} = \overline{\mathcal{B}}$ □

Proposition 3.4. If a risk measure ρ is coherent, then the acceptance set $\mathcal{A}_\rho$ is closed and satisfies the axioms of Definition 3.6 above. Moreover $\rho = \rho_{\mathcal{A}_\rho,r}$

Proof: Let us show one by one that the acceptance set of a coherent risk measure satisfies the axioms of Definition 3.6.

1. Subadditivity and positive homogeneity ensures that ρ is a convex function on $\mathcal{G}$, and continuous, since for $\alpha \in [0,1]$ we have $\rho(\alpha.x+(1-\alpha).y) \leq \alpha.\rho(X)+(1-\alpha).\rho(Y)$. This implies that $\mathcal{A}_\rho = \{x \mid \rho(X) \leq 0\}$ is closed. Moreover if X,Y $\in \mathcal{A}_\rho$ i.e $\rho(X) \leq 0$ and $\rho(Y) \leq 0$, then $\rho(\alpha.X + (1-\alpha).Y) \leq 0$. Thus $\mathcal{A}_\rho$ is convex. If $\rho(X) \leq 0$ and $\lambda > 0$, then $\rho(\lambda.X) = \lambda.\rho(X) \leq 0$. Therefore if X $\in \mathcal{A}_\rho$, then $\lambda.X \in \mathcal{A}_\rho$, implying positive homogeneity property of $\mathcal{A}_\rho$. Consequently $\mathcal{A}_\rho$ *is a closed, convex and positively homogenous cone.*

2. Positive homogeneity and continuity of ρ implies that $\rho(0) = 0$, because:
if $\lambda_n > 0 \downarrow 0$ as $n \to \infty$, then $\lambda X_n \to 0$, thus $\rho(\lim_{n\to\infty} \lambda_n X) = \lim_{n\to\infty} \lambda_X \rho(x)$
By monotonicity, if $X \geq 0$, then $\rho(X) \leq \rho(0) = 0$. Therefore $\mathcal{L}_+$ *is contained in* $\mathcal{A}_\rho$.

3. Let X be in $\mathcal{L}_{--}$ with $\rho(X) < 0$. Monotonicity of ρ implies that $\rho(0) < 0$, this is a contradiction. If $\rho(X) = 0$, then we find $\alpha > 0$ such that $X + \alpha.r \in \mathcal{L}_{--}$. Translation invariance of ρ provides $\rho(X+\alpha.r) = \rho(X)-\alpha < 0$, a contradiction. Hence $\rho(X) > 0$ that is X$\notin \mathcal{A}_\rho$. This proves that $\mathcal{A}_\rho$ *does not contain* $\mathcal{L}_{--}$.

4. To prove the second part of the proposition, let δ be any number with $\rho_{\mathcal{A}_\rho,r}(X) < \delta$ for each X. Then X+δ.r$\in \mathcal{A}_\rho$. Hence $\rho(X+\delta.r) \leq 0$. By translation invariance $\rho(X) - \delta \leq 0$, then $\rho(X) \leq \delta$. Therefore $\rho(X) \leq \rho_{\mathcal{A}_\rho,r}(X)$, hence $\rho \leq \rho_{\mathcal{A}_\rho,r}$.

5. Now for each X, let $\delta > \rho(X)$, then $\rho(X + \delta.r) < 0$ and $X + \delta.r \in \mathcal{A}_\rho$, hence $\rho_{\mathcal{A}_\rho,r} \leq 0$. By translation invariance $\rho_{\mathcal{A}_\rho,r} \leq \delta$ and so $\rho_{\mathcal{A}_\rho,r}(X) \leq \rho(X)$, that is $\rho_{\mathcal{A}_\rho,r} \leq \rho$. This proves that $\rho_{\mathcal{A}_\rho,r} = \rho$. □

These two propositions prove that there is a one to one correspondence between the coherent risk measures and regulator's manner on risk perception. In addition to these two fundamental propositions there is a complementary one, which shows the relation between relevance axiom and the axiom 2' of the Definition 3.6

Proposition 3.5. If a set $\mathcal{B}$ satisfies the axioms 1,2',3 and 4 in Definition 3.6, then the coherent risk measure $\rho_{\mathcal{B},r}$ satisfies the *relevance* axiom. If a coherent risk measure ρ satisfies the relevance axiom, then the acceptance set $\mathcal{A}_{\rho_{\mathcal{B},r}}$ satisfies the axiom 2' of Definition 3.6

Proof: **1.** If $X \leq 0$ and $X \neq 0$, then we know that $X \in \mathcal{L}_-$ and since $X \neq 0$, by axiom 2', $X \notin \mathcal{B}$, which means $\rho_{\mathcal{B},r}(X) > 0$.
2. For X $\in \mathcal{L}_-$ and $X \neq 0$ relevance axiom provides $\rho(X) > 0$ and $X \notin \mathcal{B}$ □

After these propositions, let us continue with the representation theorem for the most general coherent risk measure.

Proposition 3.6. Given the total return r on a reference investment, a risk measure ρ is coherent if and only if there exists a family $\mathcal{P}$ of probability measures on the set of states of nature, such that

$$\rho(X) = \sup\{E_P[-X/r] \mid P \in \mathcal{P}\}$$

Proof: **i)** The if part of the proof is obvious and basically depends on the properties of supremum function.
1. Let $\alpha \in \mathcal{R}$, then

$$\begin{aligned}\rho(X + \alpha.r) &= \sup\{E_P[-\frac{(X + \alpha.r)}{r}] \ \mid P \in \mathcal{P}\} \\ &= \sup\{E_P[-\frac{X}{r} - \alpha] \mid P \in \mathcal{P}\} \\ &= \sup\{E_P[-\frac{X}{r}] \mid P \in \mathcal{P}\} - \alpha \\ &= \rho(X) - \alpha\end{aligned}$$

Which proves that ρ has the *translation invariance* property.
2. Let X,Y $\in \mathcal{G}$ such that $X \geq Y$, then $\sup\{E_P[-\frac{X}{r}] \mid P \in \mathcal{P}\} \leq sup\{E_P[-\frac{Y}{r}] \mid P \in \mathcal{P}\}$. This implies *monotonicity*.

3. Let λ be any number such that $\lambda > 0$, then

$$\begin{aligned}\rho(\lambda.X) &= \sup\{E_P[-\frac{(\lambda.X)}{r}] \ \mid P \in \mathcal{P}\} \\ &= \sup\{\lambda E_P[-\frac{X}{r}] \mid P \in \mathcal{P}\} \\ &= \lambda \sup\{E_P[-\frac{X}{r}] \mid P \in \mathcal{P}\} \\ &= \lambda\rho(X)\end{aligned}$$

Therefore ρ is *positively homogenous.*

4. Let X,Y $\in \mathcal{G}$, then $\rho(X+Y) = \sup\{E_P[-\frac{(X+Y)}{r}] \mid P \in \mathcal{P}\}$, $\rho(X) + \rho(Y) = \sup\{E_P[-\frac{X}{r}] \mid P \in \mathcal{P}\} + \sup\{E_P[-\frac{Y}{r}] \mid P \in \mathcal{P}\}$. And it is obvious from the equations above that $\rho(X+Y) \leq \rho(X) + \rho(Y)$. Thus, the *subadditivity* of ρ follows.

ii) Now suppose ρ is coherent, let M be the set of all probability measures on Ω and define

$$\mathcal{P}_\rho = \{P \in M \mid \forall\ X \in \mathcal{G}, E_P[-\frac{X}{r}] \leq \rho(X)\}$$

We have to show that

$$\rho(X) = \sup_{P \in \mathcal{P}_\rho} \{E_P[-\frac{X}{r}]\}$$

1. The set of probabilities M is a compact set in $\mathcal{R}^n$ where n=$card(\Omega)$. In fact

$$M = \{P \in \mathcal{R}^n \mid \forall\ \omega, P(\omega) \geq 0, \sum_\omega P(\omega) = 1\}$$

This is a closed subset of the unit ball in $\mathcal{R}^n$ which is compact. Therefore M is compact.

2. Given X$\in \mathcal{R}$, $E_P[-\frac{X}{r}]$ is a continuous mapping from M in to $\mathcal{R}$. We know that the image of a continuous mapping of a compact set is again compact. Thus $\{E_P(-\frac{X}{r} \mid P \in M)$ is compact in $\mathcal{R}$. This implies that $\{E[-X] \mid P \in M \cap \{a \in \mathcal{R} : a \leq \rho(X)\}\}$ is compact because the closed subset of a compact metric space is again compact. Therefore $\rho(X) = sup\{E_P[-X] \mid P \in \mathcal{P}_\rho\}$. This concludes the proof. □

To sum up, according to ADEH the risk value $\rho(X)$ of a future net worth X can be determined coherently by applying the following steps:

- Compute, under each test probability $P \in \mathcal{P}$, the average of the future net worth X of the position, in formula $E_P[-\frac{X}{r}]$,
- Take the maximum number of all numbers found above, which correspond to the formula $\rho(X) = \sup\{E_P[-\frac{X}{r}] \mid P \in \mathcal{P}\}$.

Being inspired from the study [ADEHKu02] of Artzner et. al. this representation result might also be interpreted as follows: For any acceptance set, there exists a set $\mathcal{P}$ of probability distributions, called generalized scenarios (simply speaking, a set of loss values and their associated probabilities) or test probabilities, on the space Ω of states of nature, such that a given position with future random value denoted by X, is acceptable if and only if:

> For each test probability $P \in \mathcal{P}$, the expected value of the future net worth under P, i.e $E_P[-\frac{X}{r}]$ is non-positive.

Up to know we assumed that Ω, the set of states of nature, is finite. We worked on the set of risks $\mathcal{G}$ which is identified with $\mathcal{R}^\Omega$. In fact, by doing so we restricted the number of values that our position may take in the date T. In other words we restricted the number of scenarios that may become true with the boundaries of our mind, our experiences. However we should admit that in such a place as the world we live in, nothing is impossible. To take this fact into account, in his study [D00], Delbaen extended the definition of coherent risk measures to general probability spaces by removing the finiteness assumption of Ω and show how to define such a measure on the space of all random variables. The following part of this section will refer to this fundamental article of Delbaen. Before starting the underlying theory of this approach it is appropriate to give some information on the mathematical background and notation.

Given a probability space $(\Omega, \mathcal{F}, P)$, the σ -algebra $\mathcal{F}$ describes all the events that become known at the end of the observation period. The measure P describes in what probability events might occur. In finance such probabilities are not objective. Thus may differ among agents, institutions, etc. However it is argued that the class of negligible sets and consequently the class of probability measures that are equivalent to P remain the same. This can be expressed by stating that only the knowledge of the events of probability zero is important. So we need an agreement on the possibility that events might occur, not on the actual value of the probability [D00].

It might be useful to refresh our minds and remember the definitions of some special $\mathcal{L}^p$ spaces that will be used in this part (For detailed explanation see, Preliminaries Def: 2.18). With $\mathcal{L}^\infty$ we mean the space of all equivalence classes of *bounded real valued* random variables. The space $\mathcal{L}^0$ denotes the space of all equivalence classes of *real valued* random variables and $\mathcal{L}^1$ is the space of all equivalence classes of *integrable* random variables. Since we only have the knowledge of the events of probability zero, to define a risk measure we need spaces of random variables such that they remain the same when we change the underlying probability to an equivalent one. $\mathcal{L}^0$ and $\mathcal{L}^\infty$are two natural spaces of this kind. Therefore mainly these two will be considered throughout this subsection. Other notations that will be used in this part are as follows: $(\mathcal{L}^1)'$ represents the dual space (see, Preliminaries Def: 2.24) of $\mathcal{L}^1$ and is equal to $\mathcal{L}^\infty$. $(\mathcal{L}^\infty)'$ indicates the dual of $\mathcal{L}^\infty$ which is equal to the Banach space $\mathbf{ba}(\Omega, \mathcal{F}, P)$ of all bounded finitely additive measures (see, Preliminaries Def:2.6) μ (S) on $(\Omega, \mathcal{F})$ with the property that $P(A) = 0 \rightarrow \mu(A) = 0$. Finally as a remark, a positive element $\mu \in ba(P)$ such that $\mu(1) = 1$ is called a *finitely additive* probability measure.

In his study, Delbaen firstly extended the set $\mathcal{G}$ to $\mathcal{L}^\infty(\Omega, \mathcal{F}, P)$ with respect to a fixed probability measure P on a measurable space $(\Omega, \mathcal{F})$, then to allow infinitely high risks, which means something like a risk that cannot be insured, he defined the risk measure on $\mathcal{L}^0$ since this space consists of all equivalence classes of both bounded and unbounded real valued random variables. Finally, in this part, the rate of return for the reference instrument is assumed to be zero. Now we are ready to review the theory of coherent risk measures on general probability spaces.

Definition 3.9. A mapping $\rho : \mathcal{L}^\infty(\Omega, \mathcal{F}, P) \to \mathcal{R}$ is called a coherent risk measure if the following properties hold.

1. If $X \geq$ then $\rho(X) \leq 0$.

2. Subadditivity: $\rho(X_1 + X_2) \leq \rho(X_1) + \rho(X_2)$.

3. Positive homogeneity: for $\lambda \geq 0$ we have $\rho(\lambda X) = \lambda \rho(X)$.

4. For every constant function we have that $\rho(\alpha + X) = \rho(X) - \alpha$.

The interpretation of these axioms are similar to ones stated in the previous part. Here the only difference is that this model assumes no interest rate.

Theorem 3.1. Suppose that $\rho : \mathcal{L}^\infty(P) \to \mathcal{R}$ is a coherent risk measure. Then there is a convex $\sigma(\mathbf{ba}(P), \mathcal{L}^\infty)$-closed set $\mathcal{P}_{\mathbf{ba}}$ of finitely additive probabilities, such that

$$\rho(X) = \sup_{\mu \in \mathcal{P}_{\mathbf{ba}}} E_\mu[-X]$$

Proof: The set $C = \{X \mid \rho(X) \leq 0\}$ is clearly a convex and norm closed positively homogenous cone in the space $\mathcal{L}^\infty(P)$. The polar set $C^o = \{\mu \mid \forall X \in C : E_\mu[X] \geq 0\}$ is also a convex cone, closed for the weak* topology on **ba**(P). All elements in C^o are positive since $\mathcal{L}^\infty_+ \subset C$. This implies that for the set $\mathcal{P}_{\mathbf{ba}}$, defined as $\mathcal{P}_{\mathbf{ba}} = \{\mu \mid \mu \in C \text{ and } \mu(1) = 1\}$, we have $C^o = \cup_{\lambda \geq 0} \lambda \mathcal{P}_{\mathbf{ba}}$. The duality theory, more precisely, the bipolar theorem (see, Preliminaries Thm: 2.11), then implies that $C = \{X \mid \forall \mu \in \mathcal{P}_{\mathbf{ba}} : E_\mu[X] \geq 0\}$. This means that

$\rho(X) \leq 0$ if and only if $E_\mu[X] \geq 0$ for all $\mu \in \mathcal{P}_{\mathbf{ba}}$. Since $\rho(X - \rho(X)) = 0$ we have that $X + \rho(X) \in C$ and for all μ in $\mathcal{P}_{\mathbf{ba}}$ we find that $E_\mu[X + \rho(X)] \geq 0$. This can be reformulated as

$$\sup_{\mu \in \mathcal{P}_{\mathbf{ba}}} E_\mu[-X] \leq \rho(X)$$

Since for arbitrary $\varepsilon > 0$, we have $\rho(X + \rho(X) - \varepsilon > 0)$, we obtain $X + \rho(X) - \varepsilon \notin C$. Therefore there is a $\mu \in \mathcal{P}_{\mathbf{ba}}$ such that $E_\mu[X + \rho(X) - \varepsilon] < 0$ which leads to the opposite inequality and hence to:

$$\rho(X) = \sup_{\mu \in \mathcal{P}_{\mathbf{ba}}} E_\mu[-X]$$

□

The previous theorem shows that coherent risk measures can be represented in terms of finitely additive probabilities. The financial importance of this representation is that any coherent risk measure can be obtained as the supremum of the expected loss $E_\mu[-X]$ over a set $\mathcal{P}$ of generalized scenarios [FrR02]. Besides, from the proof of the previous theorem it is seen that there is a one to one correspondence between coherent risk measure ρ, weak* closed convex sets of finitely additive probability measures $\mathcal{P}_{\mathbf{ba}} \subset \mathbf{ba}(P)$ and $\|.\|_\infty$ closed convex cones $C \subset \mathcal{L}^\infty$ such that $\mathcal{L}^\infty_+ \subset C$. The relation between C and ρ is given by

$$\rho(X) = \inf\{\alpha \mid X + \alpha \in C\}$$

Here, the set C represents the *set of acceptable positions.*

If μ is a purely finitely additive measure, the expression $\rho(X) = E_\mu[-X]$ gives a coherent risk measure. This functional that is represented by a finitely additive measure cannot be directly described by a σ-additive probability measure. This requires an additional hypothesis. After the characterization of a risk measure in terms of finitely additive probabilities Delbaen gave the representation of coherent risk measures in terms of **σ-additive** probabilities by adding a continuity property (called the Fatou property). This property is equivalent to the hypothesis that the acceptable positions $\{X \in \mathcal{L}^\infty : \rho(X) \leq 0\}$ is $\sigma(\mathcal{L}^\infty, \mathcal{L}^1)$-closed.

Definition 3.10. A coherent risk measure $\rho : \mathcal{L}^\infty \to \mathcal{R}$ is said to satisfy the Fatou property if $\rho(X) \leq \liminf \rho(X_n)$, for any sequence, $(X_n)_{n\geq 1} \subset \mathcal{L}^\infty$, uniformly bounded by 1 and converging to X in probability.

Theorem 3.2. For a coherent risk measure ρ on $\mathcal{L}^\infty$, the following properties are equivalent:

1. There is an L_1-closed, convex set of probability measures $\mathcal{P}_\sigma$, all being absolutely continuous with respect to P and such that for X$\in \mathcal{L}^\infty$:

$$\rho(X) = \sup_{Q \in \mathcal{P}_\sigma} E_Q(-X)$$

2. The convex cone $C = \{X \mid \rho(X) \leq 0\}$ is weak*, i.e $\sigma(\mathcal{L}^\infty(P), \mathcal{L}^1(P))$ closed.

3. ρ satisfies the Fatou property.

4. If X_n is a uniformly bounded sequence that decreases to X a.s. then $\rho(X_n)$ tends to $\rho(X)$.

For the proof of the theorem, see [D00].

After characterizing the risk measure ρ in terms of σ-additive probabilities, the next task is to extend the domain of a coherent risk measure defined on $\mathcal{L}^\infty$ to Ł^o of all equivalence classes of measurable functions. There is a well known negative result that if the space $(\Omega, \mathcal{F}, P)$ is atomless, then there is no real-valued coherent risk measure ρ on $\mathcal{L}^o$ which means that there is no mapping $\rho : \mathcal{L}^o \to \mathcal{R}$ such that the coherency properties hold. For the proof see [D00, p:16]. This is a problem and the solution given is to extend the risk measure in such a way that it can take the value $+\infty$ but it cannot take the value $-\infty$. The former $(+\infty)$ means that the risk is so high that no matter what amount of capital is added, the position will remain unacceptable. The latter implies that the position is so safe that an arbitrary amount of capital could be withdrawn without endangering the company. Since regulators and risk managers are conservative it is normal to exclude the latter situation. Therefore it is economically reasonable to extend the range of the coherent risk measures.

Definition 3.11. A mapping $\rho : \mathcal{L}^o \rightarrow \mathcal{R} \cup \{+\infty\}$ is called a coherent risk measure if

1. If $X \geq$ then $\rho(X) \leq 0$.

2. Subadditivity: $\rho(X_1 + X_2) \leq \rho(X_1) + \rho(X_2)$.

3. Positive homogeneity: for $\lambda \geq 0$ we have $\rho(\lambda X) = \lambda\rho(X)$.

4. For every constant function we have $\rho(\alpha + X) = \rho(X) - \alpha$.

For the construction of a coherent risk measure, the first idea could be to define it as

$$\sup_{Q \in \mathcal{P}_\sigma} E_Q[-X]$$

but this does not work because the random variable X may not be integrable with respect to $Q \in \mathcal{P}_\sigma$ due to the fact that X is defined on $\mathcal{L}^o$. We can try the following to ensure that X is integrable for the measure $Q \in \mathcal{P}_\sigma$:

$$\sup\{E_Q[-X] \mid Q \in \mathcal{P}_\sigma; X \in \mathcal{L}^1(Q)\}$$

In the definition above, the set over which supremum is taken depends on the random variable X. This poses problems when we try to compare risk measures of different positions in a portfolio optimization process. Therefore we need another definition.

Definition 3.12. For a given, closed convex set, $\mathcal{P}_\sigma$, of probability measures, all absolutely continuous with respect to P and $n \geq 0$ we define ρ as:

$$\rho(X) = \lim_{n\to+\infty} \sup_{Q\in\mathcal{P}_\sigma} E[-(X \wedge n)]$$

In the Definition 3.12 the idea is to truncate the random variable X from above by n. This means that firstly the possible future wealth up to level n is taken into account. Then by using the supremum of all expected values the risk measure is calculated. Finally n is let to go to infinity. In this process, high future values play a role, but their effect only enters through a limit procedure. By doing so a conservative viewpoint is followed. Notice that in Definition 3.12 we have to ensure that $\rho(X > -)\infty$ for all $X \in \mathcal{L}^o$. This is achieved by the following theorem.

Theorem 3.3. The following properties are equivalent:

1. For each X $\in \mathcal{L}^o$ we have $\rho(X) > -\infty$.

2. Let $\phi(X) = -\rho(X)$, for each $f \in \mathcal{L}^o_+$ we have

$$\phi(X) = \lim_{n} \inf_{Q\in\mathcal{P}_\sigma} E_Q[f \wedge n] < +\infty$$

3. There is a $\gamma > 0$ such that for each A with $P[A] \leq \gamma$ we have

$$\inf_{Q\in\mathcal{P}_\sigma} Q[A] = 0$$

Proposition 3.7. The hypothesis of Theorem 3.3 is satisfied if for each non-negative function $f \in \mathcal{L}^o$, there is $Q \in \mathcal{P}_\sigma$ such that $E_Q(f) < \infty$

Theorem 3.4. The properties in Theorem 3.3 are also equivalent with:

4. For every $f \in \mathcal{L}^o_+$ there is $Q \in \mathcal{P}_\sigma$ such that $E_Q[f] < \infty$.

5. There is a $\delta > 0$ such that for every set A with $P[A] < \delta$, we can find an element $Q \in \mathcal{P}_\sigma$ such that $Q[A] = 0$.

6. There is a $\delta > 0$, as well as a number $\mathcal{K}$ such that for every set A with $P[A] < \delta$, we can find an element $Q \in \mathcal{P}_\sigma$ such that $Q[A] = 0$ and $\|\frac{dQ}{dP}\|_\infty \leq \mathcal{K}$

For the proof of the previous two theorems see [D00, p:18]

After talking about the theory of coherent risk measures both for finite and general probability spaces, we will observe risk measures globally to see the advantages and disadvantages of each. The first advantage of this system is that the way it defines risk coincides with the regulator's perspective since it is a monetary measure of risk and satisfies subadditivity.

A coherent risk measure defines the risk of a portfolio as a cash requirement that makes the position acceptable to a regulator. Thus it implicitly assumes that the firm has already made its capital budgeting decision and does not allow the portfolio composition to change. This situation may be seen as a problem when we look at the risk assessment problem from a firm's perspective. Besides, allowing only the usage of risk-free assets to make a position acceptable may oblige a firm to hold too much regulatory capital, because adding capital may not be the efficient way of reducing risk. Taking this fact into account, Jarrow et. al. represents the generalized coherent risk measures in [JP02] to make coherent risk measures more consistent with a firm's perspective. In the next section we will focus on this measure in detail.

ADEH made the following interesting remark in their study [ADEH99]

Remark 3.3. Model risk can be taken into account by including into the set $\mathcal{P}$ a family of distributions for the future prices, possibly arising from other models.

The remark above indicates that when we use coherent risk measures we are able to quantify the model risk in the risk assessment process. This is another advantage of coherent risk measures. However, as stated before, coherent risk measures fail to consider the liquidity risk arising from the fact that markets may not be liquid on the date T. In [FS02a], Föllmer et. al. introduce the Convex Risk Measures as a solution to this problem. In this context we will also consider convex measures of risk in one of the following sections.

Generally in financial markets there is a tendency to perceive the coherence of a risk measures not as a necessity but as an optional property. This is because coherent risk measures do not seem easy to apply as VaR. This may be seen as the disadvantage of coherent risk measures when compared with VaR. However, comparing VaR with coherent risk measures does not make sense due to the fact that VaR has failures that can not be accepted by finance theory.

3.3 Generalized Coherent Risk Measures

Risk management display differences when we look at the problem from different perspectives. From a regulator's perspective, the problem is to determine the amount of capital the firm must add to its initial position to make it acceptable. In this process, the regulator takes the firm's portfolio composition as fixed, thus implicitly assumes that the firm has already made its decision on capital budgeting. However, from a firm's perspective, the goal in risk management process is to choose the portfolio composition so as to maximize its return, subject to any regulatory capital requirement. Coherent risk measures of ADEH approach the problem from the regulator's perspective. To bring firm's perspective in to the discussion of coherent risk measures, in [JP02], Jarrow et. al. introduce Generalized Coherent Risk Measures as the 'minimum quantity invested in any marketable security such that the original portfolio, along with the modified security, becomes acceptable'. This section will mainly refer to this article.

Suppose there is a single time period. The future net worth of any portfolio is denoted as the random variable X on the space Ω. The space Ω is not necessarily finite. In the case where Ω is infinite there is also a given sigma algebra $\mathcal{F}$ and a probability measure P defined on $\mathcal{F}$. Let $\mathcal{G}$ be the space of all random variables. $\mathcal{G}$ is identified with $\mathcal{R}^n$, where $n = card(\mathcal{G})$. A risk function ρ maps the random variables, $\mathcal{G}$, into the *non-negative* real line. There introduced a norm$\| \, . \, \|$ on the space of random variables $\mathcal{G}$ to provide a tool for characterizing two similar portfolios. Such a norm is induced in an economy by investors' preferences.

Definition 3.13. Generalized Coherent Risk Measures
A generalized coherent risk measure satisfies the following five axioms:

1. **Subadditivity:** $\rho(X_1+X_2) \leq \rho(X_1)+\rho(X_2)$ for all $X_1, X_2 \in \mathcal{G}$.

2. **Monotonicity:** $\rho(X_1) \leq \rho(X_2)$ if $X_1 \geq X_2$.

3. **Positive Homogeneity:** $\rho(\lambda X) = \lambda\rho(X)$ for $\lambda \geq 0$.

4. **Relevance:** $\rho(X) > 0$ iff $x \leq 0$ and $X \neq 0$.

5. **Shortest Path:** If $\mathcal{H}_\rho$ is defined as

$$\mathcal{H}_\rho \equiv \{X \in \mathcal{G} : \rho(X) = 0\}.$$

For each $X \in \mathcal{G}$ there exists a portfolio $X^* \in \mathcal{H}_\rho$ such that

a) it is the point of shortest distance,$\| X - X^* \|$, from X to the set $\mathcal{H}_\rho$, and

b) for any scaler $0 \leq \lambda \leq \rho(X)$,

$$\rho(X + \lambda.u) = \rho(X) - \lambda$$

where u is the unit vector in the direction $(X^* - X)$, i.e.,

$$u = \frac{X^* - X}{\| X^* - X \|}$$

The properties (1-4) and their interpretations are the same as those in the definition of coherent risk measures of ADEH. The key difference is the shortest path axiom. The shortest path may be interpreted as the efficient way (investment) of reducing risk. SP axiom implies that the risk of a position is the distance $\| X - X^* \|$, which is the minimum re-balancing of the original portfolio X needed such that the resulting portfolio X^* belongs to the acceptance set. This SP axiom characterizes the difference between the firm's and the regulator's perspective. It induces the firm's perspective since the asset added to the portfolio (u) need not be capital, it may be a derivative instrument or an insurance contract instead. Note that if $u = r$ in the shortest path axiom, then SP axiom turns out to be the translation invariance axiom and the generalized coherent risk measure reduces to the coherent risk measure. Also, an immediate consequence of this axiom is that, by adding an amount equal to $\rho(X)$ along the shortest path, the risk of the portfolio reduces to zero, i.e. $\rho(X + \rho(X)u) = \rho(X) - \rho(X) = 0$.

Remark 3.4. For any portfolio with $X \geq 0$, we have $\rho(X) = 0$. This is because of the fact that, for an acceptable position we do not need any re-balancing on the initial portfolio. Therefore the shortest distance is equal to zero. Mathematically speaking, by using positive homogeneity and monotonicity we have;

$$2\rho(X) = \rho(2X) \leq \rho(X) \rightarrow \rho(X) \leq 0.$$

Since we defined $\rho(X)$ as the minimum distance, it cannot take negative values. Thus $\rho(X) = 0$ follows.

Definition 3.14. Acceptable Set
The set of acceptable portfolio holdings $\mathcal{A} \subset \mathcal{G}$ has the following three properties:

1. $X \in \mathcal{A}$ if and only if $X \geq 0$.

2. $\mathcal{A}$ is a convex set.

3. $\mathcal{A}$ is closed under multiplication by $\lambda \geq 0$ (i.e. it is a positively homogenous cone)

Definition 3.15. Induced Acceptance Set
The acceptance set induced by ρ equals

$$\mathcal{A}_\rho = \{X \in \mathcal{G} : \rho(X) = 0\}.$$

Notice that although the acceptance set axioms are very similar to those used by ADEH, the definition of the acceptance set induced by a generalized coherent risk measure ρ is different from the one induced by a coherent risk measure. This difference arises due to the fact that the two measures are constructed from different starting points. One can understand the reason of this difference by considering Remark 3.4 again.

Definition 3.16. Induced Risk Function
The risk function induced by the acceptance set $\mathcal{A}$ equals

$$\rho_{\mathcal{A}}(X) = \inf\{\| X - X' \| : \; X' \in \mathcal{A}\}.$$

The following two propositions relate the properties of a risk measure to the corresponding acceptance set, and vice versa.

Proposition 3.8. If acceptance set $\mathcal{B}$ satisfies the axioms in Definition 3.14, then the induced risk measure $\rho_{\mathcal{B}}$ in Definition 3.16 is a generalized coherent risk measure. Moreover $\mathcal{A}_{\rho_{\mathcal{B}}} = \mathcal{B}$

Proof: First assume that acceptance set $\mathcal{B}$ satisfies the axioms in Definition 3.14.

Let X and $Y \in \mathcal{G}$ and X^*, Y^* respectively be the closest portfolio on the acceptance set $\mathcal{B}$ under the given norm, i.e., $X^* = arg\inf\{\| X - X' \|: X' \in \mathcal{B}\}$ and $Y^* = arg\inf\{\| Y - Y' \|: Y' \in \mathcal{B}\}$. By definition

$$\rho_{\mathcal{B}}(X) = \| X - X^* \|$$

$$\rho_{\mathcal{B}}(Y) = \| Y - Y^* \|$$

Now consider the portfolio $(X + Y)$ and let $(X + Y)^*$ be the closest acceptable portfolio for $(X + Y)$. Notice that $X^* + Y^*$ belongs to the acceptance set $\mathcal{B}$ because set $\mathcal{B}$ is convex and positively homogenous. Since $(X+Y)^*$ is an element of $\mathcal{B}$ and it is the optimal one we have

$$\| X + Y - (X + Y)^* \| \leq \| X + Y - (X^* + Y^*) \|,$$

By definition of $\rho_{\mathcal{B}}$ and triangle inequality

$$\rho_{\mathcal{B}}(X + Y) \leq \| X - X^* + Y - Y^* \| \leq \| X - X^* \| + \| Y - Y^* \| = \rho_{\mathcal{B}}(X) + \rho_{\mathcal{B}}(Y)$$

This proves the *subadditivity* for $\rho_{\mathcal{B}}$.

To show the *monotonicity*, suppose $X \leq Y$ and consider a random variable $Y - X$. Since $Y - X \geq 0$ we have $Y - X \in \mathcal{B}$. Since $Y = X + Y - X$, subadditivity of $\rho_{\mathcal{B}}$ implies $\rho_{\mathcal{B}}(Y) \leq \rho_{\mathcal{B}}(X) + \rho_{\mathcal{B}}(Y - X)$. We know that $\rho(Y - X) = 0$ since it is contained in $\mathcal{B}$. By using this information $\rho_{\mathcal{B}}(Y) \leq \rho_{\mathcal{B}}(X)$ follows.

For the *positive homogeneity*, consider a risk X and a scalar $\lambda \geq 0$. Then

$$\lambda \rho_{\mathcal{B}}(X) = \lambda \| X - X^* \| = \| \lambda X - \lambda X^* \| \geq \| \lambda X - (\lambda X)^* \| = \rho_{\mathcal{B}}(\lambda X)$$

since λX^* is in the acceptance set (remember that the acceptance set is a homogenous cone) but need not be the optimal one in terms of minimizing the distance of portfolio λX to the acceptance set. Now to indicate the reverse direction, define $\rho_{\mathcal{B}}(\lambda X) = \parallel \lambda X - (\lambda X)^* \parallel = \lambda \parallel X - \frac{(\lambda X)^*}{\lambda} \parallel \geq \lambda \rho_{\mathcal{B}}(X)$ since $\frac{1}{\lambda}(\lambda X)^*$ is an element of $\mathcal{B}$ but need not be the optimal one. Thus the equality $\rho_{\mathcal{B}}(\lambda X) = \lambda \rho_{\mathcal{B}}(X)$ is proved.

Now consider $X \in \mathcal{G}$ such that $X \ngeq 0$. To prove the *relevance* property it must be shown that $\rho_{\mathcal{B}}(X) > 0$. Contrarily suppose that $\rho_{\mathcal{B}}(X) = 0$ implying that $X \in \mathcal{B}$. The first axiom of $\mathcal{B}$ guarantees that $X \geq 0$ for any $X \in \mathcal{B}$,a contradiction. Thus $X \ngeq 0$.

Set $\mathcal{H}_{\rho_{\mathcal{B}}}$ is defined as the set of random variables such that $\rho_{\mathcal{B}}(X) = 0$. Thus set $\mathcal{H}_{\rho_{\mathcal{B}}}$ is the same as $\mathcal{B}$. To show the *shortest path* axiom for $\rho_{\mathcal{B}}$ one has to show that for every X there exists a point $X^* \in \mathcal{B}$ such that there is a linear reduction in risk along the path $X - X^*$. $\mathcal{B}$ is a closed convex set and assume X is a point outside this set, then by the Separating Hyperplane Theorem (see, Preliminaries, Thm: 2.8) there exists a point X^* on the boundary of $\mathcal{B}$ such that $\parallel X - X^* \parallel$ is the unique minimum distance of X from $\mathcal{B}$. Suppose λ is a scaler and u is the unit vector in the direction $X^* - X$. $X + \lambda * u$ is a point along the path of minimum distance and thus $\rho_{\mathcal{B}}(X + \lambda u) = \parallel X + \lambda * u - X \parallel = \parallel X - X^* \parallel - \parallel \lambda * u \parallel = \parallel X - X^* \parallel - \lambda = \rho_{\mathcal{B}} - \lambda$. This proves the SP property.

Now let $\mathcal{B}$ satisfy the properties of an acceptance set and $X \in \mathcal{B}$. Then by Definition 3.16 $\rho_{\mathcal{B}}(X) = 0$. Therefore, by Definition 3.15 $X \in \mathcal{A}_{\rho_{\mathcal{B}}}$. From the other direction one can get the equation in a similar way. Thus the equality of $\mathcal{B}$ and $\mathcal{A}_{\rho_{\mathcal{B}}}$ follows. □

Proposition 3.9. If a risk measure ρ is a generalized coherent risk measure, then the acceptance set induced by ρ is closed and satisfies the acceptance set axioms in Definition 3.14. Moreover $\rho = \rho_{\mathcal{A}_\rho}$.

Proof: To show that $X \in \mathcal{A}_\rho$ iff $X \geq 0$ assume $\rho(X)$ satisfies the axioms of the Definition 3.13. Subadditivity and positive homogeneity ensures that ρ is a convex function on the set of random variables. Thus it is a continuous function. Notice that the set $\rho(X) = 0$ is closed. Therefore $\mathcal{A}_\rho$ is an inverse image of a closed set under continuous mapping. Thus it is a closed set.

Assume $X \in \mathcal{A}_\rho$, so $\rho(X) = 0$. Now assume that $X \geq 0$ is not true. Thus, by the relevance axiom we must have $\rho(X) > 0$, a contradiction. Hence we must have $X \geq 0$. For the inverse implication, assume $X \geq 0$. One has to show that $X \in \mathcal{A}_\rho$. Remember that for $X \geq 0$, $\rho(X) = 0$. Thus $X \in \mathcal{A}_\rho$.

To show that $\mathcal{A}_\rho$ is convex, let $X_1, X_2 \in \mathcal{A}_\rho$ with $\rho(X_1)$ and $\rho(X_2)$ both equal to zero. Then,

$$\rho(\lambda X_1 + (1-\lambda)X_2) \leq \rho(\lambda X_1) + \rho((1-\lambda)X_2) = \lambda\rho(X_1) + (1-\lambda)\rho(X_2) = 0.$$

This indicates that $\lambda X_1 + (1-\lambda)X_2 \in \mathcal{A}_\rho$. Thus $\mathcal{A}_\rho$ is convex.

Let $X \in \mathcal{A}_\rho$. Therefore $\rho(\lambda X) = \lambda\rho(X) = 0$. Consequently $\lambda X \in \mathcal{A}_\rho$ for $\lambda \geq 0$. This proves that $\mathcal{A}_\rho$ is a positively homogenous cone.

Finally let us show that $\rho(X) = \rho_{\mathcal{A}_\rho}(X)$. Let $X \in \mathcal{A}_\rho$, then by definition $\rho(X) = 0$. Thus $\rho_{\mathcal{A}_\rho} = 0$ as well. Now consider $X \notin \mathcal{A}_\rho$ and, letX^* be a random variable as defined in the shortest path axiom. Thus X^* is on the boundary of set $\mathcal{A}_\rho$ such that $\rho_{\mathcal{A}_\rho}(X) = \parallel X - X^* \parallel$. Define u as the unit vector in the direction of $X^* - X$. Notice that $X + \parallel X - X^* \parallel u \in \mathcal{A}_\rho$ implying that $\rho(X + \parallel X - X^* \parallel u) = 0$. By the shortest path axiom we have $\rho(X) - \parallel X - X^* \parallel = 0$ i.e. $\rho(X) = \parallel X - X^* \parallel = \rho_{\mathcal{A}_\rho}(X)$. This completes the proof. □

Remark 3.5. Under the ADEH framework only the risk free asset is allowed to be added to the original portfolio to produce an acceptable portfolio. Thus the risk of a portfolio is measured as the distance from an acceptable portfolio along the path of the riskless rate. However, generalized coherent risk measures quantify the risk as the distance along the shortest path. Therefore it immediately follows that $\rho_{\mathcal{A}}^{ADEH}(X) \geq \rho_{\mathcal{A}}(X)$.

3.4 Convex Risk Measures

As stated in previous sections, coherent risk measures may fail to be an adequate measure to quantify the risk of a position in the case of illiquid market conditions. In such a market condition there may arise an additional liquidity risk if a position is multiplied by a large factor. Taking this fact into account Föllmer et all. introduced convex risk measures by setting the weaker property of convexity instead of the conditions positive homogeneity and subadditivity. First of all, Föllmer et. al. represent convex risk measures for a finite set Ω in the study [FS02a]. Then, in [FS02c] they characterize the risk measures in a situation of uncertainty, without referring to a given a priori measure. Finally, financial positions are modelled as the functions in space $\mathcal{L}^\infty$ with respect to a fixed probability measure P on a measurable space $(\Omega, \mathcal{F})$. In this section all of these three cases will be overviewed by considering the studies of Föllmer et. al.

Let Ω be a fixed set of scenarios. A discounted financial position is denoted by the mapping $X : \Omega \to \mathcal{R}$ and $\mathcal{X}$ is the linear space of functions on a given set Ω of possible scenarios. Assume that $\mathcal{X}$ contains all constant functions and is closed under the addition of constants. Also, assume that there is no priori probability measure given in the set Ω.

Definition 3.17. Convex Measure of Risk
$\rho : \mathcal{X} \to \mathcal{R}$ is called a convex measure of risk if it satisfies the following conditions for all $X, Y \in \mathcal{X}$.

1. Convexity: $\rho(\lambda X + (1-\lambda)Y) \leq \lambda\rho(X) + (1-\lambda)\rho(Y)$, for $0 \leq \lambda \leq 1$.

2.Monotonicity: If $X \leq Y$, then $\rho(X) \geq \rho(Y)$.

3.Translation Invariance: If $m \in \mathcal{R}$, then $\rho(X+m) = \rho(X) - m$.

Convexity indicates that the risk of a diversified position $\lambda X + (1-\lambda)Y$ is less or equal to the weighted average of the individual risks. Therefore, convex risk measures support portfolio diversification. Notice that convexity is equivalent to subadditivity when we assume positive homogeneity. Thus, a *convex measure of risk* is *coherent* if it satisfies positive homogeneity. Although in some situations it is convenient not to insist on it, assume for now that ρ satisfies the *normalization* property, which means $\rho(0) = 0$. This property enables us to interpret the quantity $\rho(X)$ as a 'margin requirement'.

Let $\mathcal{A} \subset \mathcal{X}$ represents acceptable positions, $\mathcal{A}_\rho$ represents the acceptance set induced by the risk measure ρ and $\rho_{\mathcal{A}}$ is the quantitative risk measure induced by the acceptance set $\mathcal{A}$. In the following two propositions, one can see the correspondence between the convex risk measures and its acceptance sets.

Proposition 3.10. Suppose $\rho : \mathcal{X} \to \mathcal{R}$ is a convex measure of risk with associated acceptance set $\mathcal{A}_\rho$. Then $\rho_{\mathcal{A}_\rho} = \rho$. Moreover, $\mathcal{A} := \mathcal{A}_\rho$ satisfies the following properties:

1. $\mathcal{A}$ is non-empty, convex and satisfies the following property:

$$\inf\{m \in \mathcal{R} \mid m \in \mathcal{A}\} > -\infty$$

2. If $X \in \mathcal{A}$ and $Y \in \mathcal{X}$ satisfies $Y \geq X$, then $Y \in \mathcal{A}$

3. If $X \in \mathcal{A}$ and $Y \in \mathcal{X}$, then

$$\{\lambda \in [0, 1] \mid \lambda X + (1 - \lambda)Y \in \mathcal{A}\}$$

is closed in $[0, 1]$.

Proof: **i)** To show that $\rho_{\mathcal{A}_\rho}(X) = \rho(X)$ for all X, first recall that

$$\rho_{\mathcal{A}_\rho}(X) = \inf\{m \mid m + X \in \mathcal{A}_\rho\}$$

then from the definition of $\mathcal{A}_\rho$

$$\rho_{\mathcal{A}_\rho}(X) = \inf\{m \mid \rho(m + X) \leq 0\}$$

finally, the translation invariance of ρ implies that

$$\rho_{\mathcal{A}_\rho} = \inf\{m \mid \rho(X) \leq m\} = \rho(X)$$

ii) 1. Firstly we will prove that $\mathcal{A}$ is non-empty. It is previously assumed that $\mathcal{X}$ contains the constants therefore translation invariance implies $\rho(0+\rho(0)) = \rho(0) - 0 \leq 0$, thus $0 \in \mathcal{A}$. To see the convexity of $\mathcal{A}$,let $X, Y \in \mathcal{A}$, then $\rho(X) \leq 0$ *and* $\rho(Y) \leq 0$. Since ρ is convex $\rho(\lambda X + (1-\lambda)Y) \leq \lambda\rho(X) + (1-\lambda)\rho(Y) \leq 0$ *for* $\lambda \in [0, 1]$. Therefore $\rho(\lambda X + (1-\lambda)Y) \in \mathcal{A}$ which proves the convexity of $\mathcal{A}$. Now assume that $\inf\{m \in \mathcal{R} \mid m \in \mathcal{A}\} = -\infty$. This implies the existence of a position such that whatever amount drawn from the position is not able to make the final net worth unacceptable. This is not compatible with finance theory.

2. Let $X \in \mathcal{A}$ and $Y \in \mathcal{X}$ satisfying $Y \geq X$, then by monotonicity $\rho(Y) \leq \rho(X)$. Since $X \in \mathcal{A}$, $\rho(X) \leq 0$. Thus one can conclude that $\rho(Y) \leq 0$, $Y \in \mathcal{A}$.

3. Only taking finite values and being convex ensures the convexity of the function $\lambda \to \rho(\lambda X + (1-\lambda)Y)$. Therefore the set of $\lambda \in [0, 1]$ such that $\rho(\lambda X + (1-\lambda)Y) \leq 0$ is closed. □

Proposition 3.11. Let $\mathcal{A} \neq \emptyset$ be a convex subset of $\mathcal{X}$ which satisfies the property 2 of Proposition 3.10 and denote by $\rho_{\mathcal{A}}$ the functional associated to the acceptance set $\mathcal{A}$ by $\rho_{\mathcal{A}}(X) := \inf\{m \in \mathcal{R} \mid m + X \in \mathcal{A}\}$. If $\rho_{\mathcal{A}}(0) > -\infty$ then,

1. $\rho_{\mathcal{A}}$ is a convex risk measure.

2. $\mathcal{A}$ is a subset of $\mathcal{A}_{\rho_{\mathcal{A}}}$. Moreover, if $\mathcal{A}$ satisfies property 3 of the Proposition 3.10, then $\mathcal{A} = \mathcal{A}_{\rho_{\mathcal{A}}}$.

Proof: **1.** Let $\alpha \in \mathcal{R}$, then $\rho_{\mathcal{A}}(X) := \inf\{m \in \mathcal{R} \mid m + X \in \mathcal{A}\}$ and $\rho_{\mathcal{A}}(X + \alpha) := \inf\{m \in \mathcal{R} \mid m + X + \alpha \in \mathcal{A}\} = \inf\{m - \alpha \in \mathcal{R} \mid m + X + \alpha - \alpha \in \mathcal{A}\} = \rho_{\mathcal{A}}(X) - \alpha$. This shows that $\rho_{\mathcal{A}}$ satisfies the *translation invariance* property. Let, $X, Y \in \mathcal{X}$ such that $X \leq Y$, then as in 3 of proof of Proposition

3.3 $\rho_{\mathcal{A}}(X) := \inf\{m \in \mathcal{R} \mid m + X \in \mathcal{A}\} \geq \inf\{m \in \mathcal{R} \mid m + Y \in \mathcal{A}\} := \rho_{\mathcal{A}}(Y)$, implies the *monotonicity*. To indicate the convexity property, suppose that $X, Y \in \mathcal{X}$ *and* $m, n \in \mathcal{R}$ are such that $m + X$ and $n + Y \in \mathcal{A}$. If $\lambda \in [0, 1]$, then the convexity of $\mathcal{A}$ implies that $\lambda(m + X) + (1 - \lambda)(n + Y) \in \mathcal{A}$. Therefore, by the translation invariance of $\rho_{\mathcal{A}}$, $0 \geq \rho_{\mathcal{A}}(\lambda(X + n) + (1 - \lambda)(Y + n)) = \rho_{\mathcal{A}}(\lambda X + (1 - \lambda)Y) - (\lambda m + (1 - \lambda)n)$, and the *convexity* property for $\rho_{\mathcal{A}}$ follows. Now we will show that $\rho_{\mathcal{A}}$ takes only finite values. Fix Y in $\mathcal{A}$. For $X \in \mathcal{X}$ given, there exists a finite number such that $m + X > Y$ because $\mathcal{X}$ is assumed to be the linear space of bounded positions. $m + X \in \mathcal{A}$, so $\rho_{\mathcal{A}}(X + m) = \rho_{\mathcal{A}}(X) - m \leq \rho_{\mathcal{A}}(Y) \leq 0$ by monotonicity and the translation invariance. Thus $\rho_{\mathcal{A}}(X) \leq m < \infty$. To show that $\rho_{\mathcal{A}} > -\infty$ for arbitrary $X \in \mathcal{X}$,take m' such that $X + m' \leq 0$, then by monotonicity $\rho_{\mathcal{A}}(X + m') \geq \rho_{\mathcal{A}}(0)$ and by translation invariance $\rho_{\mathcal{A}}(X) \geq \rho_{\mathcal{A}}(0) + m' > -\infty$. This concludes the proof of the first part.

2. Assume that $\mathcal{A}$ satisfies the closure property in Proposition 3.10. We have to show that $X \notin \mathcal{A}$ implies that $\rho_{\mathcal{A}}(X) > 0$. To this end take $m > \rho_{\mathcal{A}}(0)$. By property 3 of Proposition 3.10, there exists an $\varepsilon \in (0, 1)$ such that $\varepsilon m + (1 - \varepsilon)X \notin \mathcal{A}$. Thus,

$$\varepsilon m \leq \rho_{\mathcal{A}}((1 - \varepsilon)X) = \rho_{\mathcal{A}}(\varepsilon.0 + (1 - \varepsilon)X)$$

$$\leq \varepsilon \rho_{\mathcal{A}}(0) + (1 - \varepsilon)\rho_{\mathcal{A}}(X)$$

Hence

$$\rho_{\mathcal{A}}(X) \geq \frac{\varepsilon(m - \rho_{\mathcal{A}}(0))}{1 - \varepsilon} > 0,$$

and property 2 follows. □

After realizing the correspondence between convex risk measures and their acceptance sets, we are ready to concentrate on the structure theorems of convex risk measures for different cases. Firstly we will consider the case in which $\mathcal{X}$ is the space of all real-valued functions on some **finite** set Ω. Then we will review the case in which financial positions are modelled in the space of all bounded measurable functions on a measurable space $(\Omega, \mathcal{F})$ without referring to a given priori measure. The final case will be the one in which $\mathcal{X}$ is identified with $\mathcal{L}^\infty(\Omega, \mathcal{F}, P)$ under the assumption that we have a probabilistic model given by a probability measure P on $(\Omega, \mathcal{F})$.

Theorem 3.5. Suppose $\mathcal{X}$ is the space of all real-valued functions on a finite set Ω and $\mathcal{P}$ is the set of all probability measures on Ω. Then $\rho : \mathcal{X} \to \mathcal{R}$ is a convex measure of risk if and only if there exists a "penalty function" $\alpha : \mathcal{P} \to (-\infty, \infty]$ such that

$$\rho(Z) = \sup_{Q \in \mathcal{P}} (E_Q[-Z] - \alpha(Q)).$$

The function α satisfies $\alpha(Q) \geq -\rho(0)$ for any $Q \in \mathcal{P}$.

Proof: **i)** To prove the 'if' part of the theorem, one has to show that for each $Q \in \mathcal{P}$ the functional

$$X \to E_Q[-X] - \alpha(Q)$$

is convex, monotone and translation invariant.

1. Let $X, Y \in \mathcal{X}$ and $\lambda \in [0, 1]$, then

$$E_Q[-(\lambda X + (1-\lambda)Y)] - \alpha(Q) \leq \lambda(E_Q[-X] - \alpha(Q)) + (1-\lambda)(E_Q[-Y] - \alpha(Q))$$

Thus, *convexity* for the functional follows.

2. For $X, Y \in \mathcal{X}$ such that $X \leq Y$

$$E_Q[-X] - \alpha(Q) \geq E_Q[-Y] - \alpha(Q)$$

indicating the monotonicity property.

3. For $m \in \mathcal{R}$

$$E_Q[-(X + m)] - \alpha(Q) = E_Q[-X] - \alpha(Q) - m$$

proves the translation invariance property.

After proving the convexity, monotonicity and translation invariance for the functional, we can conclude that

$$\rho(Z) = \sup_{Q \in \mathcal{P}} (E_Q[-Z] - \alpha(Q))$$

also satisfies convexity, monotonicity and translation invariance because these three are preserved under the operation of taking suprema.

ii) To prove the converse part, for $Q \in \mathcal{P}$, $\alpha(Q)$ is defined by

$$\alpha(Q) := \sup_{X \in \mathcal{X}} (E_Q[-X] - \rho(X)) \tag{3.4.13}$$

Then it is claimed that

$$\alpha(Q) = \sup_{X \in \mathcal{A}_\rho} E_Q[-X] \tag{3.4.14}$$

Notice that function (3.4.14) represents the minimum acceptable amount for the acceptance set $\mathcal{A}_\rho$. In the equation (3.4.14) denote the righthand side by $\hat{\alpha}(Q)$. By using the definition of $\mathcal{A}_\rho$, it can be found that

$$\alpha(Q) \geq \hat{\alpha}(Q) \tag{3.4.15}$$

To establish the converse inequality, take an arbitrary $X \in \mathcal{X}$ and recall that $X' := \rho(X) + X \in \mathcal{A}_\rho$. From the definitions of $\alpha(Q)$ and $\hat{\alpha}(Q)$ the following inequality holds for all $X \in \mathcal{X}$.

$$\hat{\alpha}(Q) \geq E_Q[-X'] = E_Q[-X] - \rho(X)$$
$$\hat{\alpha}(Q) \geq \sup_{X \in \mathcal{X}} (E_Q[-X] - \rho(X)) = \alpha(Q) \tag{3.4.16}$$

Equation (3.4.15) together with (3.4.16) shows that $\alpha(Q) = \hat{\alpha}(Q)$ and proves the claim.

Now fix some $Y \in \mathcal{X}$ and take $\alpha(.)$ as in the equation (3.4.13), then

$$E_Q[-Y] - \rho(Y) \leq \sup_{X \in \mathcal{X}} (E_Q[-X] - \rho(X)) = \alpha(Q)$$

Thus

$$\rho(Y) \geq E_Q[-Y] - \alpha(Q).$$

Taking the supremum preserves the inequality. Therefore

$$\rho(Y) \geq \sup_{Q \in \mathcal{P}} (E_Q[-Y] - \alpha(Q)) \tag{3.4.17}$$

Now take $m \in \mathcal{R}$ such that

$$m > \sup_{Q \in \mathcal{P}} (E_Q[-Y] - \alpha(Q))$$

Then one has to show that $m \geq \rho(Y)$ or, equivalently, $m + Y \in \mathcal{A}_\rho$. Suppose that the inverse is true and $m + Y \notin \mathcal{A}_\rho$. Since ρ is by definition a convex function on the space $\mathcal{R}^\Omega$ and only takes finite values, ρ is continuous (see Preliminaries, Thm:2.3). Hence $\mathcal{A}_\rho = \{\rho \leq 0\}$ is a closed convex set. (Notice that here we are using the assumption that Ω is finite, in order to obtain the closedness of the acceptance set $\mathcal{A}_\rho$)

Since $\mathcal{A}_\rho$ is a convex set, by using the separation theorem (see, Preliminaries Thm: 2.9), one can find a linear functional l *on* $\mathcal{R}^\Omega$ such that

$$\beta := \sup_{X \in \mathcal{A}_\rho} l(X) < l(m + Y) =: \gamma < \infty \tag{3.4.18}$$

The axioms of monotonicity and normalization imply that

$$\rho(X) \leq \rho(0) \ for \ X \geq 0.$$

Thus if $X \in \mathcal{X}$ satisfies $X \geq 0$, then $\lambda X + \rho(0) \in \mathcal{A}_\rho$ for all $\lambda \geq 1$, and hence

$$\gamma > l(\lambda X + \rho(0))$$

Since l is a linear functional we have

$$\gamma > \lambda l(X) + \rho(0)$$

Taking $\lambda \uparrow \infty$ yields that $l(X) \leq 0$. Thus it follows that l is a negative functional. Assume that $l(1) = -1$, then by using the assumption one can define the probability measure $Q \in \mathcal{P}$ as

$$Q[A] := l(-I_A)$$

Then the equation $E_Q[-X] = l(X)$ holds for $X \in \mathcal{A}_\rho$. By using the equations (3.4.14) and (3.4.18)

$$\alpha(Q) = \sup_{X \in \mathcal{A}_\rho} E_Q[-X] = \beta$$

but

$$E_Q[-Y] - m = l(m + Y) = \gamma > \beta = \alpha(Q)$$

which implies

$$\rho(Y) > m$$

which is a contradiction to the choice of m. Therefore, we must have $m + Y \in \mathcal{A}_\rho$ and thus, $m \geq \rho(Y)$ □

Remark 3.6. Theorem 3.5 above includes the structure theorem for coherent risk measures as a special case. ρ will possess the property of positive homogeneity, i.e., ρ will be a coherent measure of risk, if and only if the penalty function $\alpha(Q)$ only takes values 0 and $+\infty$. In this case, the theorem above implies the representation of coherent risk measures in terms of the set

$$\mathcal{Q} = \{Q \in \mathcal{P} \mid \alpha(Q) = 0\}$$

From now on assume that $\mathcal{X}$ is the **linear space of all bounded measurable functions on a measurable space** $(\Omega, \mathcal{F})$. $\mathcal{M}_1 := \mathcal{M}_1(\Omega, \mathcal{F})$ indicates the class of all probability measures on $(\Omega, \mathcal{F})$. Moreover the larger class $\mathcal{M}_{1,f}$ represents all finitely additive and non-negative set functions $Q : \mathcal{F} \to [0, 1]$ which are normalized to $Q[\Omega] = 1$. Here, risk measures are characterized in a situation of uncertainty without referring to any probability on $(\Omega, \mathcal{F})$ given a priori. In this case, firstly, convex risk measures are represented by finitely additive set functions. Later on, a criteria which guarantees that a measure of risk can be represented in terms of σ-additive probability measures is obtained. Let $\alpha : \mathcal{M}_{1,f} \to \mathcal{R} \cup \{+\infty\}$ be any functional which is bounded below and which is not identically equal to $+\infty$, then for each $Q \in \mathcal{M}_{1,f}$ the functional $X \to E_Q[-X] - \alpha(Q)$is convex, monotone and translation invariant on $\mathcal{X}$, and these properties are preserved when taking the supremum over $Q \in \mathcal{M}_{1,f}$. Thus

$$\rho(X) := \sup_{Q \in \mathcal{M}_{1,f}} (E_Q[-X] - \alpha(Q)) \qquad (3.4.19)$$

defines a convex risk measure on $\mathcal{X}$ such that

$$\rho(0) = - \inf_{Q \in \mathcal{M}_{1,f}} (\alpha(Q))$$

Theorem 3.6. Any convex measure of risk ρ on $\mathcal{X}$ is of the form

$$\rho(X) = \max_{Q \in \mathcal{M}_{1,f}} (E_Q[-X] - \alpha_{\min}(Q)), \qquad X \in \mathcal{X}, \qquad (3.4.20)$$

where the penalty function $\alpha_{\min}$ is given by

$$\alpha_{\min}(Q) := \sup_{X \in \mathcal{A}_\rho} E_Q[-X] \ for \ Q \in \mathcal{M}_{1,f}$$

Moreover, $\alpha_{\min}$ is the minimal penalty function which represents ρ, i.e., any penalty function α for which (3.4.19) holds satisfies $\alpha(Q) \geq \alpha_{\min}(Q)$ for all $Q \in \mathcal{M}_{1,f}$

Proof: As a first step, it will be shown that

$$\rho(X) \geq \sup_{Q \in \mathcal{M}_{1,f}} (E_Q[-X] - \alpha(Q)) \; \forall \; X \in \mathcal{X}$$

Recall that $X' := \rho(X) + X \in \mathcal{A}_\rho$. Also recall that

$$\alpha_{\min}(Q) := \sup_{X \in \mathcal{A}_\rho} E_Q[-X]$$

Therefore, $\alpha_{\min}(Q) \geq E_Q[-X'] = E_Q[-X] - \rho(X)$, then

$$\alpha_{\min}(Q) \geq E_Q[-X] - \rho(X)$$

and

$$\rho(X) \geq E_Q[-X] - \alpha_{\min}(Q) \tag{3.4.21}$$

Now, as a second step, for a given X, some $Q_X \in \mathcal{M}_{1,f}$ will be constructed such that

$$\rho(X) \leq E_{Q_X}[-X] - \alpha_{\min}(Q_X)$$

which, in view of the previous step, will prove the representation (3.4.20). By the translation invariance property it is sufficient to prove this only for $X \in \mathcal{X}$ with $rho(X) = 0$. Moreover, the normalization assumption, $\rho(0) = 0$ is still valid. Therefore such an X is not contained in the convex set

$$\mathcal{B} := \{Y \in \mathcal{X} \mid \rho(Y) < 0\}$$

Since ρ is normalized, $\mathcal{B}$ contains the open ball

$$\mathcal{B}_1(1) = \{Y \in \mathcal{X} \mid \|Y - 1\| < 1\}$$

This ensures that $\mathcal{B}$ has a non-empty interior. It is known that $\rho(X) = 0$ and $\rho(Y) < 0$ then, by monotonicity, $Y > X$. Since $\mathcal{B}$ is a convex set with non-empty interior and $Y > X$ is known, one can apply the separation theorem, which yields a non-zero continuous functional l on $\mathcal{X}$ such that

$$l(X) \leq \inf_{Y \in \mathcal{B}} l(Y) := b. \tag{3.4.22}$$

This implies that we are able to separate the space of positions $\mathcal{X}$ into the sets of acceptable and unacceptable positions.
Now, the claim is that, $l(Y) \geq 0$ if $Y \geq 0$. The translation invariance of ρ implies that $\rho(1 + \lambda Y) = \rho(\lambda Y) - 1$. Besides, from monotonicity, for $\lambda > 0$, $\rho(Y) > \rho(\lambda Y)$ is true. It is assumed that $\rho(Y) < 0 \; for \; Y \in \mathcal{B}$. Putting all these together, one can conclude that $\rho(\lambda Y) - 1 < 0$, and so $1 + \lambda Y \in \mathcal{B}$ for any $\lambda > 0$. Therefore, by using (3.4.22) and the linearity of l

$$l(X) \leq l(1 + \lambda Y) = l(1) + \lambda l(Y), \; \forall \; \lambda > 0$$

This proves the claim since the equation above could not be true if $l(Y) < 0$.
The second claim is that $l(1) > 0$. Since l does not vanish identically and $\mathcal{B}$ contains the unit ball, there must be some $Y \in \mathcal{B}_1(0)$ such that $0 < l(Y) = l(Y^+) - l(Y^-)$. The previous claim implies that $l(Y^+) > 0$ and since we choose Y from $\mathcal{B}_1(0)$, $l(1 - Y^+) \geq 0$. Hence $l(1) = l(1 - Y^+) + l(Y^+) > 0$. By the two preceding steps and Theorem 2.4 of Preliminaries, there exists some $Q_X \in \mathcal{M}_{1,f}$ such that

$$E_{Q_X}[Y] = \frac{l(Y)}{l(1)} \text{forall } Y \in \mathcal{X}. \tag{3.4.23}$$

In the equation above, to define the expectation, l is normalized by dividing it into $l(1)$.
$\mathcal{A}_\rho = \{Y \mid \rho(Y) \leq 0\}$ and $\mathcal{B} = \{Y \mid \rho(Y) < 0\}$, so it is clear that $\mathcal{B} \subset \mathcal{A}_\rho$, and so

$$\alpha_{min}(Q_X) = \sup_{Y \in \mathcal{A}_\rho} E_{Q_X}[-Y] \geq \sup_{Y \in \mathcal{B}} E_{Q_X}[-Y] = -\frac{b}{l(1)}$$

On the other hand, $Y+\varepsilon \in \mathcal{B}$ for any $Y \in \mathcal{A}_\rho$ and each $\varepsilon > 0$. This shows that

$$\alpha_{\min}(Q_X) = -\frac{b}{l(1)} \tag{3.4.24}$$

By using (3.4.23)and (3.4.24),

$$E_{Q_X}[-X] - \alpha_{\min}(Q_X) = -\frac{l(X)}{l(1)} + \frac{b}{l(1)} = \frac{1}{l(1)}(b - l(X)) \geq 0$$

Here, the righthand side of the equation is non-negative, since $l(1) > 0$ is proved previously and $(b - l(X)) \geq 0$ due to (3.4.22). Remember that in the beginning of the proof X were chosen such that $\rho(X) = 0$ Thus

$$E_{Q_X}[-X] - \alpha_{\min}(Q_X) \geq 0 = \rho(X) \tag{3.4.25}$$

Therefore (3.4.25) together with (3.4.21) indicates that

$$\rho(X) = E_Q[-X] - \alpha_{\min}(Q)$$

which concludes the proof of representation (3.4.20).
Finally, let α be any penalty function for ρ. Then for all $Q \in \mathcal{M}_{1,f}$ and $X \in \mathcal{X}$

$$\rho(X) \geq E_Q[-X] - \alpha(Q)$$

and hence

$$\begin{aligned} \alpha(Q) &\geq \sup_{X \in \mathcal{X}} (E_Q[-X] - \rho(X)) \\ &\geq \sup_{X \in \mathcal{A}_\rho} (E_Q[-X] - \rho(X)) \\ &\geq \alpha_{\min}(Q) \end{aligned} \tag{3.4.26}$$

This shows that α dominates α_{min}. □

Remark 3.7. a. If we take $\alpha = \alpha_{\min}$ in (3.4.26) we obtain an alternative formula for $\alpha_{\min}$:

$$\alpha_{min} = \sup_{X \in \mathcal{X}} (E_Q[-X] - \rho(X))$$

b. If ρ is defined by $\rho = \rho_{\mathcal{A}}$ for a given acceptance set $\mathcal{A} \subset \mathcal{X}$, then the predetermined acceptance set $\mathcal{A}$ determines $\alpha_{\min}$:

$$\alpha_{\min}(Q) = \sup_{Q \in \mathcal{Q}} E_Q[-X], \ \ for \ all \ Q \in \mathcal{M}_{1,f}$$

In fact, this is because of the fact that $X \in \mathcal{A}$ implies $\varepsilon + X \in \mathcal{A}_\rho$ for all $\varepsilon > 0$.

The following corollary gives the representation of coherent risk measures in terms of the minimal penalty function.

Corollary 3.1. The minimal penalty function α_{min} of a coherent risk measure only takes values 0 and $+\infty$. In particular,

$$\rho(X) = \max_{Q \in \mathcal{Q}_{max}} E_Q[-X], \quad X \in \mathcal{X},$$

for the set

$$\mathcal{Q}_{max} := \{Q \in \mathcal{M}_{f,1} \mid \alpha_{min}(Q) = 0\}$$

Proof: Recall that the acceptance set of a coherent risk measure is a cone. Thus the minimal penalty function satisfies

$$\alpha_{min}(Q) = \sup_{X \in \mathcal{A}_\rho} E_Q[-X] = \sup_{\lambda X \in \mathcal{A}_\rho} E_Q[-\lambda X] = \lambda \alpha_{min}(Q)$$

For all $Q \in \mathcal{M}_{1,f}$ and $\lambda > 0$. Thus, α_{min} can only take values 0 and $+\infty$. □

After representing convex risk measures in terms of finitely additive set functions, in [FS02c] Föllmer et. al. introduced the situation in which a convex risk measure of risk admits a representation in terms of σ-additive probability measures. In such a case a convex measure of risk can be represented by a penalty function α which is infinite outside the set $\mathcal{M}_1 = \mathcal{M}_1(\Omega, \mathcal{F})$:

$$\rho(X) = \sup_{Q \in \mathcal{M}_{1,f}} (E_Q[-X] - \alpha(Q)). \tag{3.4.27}$$

A representation in terms of probability measures is closely related with the continuity properties of ρ. The following lemma shows that a convex risk measure that can be represented in the form (3.4.27) is continuous from above.

Lemma 3.1. A convex measure of risk ρ which admits a representation (3.27) on $\mathcal{M}_1$ is continuous from above in the sense that

$$X_n \searrow X \Longrightarrow \rho(X_n) \nearrow \rho(X) \tag{3.4.28}$$

Moreover,if (X_n) is a bounded sequence in $\mathcal{X}$ which converges pointwise to $X \in \mathcal{X}$, then

$$\rho(X) \leq \liminf_{n\uparrow\infty} \rho(X_n) \tag{3.4.29}$$

Proof: Firstly, assume X_n is a bounded sequence converging pointwise to X, then the dominated convergence theorem (see, Preliminaries Thm: 2.7) states that

$$E_Q[\lim_{n\to\infty} X_n] = \lim_{n\to\infty} E_Q[X_n]$$

then

$$E_Q[X] = \lim_{n\to\infty} E_Q[X_n]$$

Thus as n tends to infinity we have

$$E_Q[X_n] \to E_Q[X] \; for \; all \; Q \in \mathcal{M}_1.$$

Hence

$$\begin{aligned}
\rho(X) &= \sup_{Q\in\mathcal{M}_1} (\lim_{n\uparrow\infty} E_Q[-X_n] - \alpha(Q)) \\
&\leq \liminf_{n\uparrow\infty} \sup_{Q\in\mathcal{M}_1} (E_Q[-X_n] - \alpha(Q)) \\
&= \liminf_{n\uparrow\infty} \rho(X_n)
\end{aligned}$$

which proves (3.4.29).

To show the equivalence of (3.4.28) and (3.4.29), first assume that (3.4.29) holds. By using monotonicity of ρ, $\rho(X_n) \leq \rho(X)$ for each n if $X_n \searrow X$, and so $\rho(X_n) \nearrow \rho(X)$ follows. Now to prove the inverse implication assume continuity above. Let (X_n) be a sequence in $\mathcal{X}$ which converge pointwise to X. Define a sequence Y_m such that $Y_m := sup_{n \geq m} X_n, X_n \in \mathcal{X}$. Then clearly Y_m decreases to X. Thus, by monotonicity $\rho(X_n) \geq \rho(Y_n)$. When we take the infimum the equation preserved. Therefore by using the condition (3.4.28)

$$\liminf_{n \uparrow \infty} \rho(X_n) \geq \lim_{n \uparrow \infty} \rho(Y_n) = \rho(X)$$

□

In the following proposition Föllmer et. al. give a strong sufficient condition, continuity from below, that guarantees the concentration of any penalty function on the set $\mathcal{M}_1$ of probability measures.

Proposition 3.12. Let ρ be a convex measure of risk which is continuous from below in the sense that

$$X_n \nearrow X \Rightarrow \rho(X_n) \searrow \rho(X)$$

and suppose that α is any penalty function on $\mathcal{M}_1$ representing ρ. Then α is concentrated on the class $\mathcal{M}_1$ of probability measures, i.e.,

$$\alpha(Q) < \infty \Rightarrow Q \textit{ is } \sigma \textit{ additive.}$$

However, continuity from below in this strong form might be too restrictive in many situations. Therefore, to obtain a weaker version, Föllmer et. al. discussed the problem in a topological setting. They made the assumption that Ω is a Polish space, i.e., a separable topological space admitting a complete metric, and F is the σ-field of Borel sets. For the rest of the discussion and proof of Proposition 3.12, see [FS02b, p:171].

In the last case for the representation of convex risk measures, $\mathcal{X}$ is identified with $\mathcal{L}^\infty(\Omega, \mathcal{F}, P)$ under the assumption that P is a probability measure given a priori on $(\Omega, \mathcal{F})$. In this situation, it is naturel to think ρ such that

$$\rho(X) = \rho(Y) \qquad if \;\; X = Y \;\; P - a.s. \tag{3.4.30}$$

Lemma 3.2. Let ρ be a convex measure of risk that satisfies (3.4.30) and which is represented by a penalty function α as

$$\rho(X) := \sup_{Q \in \mathcal{M}_1} (E_Q[-X] - \alpha(Q)).$$

Then $\alpha(Q) = +\infty$ for any probability measure Q which is absolutely continuous with respect to P.

Proof: If $Q \in \mathcal{M}_1(\Omega, \mathcal{F})$ is not absolutely continuous with respect to P, then from the definition of absolute continuity (see Preliminaries, Def: 2.16), there exists $A \in \mathcal{F}$ such that $Q[A] > 0$ but $P[A] = 0$. Let $X \in \mathcal{A}_\rho$, then define $X_n := X - nI_A$. We have $X - nI_A = X$ P a.s., then $\rho(X_n) = \rho(X)$. Therefore X_n is contained in $\mathcal{A}_\rho$. Hence

$$\alpha(Q) \geq \alpha_{min}(Q) \geq E_Q[-X_n] = E_Q[-X] + nQ[A] \rightarrow \infty \qquad as \;\; n \uparrow \infty$$

□

After this Lemma, it is appropriate to conclude this section with the representation theorem of convex risk measures on $\mathcal{L}^\infty$.

Theorem 3.7. Suppose $\mathcal{X} = \mathcal{L}^\infty(\Omega, \mathcal{F}, P), \mathcal{P}$ is the set of probability measures $Q \ll P$, and $\rho : \mathcal{X} \rightarrow \mathcal{R}$ is a convex measure of risk. Then the following properties are equivalent.

1. There is a penalty function $\alpha : \mathcal{P} \rightarrow (-\infty, \infty]$ such that

$$\rho(X) = \sup_{Q \ll P} (E_Q[-X] - \alpha(Q)) \qquad \forall X \in \mathcal{X}$$

2. The acceptance set $\mathcal{A}_\rho$ associated with ρ is weak$*$-, i.e.,$\sigma(\mathcal{L}^\infty(P), \mathcal{L}_1(P))$-closed.

3. ρ posses the Fatou property: If the sequence $(X_n)_{n \in N} \subset \mathcal{X}$ is uniformly bounded, and X_n converges to some $X \in \mathcal{X}$ in probability, then $\rho(X) \leq \liminf_n \rho(X_n)$.

4. If the sequence $(X_n)_{n \in N} \subset \mathcal{X}$ decreases to $X \in \mathcal{X}$, then $\rho(X_n) \rightarrow \rho(X)$

For the proof of the theorem see [FS02a, p:436]

It is seen that convex risk measures are superior in terms of their awareness of liquidity risk. With this property convex risk measures get a lead among the risk measures we reviewed up to now. However, similar to other risk measures convex measures assume full and symmetric information among all agents of the market. This seems unrealistic since there is a discrepancy between the amount of information that the agents have. In the next section we will become familiar with a new measure of risk which takes account of the existence of asymmetric information in financial markets.

3.5 Conditional and Convex Conditional Risk Measures

Risk measures we reviewed up to this section assume that every agent in the market has full information. However in the real market condition this is not the case. In order to deal with situations of partial or asymmetric information, Bion-Nadal introduced the concept of conditional risk measures. Throughout this section the theory of conditional risk measures will be reviewed by taking the study [BN04] as a reference.

To begin with, in [BN04], partial information is defined as a situation where the set of financial positions is a linear space $\mathcal{X}$ of bounded maps on a space Ω, and where the investor does not have access to all maps defined on Ω, but only to the measurable maps relative to the σ-algebra $\mathcal{F}$. Here $\mathcal{F}$ represents all **accessible** information. Conditional risk measures has been defined under both **partial** and **complete uncertainty**. With partial uncertainty, the situation is implied in which the investor has access to partial information represented by the σ-algebra $\mathcal{F}$ and a probability measure is given on this σ-algebra $\mathcal{F}$ (implying that the agent knows at least which events of $\mathcal{F}$ are of null probability). On the other hand, the case of complete uncertainty represents the situation in which one does not know which probability is the 'good one' even on the σ-algebra $\mathcal{F}$. The main aim in defining conditional risk measures is to reach a robust representation for convex conditional risk measures. Here we will first consider the definition and properties of conditional risk measures under partial uncertainty and then continue with the complete uncertainty case. In the sequel, the representation theorems for convex conditional risk measures for the related case will be given.

Let Ω represents the states of nature. A financial position is described by a bounded map defined on the set Ω. $\mathcal{X}$ represents the linear space of financial positions. Consider a σ-algebra $\mathcal{F}$ on the space Ω. Then $(\Omega, \mathcal{F})$ is a measurable space. $\varepsilon_{\mathcal{F}}$ indicates the set of all bounded real valued $(\Omega, \mathcal{F})$ measurable maps. Let P be a a priori given probability on $\mathcal{F}$, then the aim is to define a notion of a risk measure conditional to $(\Omega, \mathcal{F}, P)$.

Definition 3.18. Conditional Risk Measure under Partial Uncertainty
A mapping

$$\rho_{\mathcal{F}} : \mathcal{X} \to \mathcal{L}^{\infty}(\Omega, \mathcal{F}, P)$$

is called a risk measure conditional to $(\Omega, \mathcal{F}, P)$ if it satisfies the following conditions:

1. **Monotonicity:** For all $X, Y \in \mathcal{X}$ if $X \leq Y$, then $\rho_{\mathcal{F}}(Y) \leq \rho_{\mathcal{F}}(X)$ P $a.s.$

2. Translation Invariance: For all $Y \in \varepsilon_{\mathcal{F}}$, for all $X \in \mathcal{X}$,

$$\rho_{\mathcal{F}}(X+Y) = \rho_{\mathcal{F}}(X) - Y \;\; P \;\; a.s.$$

3. Multiplicative Invariance: For all $X \in \mathcal{X}$, for all $A \in \mathcal{F}$,

$$\rho_{\mathcal{F}}(XI_A) = I_A \rho_{\mathcal{F}}(X) \;\; P \;\; a.s.$$

The interpretation of the first two properties are similar to the ones we considered in the previous sections. The new property, multiplicative invariance, evokes the property of conditional expectation.

Lemma 3.3.
1. Any conditional risk measure is Lipschitz continuous with a Lipschitz constant equal to 1 from $(\mathcal{X}, \| \,.\, \|)$ to $\mathcal{L}^\infty(\Omega, \mathcal{F}, P)$, i.e.

$$\| \rho_{\mathcal{F}}(X) - \rho_{\mathcal{F}}(Y) \| \leq \| X - Y \|, \forall X, Y \in \mathcal{X}$$

2. The restriction to $\varepsilon_{\mathcal{F}}$ of any risk measure conditional to $(\Omega, \mathcal{F}, P)$ is equal to the negative of the identity,i.e., $-id \; P \; a.s.$

For the proof of Lemma see [BN04, p:7]

Definition 3.19. The $\mathcal{F}$- acceptance set of a risk measure $\rho_{\mathcal{F}}$ conditional to the probability space $(\Omega, \mathcal{F}, P)$ is

$$\mathcal{A}_{\rho_{\mathcal{F}}} = \{X \in \mathcal{X} \mid \rho_{\mathcal{F}}(X) \leq 0 \;\; P \;\; a.s\}$$

The following two propositions indicate the correspondence between a conditional risk measure and a $\mathcal{F}$-acceptance set.

Proposition 3.13. The acceptance set $\mathcal{A} = \mathcal{A}_{\rho_{\mathcal{F}}}$ of the risk measure $\rho_{\mathcal{F}}$ has the following properties:

1. $\mathcal{A}$ is non empty, closed with respect to the supremum norm and has a

hereditary property : for all $X \in \mathcal{A}$, for all $Y \in \mathcal{X}$, if $Y \geq X$, then $Y \in \mathcal{A}$.

2. Bifurcation property: for all $X_1, X_2 \in \mathcal{A}$ and for all B_1, B_2 disjoint sets $\in \mathcal{F}$, $X = X_1 I_{\{B_1\}} + X_2 I_{\{B_2\}}$ is in $\mathcal{A}$.

3. Positivity: Every element $\mathcal{F}$-measurable of $\mathcal{A}$ is positive P $a.s$

4. $\rho_{\mathcal{F}}$ can be recovered from $\mathcal{A}$

$$\rho_{\mathcal{F}} = essinf\{Y \in \varepsilon_{\mathcal{F}} \mid X + Y \in \mathcal{A}\}$$

Proof: 1. By the multiplicative invariance property we have $\rho_{\mathcal{F}}(0) = 0$. Therefore by definition of $\mathcal{A}$, $0 \in \mathcal{A}$. Thus it is non empty. To show the closedness let $X \in \mathcal{X} - \mathcal{A}$ and $\epsilon > 0$ such that $P(\{\omega \in \Omega \mid \rho_{\mathcal{F}}(X)(\omega) > \epsilon\}) > 0$. Let $\mathcal{O} = \{Y \in \mathcal{X} \mid \| \rho_{\mathcal{F}}(Y) - \rho_{\mathcal{F}}(X) \| < \epsilon\}$. $\mathcal{O}$ is an open subset of $(\mathcal{X}, \| \, . \, \|)$ because $\rho_{\mathcal{F}}$ is Lipschitz continuous as stated in Lemma 3.3. Moreover $\mathcal{O}$ is contained in $\mathcal{X} - \mathcal{A}$. So $\mathcal{X} - \mathcal{A}$ is clearly open. Then its complement, $\mathcal{A}$, is closed. Hereditary property is an overt consequence of monotonicity. Indeed, for $X \in \mathcal{A}$ $\rho_{\mathcal{F}}(X) \leq 0$ from the definition of $\mathcal{A}$. Now let $Y \in \mathcal{X}$ such that $Y \geq X$. By the monotonicity of $\rho_{\mathcal{F}}$, $\rho_{\mathcal{F}}(Y) \leq \rho_{\mathcal{F}}(X) \leq 0$. Thus $Y \in \mathcal{A}$.
2. From the multiplicative invariance

$$\rho_{\mathcal{F}}(X_i I_{\{B_i\}}) = \rho_{\mathcal{F}}(X_i) I_{\{B_i\}}$$

and $\rho_{\mathcal{F}}(0) = 0$, then for $X_1, X_2 \in \mathcal{A}$ we have $\rho_{\mathcal{F}}(X) = \rho_{\mathcal{F}}(X_1) I_{\{B_1\}} + \rho_{\mathcal{F}}(X_2) I_{\{B_2\}}$ which is clearly contained in $\mathcal{A}$.
3. From Lemma 3.3 we have $\rho_{\mathcal{F}}(X) = -X$ P $a.s.$ Then from the definition of $\mathcal{A}$ we have $X \geq 0$ for $X \in \mathcal{A}$

4. Let $X \in \mathcal{X}$ and $\mathcal{B}_X = \{f \in \varepsilon_{\mathcal{F}} \mid X + f \in \mathcal{A}\}$. From the translation invariance we have $\rho_{\mathcal{F}}(X + \rho_{\mathcal{F}}) = 0$, so $\rho_{\mathcal{F}}(X + \rho_{\mathcal{F}}(X)) \in \mathcal{A}$. As a consequence, $\rho_{\mathcal{F}}(X) \in \mathcal{B}_X$. On the other hand, for every $f \in \mathcal{B}_X$, $\rho_{\mathcal{F}}(X+f) = \rho_{\mathcal{F}}(X) - f \leq 0$ $P\ a.s \rightarrow \rho_{\mathcal{F}}(X) \leq f\ P\ a.s.$. This guarantees that B_X has a minimal element which is equal to $\rho_{\mathcal{F}}(X)\ P\ a.s.$ Therefore

$$\rho_{\mathcal{F}}(X) = essinf\{f \in \varepsilon_{\mathcal{F}} \mid f + X \in \mathcal{A}\}$$

□

Proposition 3.14. Let $\mathcal{A}$ be a $\mathcal{F}$- acceptance set. For all $X \in \mathcal{X}$ consider the set

$$\mathcal{B}_X = \{Y \in \varepsilon_{\mathcal{F}}, X + Y \in \mathcal{A}\}$$

Then $\rho_{\mathcal{F}}(X) = essinf \mathcal{B}_X$ is a risk measure conditional to the probability space $(\Omega, \mathcal{F}, P)$.

For the proof see [BN04, p:9]

From the previous two propositions we see that a conditional risk measure can be defined directly or it can be defined from a $\mathcal{F}$-acceptance set. From the definition of B_X in Proposition 3.14 it is seen that to make a position acceptable we add another position to the initial one. In other words, a risk measure of a position X conditional to the σ-algebra $\mathcal{F}$ is implicitly defined as the minimal $\mathcal{F}$- measurable map (implying another position) which added to the initial position X makes the position acceptable. This is different from the cases we saw on coherent and convex risk measures; instead, it is similar to the situation in generalized coherent risk measure since in the latter, to remove the risk of a position, an agent is allowed to hold not only the risk free asset but various other kinds of assets.

Definition 3.20. Convex Conditional Risk Measures
A risk measure defined on $\mathcal{X}$ conditional to the probability space $(\Omega, \mathcal{F}, P)$ is

called **convex** if for all $X, Y \in \mathcal{X}$, for all $0 \leq \lambda \leq 1$,

$$\rho_{\mathcal{F}}(\lambda X + (1-\lambda)Y) \leq \lambda \rho_{\mathcal{F}}(X) + (1-\lambda)\rho_{\mathcal{F}}(Y) \quad P \quad a.s$$

Definition 3.21. Continuity From Below
A convex conditional risk measure is continuous from below if:
For all increasing sequence X_n of elements of $\mathcal{X}$ converging to X, the decreasing sequence $\rho_{\mathcal{F}}(X_n)$ converges to $\rho_{\mathcal{F}}(X)$ P $a.s.$.

After this introductory part we will continue with the representation theorem of convex conditional risk measures in the case of partial uncertainty.
Assume that there is a σ-algebra $\mathcal{G}$ such that $\mathcal{X}$ is the set of all bounded measurable functions on the measure space $(\Omega, \| \, . \, \|)$. Let $\mathcal{F}$ be a sub-σ-algebra of $\mathcal{G}$. $\mathcal{M}_{1,f}$ represents the set of all finitely additive set functions $\mathcal{Q} : \mathcal{G} \to [0,1]$ such that $\mathcal{Q}(\Omega) = 1$.

Theorem 3.8. Let $\rho_{\mathcal{F}}$ be a convex risk measure conditional to probability space $(\Omega, \mathcal{F}, P)$. Assume that $\rho_{\mathcal{F}}$ is continuous from below, then for all $X \in \mathcal{X}$

$$\rho_{\mathcal{F}}(X) = essmax_{Q \in \mathcal{M}}(E_Q(-X \mid \mathcal{F}) - \alpha(Q))$$

where $\alpha(Q) = esssup_{Y \in \mathcal{A}_{\rho_{\mathcal{F}}}} E_Q(-Y \mid \mathcal{F})$ and

$\mathcal{M}$ is a set of probability measures on $(\Omega, \mathcal{G})$ whose restriction to $\mathcal{F}$ is equal to P. Here $\rho_{\mathcal{F}}(X) = essmax_{Q \in \mathcal{M}}(E_Q(-X \mid \mathcal{F}) - \alpha(Q))$ means that $\rho_{\mathcal{F}}(X)$ is the *esssup* and that this ess sup is attained for one $Q \in \mathcal{M}$. Before the proof of the theorem the following lemma will be given for the sake of clarity in some parts of the proof.

Lemma 3.4. Let P be a finitely additive set function on $\mathcal{F}$; $P : \mathcal{F} \to [0,1]$ such that $P(\Omega) = 1$. For each $X \in \mathcal{X}$ there is a finitely additive set function Q_X on $\mathcal{G}$ such that the equality

$$E_{Q_X}(\rho_{\mathcal{F}}(X) I_B) = E_{Q_X}(-X I_B) - \sup_{Y \in \mathcal{A}_{\rho_{\mathcal{F}}}} E_{Q_X}(-Y I_B)$$

is satisfied and such that the restriction of Q_X to $\mathcal{F}$ is equal to P.

For the proof of the Lemma (3.4) see [BN04, p:17]

Proof: **(Theorem 3.8) 1.** Firstly it will be verified that for all $X \in \mathcal{X}$, for all B in $\mathcal{F}$ and for all Q absolutely continuous with respect to P, the inequality

$$E_Q(\rho_{\mathcal{F}}(XI_B)) \geq E_Q(-XI_B) - \sup_{Y \in \mathcal{A}_{\rho_{\mathcal{F}}}} E_Q(-YI_B). \qquad (3.5.31)$$

holds.

For all $X \in \mathcal{X}$, $\rho_{\mathcal{F}}(\rho_{\mathcal{F}}(X) + X) = \rho_{\mathcal{F}}(X) - \rho_{\mathcal{F}}(X) = 0$ due to the translation invariance property. So $\rho_{\mathcal{F}}(X) + X \in \mathcal{A}_{\rho_{\mathcal{F}}}$. Then $\sup_{Y \in \mathcal{A}_{\rho_{\mathcal{F}}}} E_Q(-Y) \geq E_Q(-(X + \rho_{\mathcal{F}}(X)))$. Therefore, obviously the equation (3.5.31) holds for all $Q \in \mathcal{M}_{1,f}$ and for all $B \in \mathcal{F}$.

2. As a second step it will be proved that there exist a finitely additive set function Q_X satisfying the equation:

$$E_{Q_X}(\rho_{\mathcal{F}}(XI_B)) = E_{Q_X}(-XI_B) - \sup_{Y \in \mathcal{A}_{\rho_{\mathcal{F}}}} E_Q(-YI_B). \qquad (3.5.32)$$

Showing this for the case $\rho_{\mathcal{F}}(X) = 0$ is satisfactory since when $\rho_{\mathcal{F}}(X) \neq 0$ one can get the result by replacing X with $X + \rho_{\mathcal{F}}(X)$. Therefore showing the equation

$$E_{Q_X}(-XI_B) = \sup_{Y \in \mathcal{A}_{\rho_{\mathcal{F}}}} E_{Q_X}(-YI_B)$$

becomes equivalent to showing (3.5.32).

Consider now the convex hull (see Preliminaries, Def: 2.22) $\tilde{C}$ of $\{(Y - X)I_B; \rho_{\mathcal{F}}(Y) < 0 \ P \ a.s.; B \in \mathcal{F} \ and \ P(B) \neq 0\}$

i) Firstly it will be shown that $\tilde{C}$ does not contain 0. For this aim, assume that there are $\lambda_i \geq 0; \sum_{i=1}^n \lambda_i = 1$ and $\sum_{i=1}^n \lambda_i(Y_i - X)I_{B_i} = 0$. Choose $J \subset \{1, 2, ...n\}$ such that $\tilde{B} = \cap_{i \in J} B_i \neq \emptyset$ and such that $\forall j \in \{1, 2, ...n\} - J$, $\tilde{B} \cap B_j = \emptyset$, then

$$\sum_{i \in J} \lambda_i(Y_i - X)I_{\tilde{B}} = 0 \qquad (*)$$

In fact from its representation the expression $(*)$ above appears like an element

of the convex hull $\tilde{C}$. Now let $\tilde{Y} = \frac{\sum_{i \in J} \lambda_i(Y_i)}{\sum_{i \in J} \lambda_i}$.
It is clear from the convexity of $\rho_{\mathcal{F}}$ that $\rho_{\mathcal{F}}(\tilde{Y}) < 0$. From the expression $(*)$ we have

$$\sum_{i \in J} \lambda_i(Y_i) I_{\tilde{B}} = \sum_{i \in J} \lambda_i(X) I_{\tilde{B}}$$

Therefore we have $\tilde{Y} I_{\tilde{B}} = X I_{\tilde{B}}$ so $\rho_{\mathcal{F}}(\tilde{Y} I_{\tilde{B}}) = \rho_{\mathcal{F}}(X I_{\tilde{B}}) = 0$ which is a contradiction because we found above that $\rho_{\mathcal{F}}(\tilde{Y}) < 0$. Therefore 0 is not contained in $\tilde{C}$.

ii) Now it will be proved that $\tilde{C}$ is non empty.
$\tilde{C}$ contains the open ball

$$B_1(1 - X) = \{Y \in \mathcal{X}; \parallel Y - (1 - X) \parallel < 1\} \qquad (**)$$

For $Y \in B_1(1 - X)$ by setting $Y = Z - X$ in $(**)$ one can get $B_1(1) = \{Z \in \mathcal{X}; \parallel Z - 1 \parallel < 1\}$ indicating that $Z \in B_1(1)$. Therefore Z is positive. Thus, by the second property in Lemma 3.3 $\rho_{\mathcal{F}}(Z) = -Z < 0$. Therefore $\tilde{C}$ is non empty.

From (i), 0 does not belong to $\tilde{C}$, and from (ii) the interior of $\tilde{C}$ is non empty so by the separation theorem (see, Preliminaries Thm: 2.9) there exists a non-zero continuous functional L on $\mathcal{X}$ such that; $0 = L(0) \leq L(Z)$ for all $Z \in \tilde{C}$. Therefore for all Y such that $\rho_{\mathcal{F}}(Y) < 0$ and for all $B \in \mathcal{F}$. $\forall Y \in \mathcal{A}_{\rho_{\mathcal{F}}}, \forall \epsilon > 0, \rho_{\mathcal{F}}(Y + \epsilon) < 0$ by monotonicity. Hence by continuity of L,

$$0 \leq L((Y - X) I_B) \quad \forall Y \in \mathcal{A}_{\rho_{\mathcal{F}}} \qquad (3.5.33)$$

Now $\forall Y \geq 0, \forall \lambda > 0$, $\rho_{\mathcal{F}}(1 + \lambda Y) < 0$. Then $(1 + \lambda Y - X) \in \tilde{C}$ which indicates that $L(1 + \lambda Y - X) = L(1) + \lambda L(Y) - L(X) \geq 0$. This implies that $\forall Y \geq 0, L(Y) \geq 0$. Thus L is a linear functional. Then it follows that $L(1) > 0$. Hence by Theorem 2.4 of Preliminaries, there exists a unique $Q_X \in \mathcal{M}_{1,f}$ defined by $E_{Q_X}(Y) = \frac{L(Y)}{L(1)}$ for all $Y \in \mathcal{X}$. Thus from this equality and (3.5.33)

$$0 \leq E_{Q_X}((Y - X) I_B) \rightarrow E_{Q_X}(-X I_B) \geq E_{Q_X}(-Y I_B) \qquad \forall Y \in \mathcal{A}_{\rho_{\mathcal{F}}}, \forall B \in \mathcal{F} \textbf{(I)}$$

The inequality (3.31) becomes $\sup_{Y\in\mathcal{A}_{\rho_{\mathcal{F}}}} E_Q(-YI_B\rho_{\mathcal{F}}(X) \geq E_Q(-XI_B)$**(II)** for X s.t $\rho_{\mathcal{F}}(X) = 0$. **(I)** together with **(II)** proves that $\forall X \in \mathcal{X}$ there exist a finitely additive set function Q_X satisfying the equation:

$$E_{Q_X}(\rho_{\mathcal{F}}(XI_B)) = E_{Q_X}(-XI_B) - \sup_{Y\in\mathcal{A}_{\rho_{\mathcal{F}}}} E_Q(-YI_B)$$

3. In the third step it will be proved that Q_X is a probability measure i.e. that Q_X is σ-additive.

Let P be a probability measure on $(\Omega, \mathcal{F})$ and $X \in \mathcal{X}$. From Lemma 3.4, there is a finitely additive set function Q_X such that the inequality (3.5.32) holds for all $B \in \mathcal{F}$ and such that the restriction of Q_X to $\mathcal{F}$ is equal to P.

Consider an increasing sequence $(A_n)_{n\in N}$ of elements of $\mathcal{G}$ whose union is equal to Ω. Then it must be shown that $Q_X(A_n)$ converges to 1. Now apply equality (3.5.32) to $B = \Omega$ and get

$$E_P(\rho_{\mathcal{F}}(X)) = E_{Q_X}(-X) - \alpha(Q_X)$$

where $\alpha(Q_X) = \sup_{Y\in\mathcal{A}_{\rho_{\mathcal{F}}}} E_{Q_X}(-Y)$ and X and $\rho_{\mathcal{F}}(X)$ are bounded so $\alpha(Q_X)$ is finite. Let $\lambda > 0$. Then apply inequality (3.5.31) to λI_{A_n}

$$E_{Q_X}(\lambda I_{A_n}) \geq -E_P(\rho_{\mathcal{F}}(\lambda I_{A_n})) - \alpha(Q_X)$$

As n tends to infinity, $\rho_{\mathcal{F}}(\lambda I_{A_n})$ tends to $\rho_{\mathcal{F}}(\lambda) = -\lambda$ so

$$\liminf_{n\to\infty} \lambda E_{Q_X}(I_{A_n}) \geq \lim_{n\to\infty}(-E_P(\rho_{\mathcal{F}}(\lambda I_{A_n})) - \alpha(Q_X))$$

$$\liminf_{n\to\infty} E_{Q_X}(I_{A_n}) \geq 1 - \frac{\alpha(Q_X)}{\lambda}$$

As λ tends to ∞, $\liminf_{n\to\infty} E_{Q_X}[I_{A_n}] = \liminf_{n\to\infty} Q_X(A_n) \geq 1$. Which proves that Q_X is a probability measure.

4. In this final step it will be proved that the probability measure Q_X is absolutely continuous with respect to a priorly given P.

Let A in $\mathcal{F}$ such that $P(A) = 0$. For all $Z \in \tilde{C}$, for all $\beta \in R$, $Z + \beta I_A = Z \;\; P \;\; a.s.$. Apply the inequality $E_{Q_X}(Z) \geq 0$ to $Z + \beta I_A$ for all $\beta \in R$. Then to satisfy the inequality $E_{Q_X}(Z) + \beta E_{Q_X}(I_A) \geq 0$ for all β, $Q_X(A) = 0$ indicating that for $A \in \mathcal{F}$ such that $P(A) = 0$, $Q_X(A) = 0$, which is the definition of absolute continuity. This ends the proof of this step and also the proof of the theorem. □

From now on we will consider the case of **complete uncertainty**, the case where no probability is given. As in the other case, a financial position is described by a bounded map defined on the set Ω of scenarios. $\mathcal{X}$ is the linear space of financial positions, $\mathcal{F}$ represents a σ-algebra on Ω. $\varepsilon_{\mathcal{F}}$ denotes the set of all bounded real valued $(\Omega, \mathcal{F})$ measurable maps. In this case, it is not assumed that a probability measure is given nor that there is any consensus on which $\mathcal{F}$-measurable sets should be null sets. In this situation, a risk measure conditional to the σ-algebra $\mathcal{F}$ is defined as the mapping defined on $\mathcal{X}$ with values in $\varepsilon_{\mathcal{F}}$ which satisfies the conditions of monotonicity, translation invariance and multiplicative invariance for all $X \in \mathcal{X}$.

Definition 3.22. Conditional Risk Measure under Complete Uncertainty
A mapping

$$\rho_{\mathcal{F}} : \mathcal{X} \rightarrow \varepsilon_{\mathcal{F}}$$

is called a risk measure conditional to σ-algebra $\mathcal{F}$ if it satisfies the following conditions:

1. **Monotonicity:** For all $X, Y \in \mathcal{X}$ if $X \leq Y$, then $\rho_{\mathcal{F}}(Y) \leq \rho_{\mathcal{F}}(X)$
2. **Translation Invariance:** For all $Y \in \varepsilon_{\mathcal{F}}$, for all $X \in \mathcal{X}$,

$$\rho_{\mathcal{F}}(X + Y) = \rho_{\mathcal{F}}(X) - Y$$

3. **Multiplicative Invariance:** For all $X \in \mathcal{X}$, for all $A \in \mathcal{F}$,

$$\rho_{\mathcal{F}}(X I_A) = I_A \rho_{\mathcal{F}}(X)$$

Definition 3.23. The acceptance set of a risk measure $\rho_{\mathcal{F}}$ is

$$\mathcal{A}_{\rho_{\mathcal{F}}} = \{X \in \mathcal{X} \mid \rho_{\mathcal{F}}(X) \leq 0\}$$

Proposition 3.15. The acceptance set $\mathcal{A} = \mathcal{A}_{\rho_{\mathcal{F}}}$ of a conditional risk measure $\rho_{\mathcal{F}}$ satisfies the following properties.

1. $\mathcal{A}$ is non empty, closed with respect to the supremum norm and has a
hereditary property : for all $X \in \mathcal{A}$, for all $Y \in \mathcal{X}$, if $Y \geq X$, then $Y \in \mathcal{A}$.

2. Bifurcation property: for all $X_1, X_2 \in \mathcal{A}$ and for all B_1, B_2 disjoint sets $\in \mathcal{F}$, $X = X_1 I_{\{B_1\}} + X_2 I_{\{B_2\}}$ is in $\mathcal{A}$.

3. Positivity: Every $\mathcal{F}$-measurable element of $\mathcal{A}$ is positive.

4. $\rho_{\mathcal{F}}$ can be recovered from $\mathcal{A}$

$$\rho_{\mathcal{F}} = \inf\{Y \in \varepsilon_{\mathcal{F}} \mid X + Y \in \mathcal{A}\}$$

The proof of the proposition is very similar to that of Proposition 3.13 and can be found in [BN04, p:14].

Now we will conclude this section with the representation theorem for the convex conditional risk measures that are continuous from below.

Theorem 3.9. Let $\rho_{\mathcal{F}}$ be a convex risk measure conditional to $\mathcal{F}$. Assume that $\rho_{\mathcal{F}}$ is continuous from below then:

1. For all $X \in \mathcal{X}$ for every probability measure Q on $(\Omega, \mathcal{G})$

$$\rho_{\mathcal{F}}(X) \geq E_Q(-X|\mathcal{F}) - \alpha(Q) \quad Q \quad a.s$$

$$\alpha(Q) = esssup_{\{Y \in \mathcal{A}_{\rho_{\mathcal{F}}}\}}(E_Q(-Y|\mathcal{F})) \quad Q \quad a.s.$$

2. For all $X \in \mathcal{X}$, for every probability measure P on $(\Omega, \mathcal{F})$ there is Q_X in $\mathcal{M}_1(\mathcal{G}, \mathcal{F}, P)$ such that

$$\rho_{\mathcal{F}}(X) = E_{Q_X}(-X|\mathcal{F}) - \alpha(Q_X) \ \ P \ \ a.s.$$

where $\mathcal{M}_1(\mathcal{G}, \mathcal{F}, P)$ is the set of all probability measures Q on $(\Omega, \mathcal{G})$ such that the restriction of Q to $\mathcal{F}$ is equal to P. **3.** For all $X \in \mathcal{X}$, a continuous from below convex conditional risk measure $\rho_{\mathcal{F}}(X)$ can be represented as follows:

$$\begin{aligned}\rho_{\mathcal{F}}(X) = \inf\{g \in \varepsilon_{\mathcal{F}}; \forall Q \in \mathcal{M}_1(\Omega, \mathcal{G}), g \geq (E_Q[-X \mid \mathcal{F}]\\ - esssup_{\{Y \in \mathcal{A}_{\rho_{\mathcal{F}}}\}} E_Q(-Y \mid \mathcal{F})) \ \ Q \ a.s.\}\end{aligned}$$

The theorem can be proved in a similar way with the proof of Theorem 3.8. It can also be found in [BN04, p:18].

Chapter 4

MORE RISK MEASURES

In the previous chapter it is observed that there is no unique definition of risk. Indeed, the concept of risk is defined implicitly by the characterizations of possible ways to measure it. Because of this, coherence axioms become the key defining properties of a consistent risk measure. Although in the literature a number of coherent risk measures has been introduced as alternatives, many practitioners think that coherence axioms belong to some ideal world and insist on using VaR. This is because it is easy to apply and interpret. However this may be dangerous and may cause irreversible losses in non-normal market conditions due to reasons we reviewed in the previous chapter. For this reason, being familiar with coherent measures and learning more of their advantages is an important issue. As a result of this, in this chapter the alternative risk measures of VaR will be discussed and evaluated in terms of their consistency and applicability. An important example for a risk measure of this kind is the Worst Conditional Expectation (WCE) introduced in [ADEH99]. Furthermore Conditional Value at Risk (CVaR) by Uryasev and Rockafeller in [RoU00] and Expected Shortfall(ES) by Acerbi et al. in 2001 [AcNSi01] are other well known examples. In addition, a recently developed measure of risk will also be introduced within this chapter.

4.1 Worst Conditional Expectation

One of the famous coherent risk measures in the literature is the worst conditional expectation and this section will review the definition and some properties of this measure. Suppose the confidence level α is given and we assume that the return of our data follows a particular distribution. Now we can produce a set of possible loss values whose probability assort with the given distribution. We then look for the maximum expected loss incurred in this set of scenarios. This gives us a risk measure. This is the informal definition of the WCE. In the well known study [ADEH99], ADEH prove that this measure is coherent and give the formal definition of the measure as:

Definition 4.1. Worst Conditional Expectation
Let $(\Omega, \mathcal{A}, P)$ be a probability space and X a random variable which represents the random loss of a position on this probability space. Assume $E[X^-] < \infty$. Then

$$WCE_\alpha(X) = -\inf\{E[X|A] : A \in \mathcal{A}, P[A] > \alpha\}$$

is the worst conditional expectation at level α of X.

At first glance, we note that WCE is always finite under the assumption that $E[X^-] < \infty$. As such, let q_α^+ is the upper quantile as defined in equation (3.1.3), then $lim_{t\to\infty} P[X \leq q_\alpha^+ + t] = 1$ implying that there is an event $A = \{X \leq q^+ + t\}$ with $P[A] > 0$. Therefore $E[X|A] < \infty$.

Secondly, from the definition of WCE, the subadditivity of WCE is obvious, i.e. for random variables X and $Y \in \mathcal{X}$

$$WCE_\alpha(X + Y) \leq WCE_\alpha(X) + WCE_\alpha(Y).$$

Moreover, Theorem 6.10 in [D00] says that WCE is the smallest coherent risk measure that dominates VaR. Besides, WCE depends on X through only its distribution in the condition that the probability space is atomless (for the definition of atom see Preliminaries), or **rich** enough.

All properties we enumerated above are very suitable to the concept of coherency and make WCE a theoretically reasonable measure of risk. However, it may not be a practical risk measure since it may fail to operate under the condition that the probability space is not rich enough. In such a situation infimum may not be attained. Thus the coherency of WCE may disappear. This observation might be an incentive for academicians to introduce new measures which are both coherent and applicable. One of such measures is the Conditional Value at Risk.

4.2 Conditional Value at Risk

The conditional value at risk, which is able to quantify the dangers beyond VaR, was first introduced by Rockafeller et. al. in [RoU00]. This measure attracts the interest of researchers and practioners since, as we will see through this section, it is both coherent and applicable in many situations.

Let X be again a random variable on $(\Omega, \mathcal{A}, P)$ representing the loss of a position and α be the given confidence level. The conditional value at risk $CVaR_\alpha$ is defined for the first time in [URY2000] as the solution of an optimization problem

$$CVaR_\alpha(X) := \inf\{a + \frac{1}{1-\alpha} E[X-a]^+\} : a \in R$$

where the function $[Y]^+ = max(0, Y)$. After that, Uryasev and Rockafeller proved that the infimum above is attained by choosing $a = VaR_\alpha(X)$. Then the usual definition of $CVaR_\alpha$ becomes:

Definition 4.2. Conditional Value at Risk
Conditional value at risk of a position X in the given confidence level α is equal to the conditional expectation of X given that $X \geq VaR_\alpha(X)$, i.e.

$$CVaR_\alpha(X) = E[X \mid X \geq VaR_\alpha(X)]$$

Alternative equivalent representations are given in [P00] as:

$$\begin{aligned} CVaR_\alpha(X) &= E[X \mid X \geq F^{-1}(\alpha)] \\ &= \frac{1}{1-\alpha}\int_\alpha^1 F^{-1}(v)dv \\ &= \frac{1}{1-\alpha}\int_{F^{-1}(\alpha)}^\infty udF(u) \end{aligned}$$

where $F(u) = P[X \leq u]$ and $F^{-1}(u)$ is the right continuous inverse of the distribution function F.

It is indicated in [P00] that $CVaR_\alpha$, in the context of the Definition 4.2 satisfies the coherency axioms. In fact this is true only for continuous loss distributions. For distributions with possible discontinuities, which is always the situation in scenario models used in applications, we need another definition for $CVaR_\alpha$ which satisfies the coherence axioms.

Uryasev and Rockafeller derived the representation of $CVaR$ which is coherent in the case of discontinues loss distributions. First of all, in such situations there arise upper and lower CVaR values which are defined in [RoU01] as:

Definition 4.3. $CVaR_\alpha^+$, $CVaR_\alpha^-$
The upper α-CVaR associated to X is the value:

$$CVaR_\alpha^+ = E[X \mid X > VaR_\alpha(X)]$$

whereas the lower α-CVaR of the loss is the value:

$$CVaR_\alpha^- = E[X \mid X \geq VaR_\alpha(X)]$$

In general we have $CVaR^- \leq CVaR \leq CVaR^+$. The equality holds only in the case of continuous loss distributions. In the presence of discreteness $CVaR$ value differs from $CVaR^+$ and $CVaR^-$ values which are lack of coherence. As a small note, the term tail-VaR is suggested for $CVaR^-$ by ADEH in [ADEH99].

Rockafeller and Uryasev gave the representation of $CVaR$ as a weighted sum of VaR and $CVaR_\alpha^+$ in Proposition 6 of [RoU01]. CVaR with this representation is coherent and applicable in general. Proof of coherency axioms can be found in [RoU01].

Proposition 4.1. Conditional Value at Risk satisfies the following decomposition.
For $\lambda = \frac{P[X \leq VaR_\alpha(X)] - \alpha}{1-\alpha} \in [0, 1]$

$$CVaR_\alpha = \lambda VaR_\alpha(X) + (1 - \lambda)CVaR_\alpha^+$$

Besides its coherency, the applicability of CVaR seems to make it one of the most challenging rivals for the VaR measure. As a tool in portfolio optimization, due to its convexity property, CVaR has many preeminent properties. Optimization problems using the CVaR objective suggest unique global solutions. Details of optimization methodology and optimization shortcuts offered by CVaR is available in [P00] and also in [RoU01].

The CVaR is not lacking criticism either. For instance, on the basis of extensive numerical comparison, it is concluded that the CVaR is not consistent with increasing tail-thickness [Pe03].

4.3 Expected Shortfall

The term expected shortfall might be used to define different measures of risk. In fact in the literature there is a confusion on this concept. To make it clear, it is pointed out that this study used this term in the context of its definition in [AcNSi01]. Acerbi et. al. has shown in [AcT02a] that a coherent alternative for VaR arises as the answer to a simple question on a specified sample of worst cases of a distribution. In this section we will see what the question is and how the expected shortfall is constructed as an answer to this question.
From now on, suppose X is the random variable describing the future value of

the profit or loss of a portfolio on some fixed time horizon T from today and $\alpha = A\% \in (0, 1)$ will be a number which represents the sample of worst cases, then the question arising in [AcT02a] is:

What is the expected loss incurred in the $A\%$ worst cases?

In fact, for continuous loss distributions Artzner et. al. had already given an answer to this question by introducing tail-VaR or equivalently the Tail Conditional Expectation.

$$TCE_\alpha(X) = -E\{X \mid X \leq q^\alpha\}$$

where q^α represents the α- quantile value.

This measure is basically the conditional expected value of losses below the VaR value and coherent for continuous distributions. However this statistic fails to be the answer of the question when we consider more general distributions. This occurs because of the fact that for distributions with jumps, the event $\{X \leq q^\alpha\}$ may happen to have a probability larger than $A\%$, thus larger than the pre-specified percentage for the worst cases. So, what should the definition of a measure that answers the question above for more general cases be?
Suppose α is the $(1-(given\ \ confidence\ \ level))$ (for instance if you are working on the 95%confidence level take $\alpha = 1 - 0.95 = 0.05$). Let n be a large number of observations $\{X_i\}_{i=1,\dots,n}$ of the random variable X and $w = [n.\alpha] = max\{m \mid m \leq n.\alpha, m \in N\}$ is an approximate value for the number of $A\%$ worst cases in the sample. Now define the ordered statistic $X_{1:n} \leq ... \leq X_{n:n}$ by sorting the sample in ascending order, then the least w worst cases can be represented by $\{X_{1:n}, ..., X_{w:n}\}$. Assume that $X_{w:n}$ is the naturel estimator for q_n^α. In [AcNSi01] the naturel estimator for the expected loss in $A\%$ worst cases is given by:

$$ES_n^{(\alpha)}(X) = -\frac{\sum_{i=1}^{w} X_{i:n}}{w} \tag{4.3.1}$$

whereas as the naturel estimator of TCE the following expression is given:

$$TCE_n^{(\alpha)}(X) = -\frac{\sum_{i=1}^{n} X_i I_{\{X_i \leq X_{w:n}\}}}{\sum_{i=1}^{n} I_{\{X_i \leq X_{w:n}\}}} \tag{4.3.2}$$

It is clearly seen from the equations above that ES answers the question of 'average of least $A\%$ outcomes'. However TCE represents the 'average of all losses that are greater than or equal to the VaR value'.

The subadditivity of the estimator ES_n can be proved easily, i.e.,

$$\begin{aligned} ES_n^{(\alpha)}(X) &= -\frac{\sum_{i=1}^{w}(X+Y)_{i:n}}{w} \\ &= -\frac{\sum_{i=1}^{w}(X_{i:n}+Y_{i:n})}{w} \\ &= ES_n^{(\alpha)}(X) + ES_n^{(\alpha)}(Y) \end{aligned}$$

Now let us continue with the expansion of the definition of $ES_n^{(\alpha)}$,

$$\begin{aligned} ES_n^{(\alpha)} &= -\frac{\sum_{i=1}^{w} X_{i:n}}{w} = -\frac{\sum_{i=1}^{n} X_{i:n} I_{\{i \leq w\}}}{w} \\ &= -\frac{1}{w}[\sum_{i=1}^{n} X_{i:n} I_{\{X_{i:n} \leq X_{w:n}\}} - \sum_{i=n}^{n} \underbrace{X_{i:n}(I_{\{X_{i:n} \leq X_{w:n}\}} - I_{\{i \leq w\}})}] \end{aligned}$$

The term indicated under the brace indicates X_i values that are equal to the quantile value. Therefore we can rearrange the expression as:

$$-\frac{1}{w}[\sum_{i=1}^{n} X_i I_{\{X_i \leq X_{w:n}\}} - X_{w:n} \sum_{i=n}^{n} (I_{\{X_{i:n} \leq X_{w:n}\}} - I_{\{i \leq w\}})]$$

Dividing and multiplying the expression by n yields

$$-\frac{n}{w}[\frac{1}{n} \sum_{i=1}^{n} X_i I_{\{X_i \leq X_{w:n}\}} - X_{w:n}(\frac{1}{n} \sum_{i=n}^{n} I_{\{X_i \leq X_{w:n}\}} - \frac{w}{n})]$$

If the equality

$$\lim_{n \to \infty} X_{w:n} = q^{\alpha} \tag{4.3.3}$$

holds with probability 1, then

$$\lim_{n\to\infty} ES_n^{(\alpha)}(X) = -\frac{1}{\alpha}(E[XI_{\{X\leq q^\alpha\}}] - q^\alpha(P[X \leq q^\alpha] - \alpha)) \tag{4.3.4}$$

The equation (4.3.3) may not be hold in general since, in the case of non-unique quantiles, the ordered statistic $X_{[n.\alpha]:n}$ may not converge to q^α. So should we conclude that the equation (4.3.4) is not the answer for the question which arose at the beginning of the section. The answer is no since Acerbi et. al. proved the following proposition in [AcT02b]. This proposition indicates that $ES^{(\alpha)}$ is the naturel answer to the question and equation (4.3.4) holds in general.

Proposition 4.2. Let $\alpha \in (0,1)$ be fixed, X a real random variable with $E[X^-] < \infty$ and $\{X_i\}_{\{i=1,\dots,n\}}$ is an independence sequence of random variables with the same distribution as X. Then with probability 1

$$\lim_{n\to\infty} \frac{\sum_{i=1}^{[n\alpha]} X_{i:n}}{[n\alpha]} = -ES^{(\alpha)}(X)$$

By the way Acerbi et. al. gave the definition of the expected shortfall as follows:

Definition 4.4. Expected Shortfall
Let X be the profit loss of a portfolio on a specified time horizon T and let $\alpha = A\% \in (0,1)$ some specified probability level. The $A\%$ expected shortfall of the portfolio is defined as

$$ES^{(\alpha)}(X) = -\frac{1}{\alpha}(E[XI_{\{X\leq q^\alpha\}}] - q^\alpha(P[X \leq q^\alpha] - \alpha)$$

In the definition above, the expression $q^\alpha(P[X \leq q^\alpha] - \alpha)$ has to be interpreted as the exceeding part to be subtracted from the expected value $E[XI_{\{X\leq q^\alpha\}}]$ when $\{X \leq q^\alpha\}$ has a probability larger then $\alpha = A\%$ [AcT02a]. With this definition, the expected shortfall is coherent and the proof of the coherency axioms can be found in [AcT02b]. Besides, expected shortfall is proved to be an insensitive risk measure in terms of changing confidence levels. This is because

ES_α is continuous with respect to the confidence level α. This property is also examined and proved by using an alternative definition of the expected shortfall in the study [AcT02a].

4.4 Distorted Risk Measures

Risk measures we investigated in this section up to now are dealing with the losses exceeding the given α-quantile values. In his study [We02], Wang argues that working only the values below VaR may cause to ignore the useful information in a large part of the loss distribution. Because of this, in the portfolio optimization process these kinds of measures may fail to reflect the risk differentials in alternative strategies. In this section we are going to review the concept of distorted risk measures in general and concentrate on the WT measure which is a specific type of distorted risk measures.

Definition 4.5. Distorted Probability Measure
Let $g : [0, 1] \to [0, 1]$ be an increasing function with $g(0) = 0$ and $g(1) = 1$. For a random variable X with cumulative distribution function F(x), the transform

$$F^*(x) = g(F(x))$$

defines a distorted probability measure where 'g' is called the distortion function.

Remark 4.1. For F and F^* to be equivalent probability measures, $g : [0, 1] \to [0, 1]$ must be continuous and one-to-one.

Definition 4.6. Distorted Risk Measure
Let X be a random variable representing profit or loss of a portfolio with cumulative distribution function F(x) and g be the distortion function, then the expression

$$E^*(X) = -\int_{-\infty}^{0} g(F(x))dx + \int_{0}^{\infty} [1 - g(F(x))]dx$$

defines a distorted risk measure.

The risk measure defined above simply represents the expected value of X under the distorted probability $F^*(x)$.

Remark 4.2. Any differentiable distortion function 'g' gives a coherent risk measure.

Wang recommended the usage of the following special distortion

$$g(u) = \Phi[\Phi^{-1}(u) - \lambda]$$

which is known as Wang Transform [Wang]. Here Φ is the standard normal cumulative distribution. The transformed and empirical cumulative distributions for an observed loss data set is visualized in the *Figure 4.1.* After this,

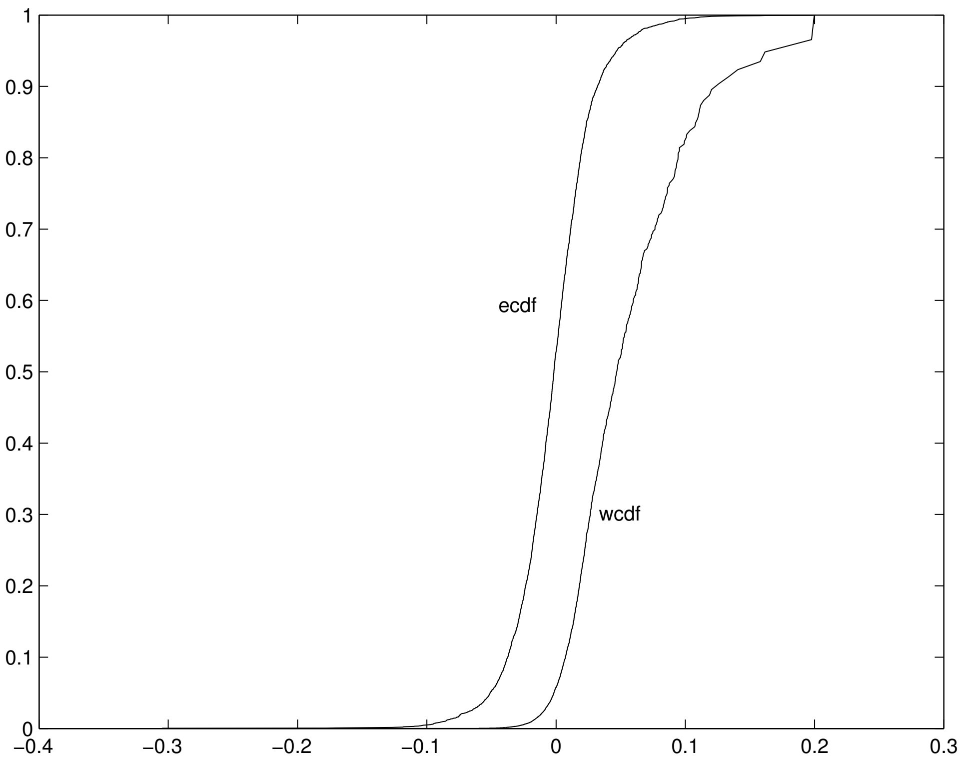

Figure 4.1: Wang CDF vs Empirical CDF

Wang defined the corresponding distorted risk measure which is differentiable and one-to-one, thus coherent.

Definition 4.7. WT-Measure

1. For a pre-specified security level α, let $\lambda = \Phi^{-1}(\alpha)$.

2. Apply the Wang Transform: $F^*(x) = \Phi[\Phi^{-1}(F(x)) - \lambda]$.

3. Set the capital requirement to be the expected value under F^*.

$$WT(\alpha) = E^*[X]$$

Chapter 5

AN IMPLEMENTATION ON ISE-100 INDEX

The main usage of the risk measures given in previous chapters is to determine a risk capital level which would be enough to prevent the position from ending with a high loss in case of significant downside movement in the market. Therefore it is reasonable to investigate the performance of risk measures in terms of their ability to capture market movements. This chapter is mainly constructed for this purpose and includes the performance comparison of VaR, ES and WT measures by employing Turkish Stock Market data. It is known that VaR is the most widely used risk measure in financial markets. As it has been expressed in section (3.1.2) there are various methods to compute VaR. Although it may be impractical in some situations, the Monte Carlo Simulation method is one of the most accepted one. Thus, this method will be included in VaR and ES estimations.

Concussions in financial markets has been evidence that asset prices can display extreme movements beyond those represented by the normal distribution. So the risk measurement techniques such as VaR failed to estimate risk in a correct manner. This happened because of the fact that normal distribution, which is the main underlying block of nearly all VaR methods, is unable to assign reasonable probabilities to extreme events. Indeed, it is well known that returns on financial assets typically exhibit higher than normal kurtosis as expressed in both higher peaks and fatter tails that can be found in normal distribution. In other words the tails of the normal distribution are too thin to capture the extreme losses. At that point, in the literature, one of the solutions suggested for this problem is the Extreme Value Theory. Therefore, in this chapter EVT will be one of the methods used in the estimation of VaR and ES.

Another characteristic of the financial data is that they show volatility clustering. Generalized Autoregressive Conditional Heteroscedasticity (GARCH) models are able to capture this property of financial data. Besides, by using the t-GARCH model we are able to estimate the implied volatility of fatter tailed data. This is a motivation for this study to include the t-GARCH method in the risk estimation process.

As has been expressed, the main aim of this chapter is to compare the relative performances of ES and VaR, which are computed under Monte Carlo Simulation Method, Extreme Value Theory and t-Garch technique, and WT measure for the Turkish Stock Market. This chapter consists of three sections. The first section is on the methodology, the following one includes data analysis and application. Finally, results and comparison section concludes.

5.1 Methodology

At the beginning it is appropriate to emphasize some properties of the financial data series. In particular, it is well known that returns on financial assets typically exhibit higher than normal kurtosis as expressed in both higher peaks and fatter tails. This is because the normal distribution curve assumes complete randomness. However, in real life, when prices are falling investors continue to sell which causes the price to fall faster. Therefore share prices see the extremes a lot quicker than the normal distribution suggests. Hence, assuming normality will result with systematic under estimation of riskiness of portfolio and increase the chance of having a hit. To get ride of such kind of problems, tails of the distribution must be modelled and a well known method in this area is using the Extreme Value Theory (EVT).

There are many kinds of methods for applying EVT. In this study a parametric model based on the Generalized Pareto Distribution (GPD) will be used for modelling the tail of the loss distribution. In GPD there are two parameters: ξ called the shape parameter and σ called the scale parameter of the distribution. The most relevant case for risk management purposes is $\xi > 0$ since in this case GPD is heavy tailed. The interested reader should consult [EmKM97], [E01] and [E00b] for the whole theory of EVT and the extensive list of the references.

In practice, our preliminary aim is to estimate the parameters; to do this firstly we have to choose a sensible threshold level u. However, here there is a dilemma since choosing u is basically a compromise between choosing a sufficiently high u so that the asymptotic theorem can be considered to be essentially exact and choosing a sufficiently low u so that we have enough material to estimate the parameters. Many solutions are suggested in [EmKM97] to this problem. After determining u a second task is to estimating the parameters with an appropriate method. As can be perceived our final aim is to estimate the tail VaR and Expected Shortfall of the finance data by using EVT. In [E00b] the formula for TVaR and ES are given as

$$VaR_p = u + \frac{\sigma}{\xi}((\frac{n}{N_u})^{-\xi} - 1)$$

$$ES_p = VaR_p + \frac{\sigma + \xi(VaR_p - u)}{1 - \xi} = \frac{VaR_p}{1 - \xi} + \frac{\sigma - \xi u}{1 - \xi}$$

Here p is the confidence level, n is the total number of observations, N_u is the number of observations over threshold level and ξ,σ are estimated parameters of GPD. In this study the procedure expressed above is applied step by step and the following section reveals the results of all steps in detail.

The second technique used in VaR and ES estimation is t-GARCH model. This model enables the derivation of conditional volatility estimates. In this study t-GARCH VaR and ES are estimated by applying the following steps.

First of all, the data is tested in terms of its appropriateness to this model. For this aim partial autocorrelations and autocorrelations are examined. The second step is model specification which is made by using the Ljung-Box test. Test results reveal that t-GARCH(1,1) is the most appropriate model for the selected data set. After this step one day ahead implied volatility is estimated by taking window length of 400 days. Finally, estimated parameters are used in VaR and ES computations.

Monte Carlo Simulation method is also used in ES and VaR estimations. About this method detailed information has already given in section (3.1.2). To summarize basically, asset prices are assumed to follow a random walk model and in each trial of simulation 3000 random numbers are drawn from the normal distribution. In this way we simulated a distribution for the one day ahead values of the return data. This distribution was then used to estimate the required risk figures.

Finally the WT measure which was introduced in section (4.4) is computed by using the corresponding Wang Transform formula. Here there is an important point to be careful about: in the WT formula we need the empirical cumulative distribution function of the data set. This task can be successfully achieved in Matlab since there is a command which calculates 'the Kaplan-Meier, or empirical, estimate of the cumulative distribution function ' of the given data set.

After estimating risk measures with corresponding methods, for testing purposes, graphical investigation and statistical tests are applied to the results. This constitutes a criteria to compare the risk measures.

5.2 Data and Application

Since it includes extreme valued data, a volatile market provides a suitable environment to compare the relative performance of different risk modelling. As an emerging market, the Turkish stock market is obviously a good candidate.

ISE became operational in 1985. With the capital account liberalization in 1989 foreign investors are allowed to trade in ISE. From that date ISE-100 index, which consists of the top 100 stocks in terms of transaction volume and liquidity, experiences several severe downturns. And most recent ones are financial crises in 2001 and 11^{th} September.

The data set is the daily closing values of the ISE-100 index from January 2^{nd}, 1990 to August 12^{th}, 2005. It is gathered from the web page of the Central Bank of the Republic of Turkey (www.tcmb.gov.tr), then divided into two parts. The first part is from January, 2^{nd} 1990 to March, 24^{th} 2005 and used as historical data for parameter estimation. The rest of the data is used to test the out of sample forecasting performance of the models. To apply MC, t-GARCH and WT, program files that are generated in Matlab are used. Naturally, in this generating process the theory of these models is considered in all steps to prevent inconsistency. However, GPD application is a little more complicated and necessitates more detailed work. It consists of three main steps:

1. Data analysis,
2. Determination of threshold level,
3. Parameter estimation and VaR, ES computations.

Firstly, in order to get an idea of the distribution of the historical data, kurtosis is computed. For our data set, the corresponding kurtosis value is 8.2671. A kurtosis higher than 3 indicates a deviation from the normal distribution. The higher the kurtosis, the higher the peak of density and higher the tails of the distribution. However, only checking kurtosis is not sufficient to conclude that the data set is heavy tailed. In addition, it has to be investigated graphically. This investigation is done through the QQ-plot of the data set versus normal distribution. If the parametric model fits the data well, the graphs must have the linear form. In other words the more linear the QQ-plot the more appropriate the model in terms of its fitness. Moreover QQ-plot makes it possible to determine how well the selected model fits the tails of the emprical distribution. Thus, the graph makes it possible to compare various estimated models to choose the most appropriate one. In the *Figure 5.1* empirical data is plotted

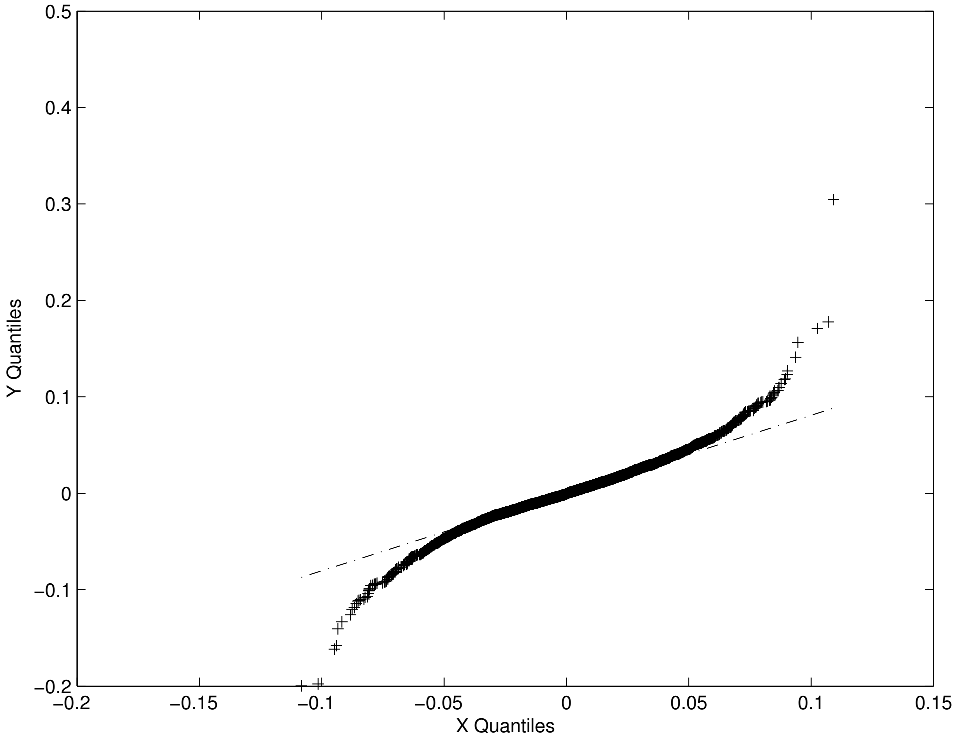

Figure 5.1: QQ-plot(normal distribution)

against normally distributed data set which has equal mean and standard deviation values with the data set. It can be seen that data fits normal distribution at the center but significantly deviates in the tails as expected. There is a curve towards the top of the right end and to the bottom of the left end indicating higher values at the right tail and lower returns at the left, emphasizing that the data is fat tailed.

The second task is determining the threshold level, is a crucial step in extreme value theory applications. The higher the threshold level, the better asymptotic features of the theory works. However, increasing the number of observations means including more data from the center which creates bias in the parameters. On the other hand, choosing a high threshold level makes the estimators more volatile since we are left with fewer observations. In this study two graphical methods are used to estimate threshold level.

Mean Excess Plot : A mean excess function is the sum of the excesses over the threshold level "u" divided by the number of observations which exceeds the threshold u. *Figure 5.2* shows the values of mean excess functions for each threshold level. To determine the threshold level we need to find the interval where the mean excess function takes a positively sloped linear form since positively sloped straight line above a certain threshold u is an indication that the data follows GPD with positive shape parameter ξ. Looking to the *Figure 5.2* threshold level for our data set is between 0.09 and 0.095 since in this interval there is a linear positive trend.

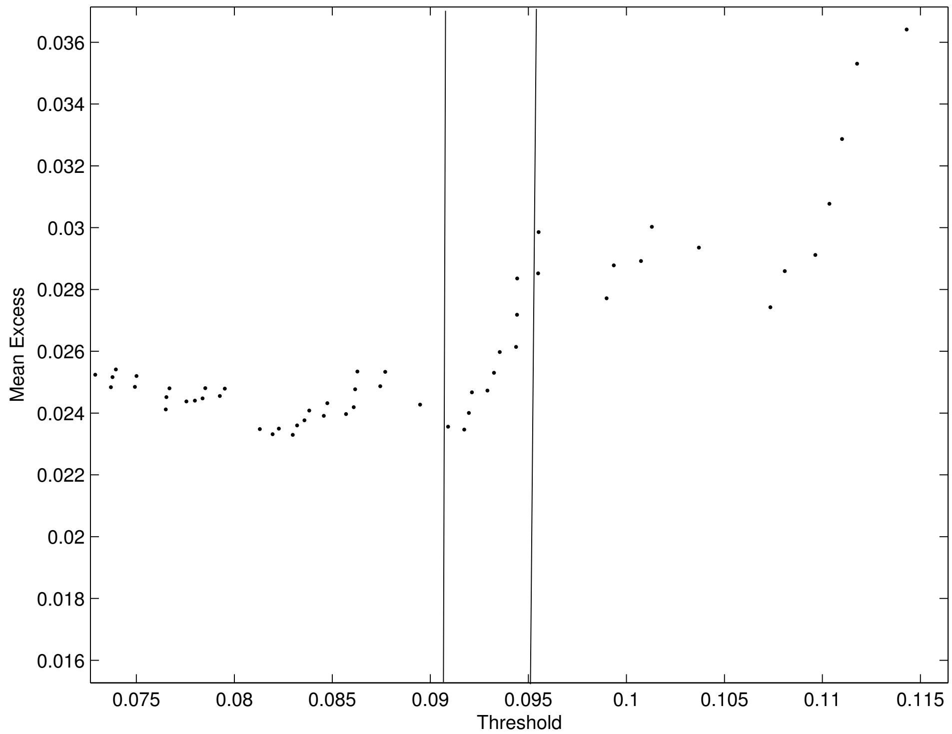

Figure 5.2: Mean Excess Plot

Hill Plot: Hill Estimator is used to compute tail index α.

$$\frac{1}{\hat{\alpha}} = \frac{1}{k}\Sigma_{i=j}^{k} lnX_{j,n} - lnX_{k,n}$$

Where $k \rightarrow \infty$ is the number of exceedances and n is the sample size. Hill plot is composed of Hill estimators for each threshold level u or, in the given context, k. A threshold level is selected from *Figure 5.3* where shape parameter ξ is fairly stable. For the logarithmic returns in the figure threshold level seems to lie in the interval $[8\% - 9\%]$. This corresponds to the number of observations 310-325. Checking the QQ -plots of GPD's constructed for each threshold level, it is seen that 8.7 is the best value for the threshold level.

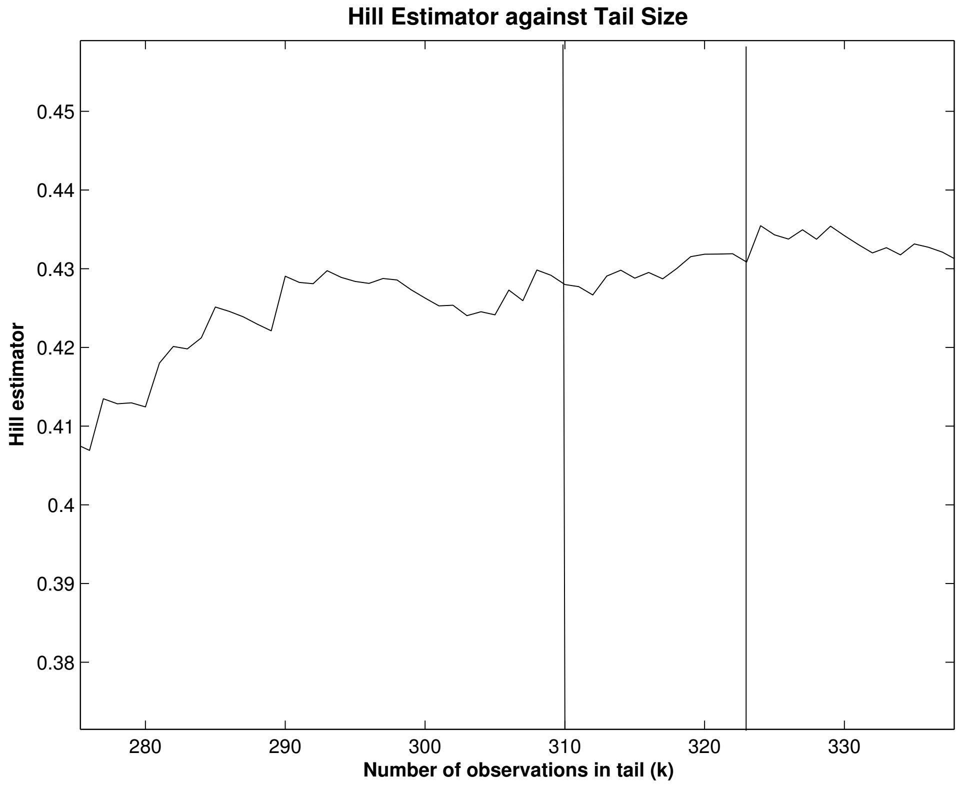

Figure 5.3: Hillplot

Thirdly, parameters are estimated in MATLAB according to the procedure given in the previous section. Shape parameter ξ is estimated as 0.1302 and estimated location parameter σ is equal to 0.0184. Using these parameter estimations VaR and ES are computed for two different confidence levels, 95%, 97%. Afterwards, VaR and ES are estimated by using the other specified methods for the confidence levels 95%, 97%. Results for corresponding VaR and ES values are given in *Table 5.1* and *Table 5.2* respectively.

	95%	97%
EVT	0.0454	0.0559
MC	0.0499	0.0576
t-GARCH(1,1)	0.0418	0.0478

Table 5.1: VaR Results

	95%	97%
EVT	0.0681	0.0802
MC	0.0637	0.0684
t-GARCH(1,1)	0.0524	0.0576

Table 5.2: ES Results

Finally WT measure is estimated by applying the mentioned procedure and for confidence levels 95% , 97% the values 0.0593 and 0.0713 are found respectively.

5.3 Results and Performance Comparison

Values given in the previous section cannot help us to evaluate the methods or to make a relevant comparison. To achieve this goal, forecast power of the models must be compared. For this reason, second part of the data set (24.03.2005-12.08.2005) is used. For each of the given 100 days VaR, ES and WT values are estimated respectively by making one day ahead forecasts. Afterwards, risk levels given by VaR, ES and WT are compared with the observed data. Firstly, results are compared by using graphical representations. Forecasted values for 95% confidence level and observations can be seen in the figures below. First lines on the top of the figures are ES forecasts. The lines in the middle represent the VaR values. In *Figure 5.4* both VaR and ES follow

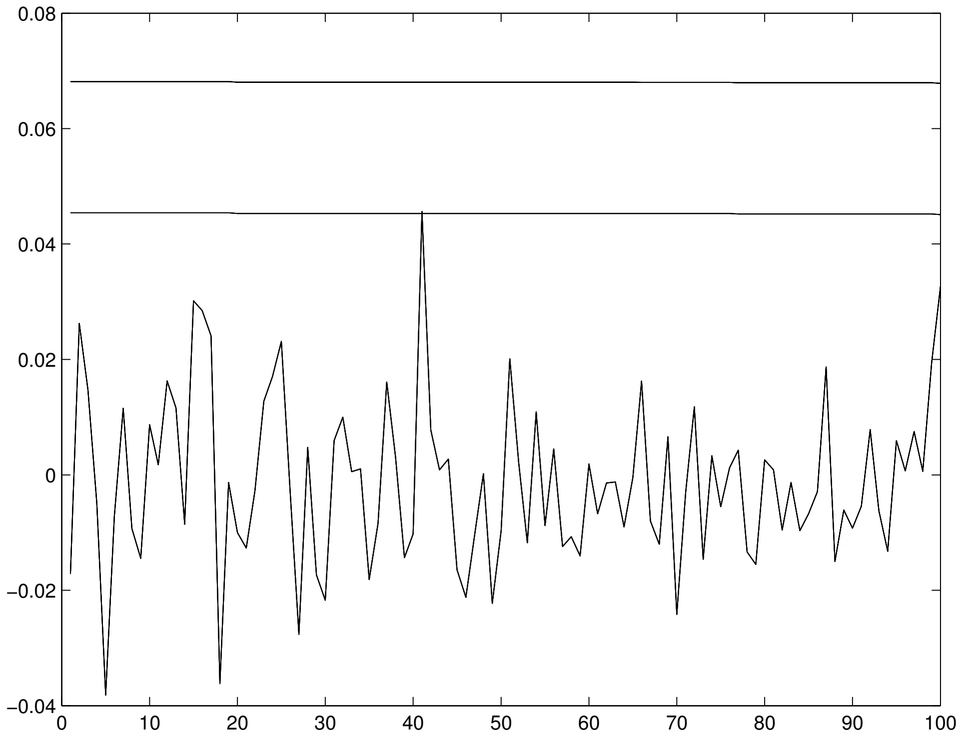

Figure 5.4: GPD VaR, ES Estimates vs Observed Data

stable paths. ES sees no hit but VaR has one hit in the tested period. However, being more conservative ES offers higher risk based capital allocation than VaR. The performance of the t-GARCH technique is demonstrated in *Figure 5.5.* t-GARCH VaR is the best in enveloping the movements of losses. Like VaR, ES has a good performance. Whilst following the path of observed data t-GARCH ES has two and VaR has three hits through the estimation period. In the case of MCS, as illustrated in *Figure 5.6*, both VaR and ES have small fluctuations parallel to the movements of observed data. Moreover they do not have any hits. Finally, Figure 5.7 represents the relative performance of the WT measure for the given period. WT is stable throughout the estimation period and sees no hit. Besides it is nearly as conservative as ES causing the burden of holding too much capital from the perspective of financial institutions.

When we consider the general of these results, we realized that with one method or another, ES is more reliable risk measure than VaR in terms of having

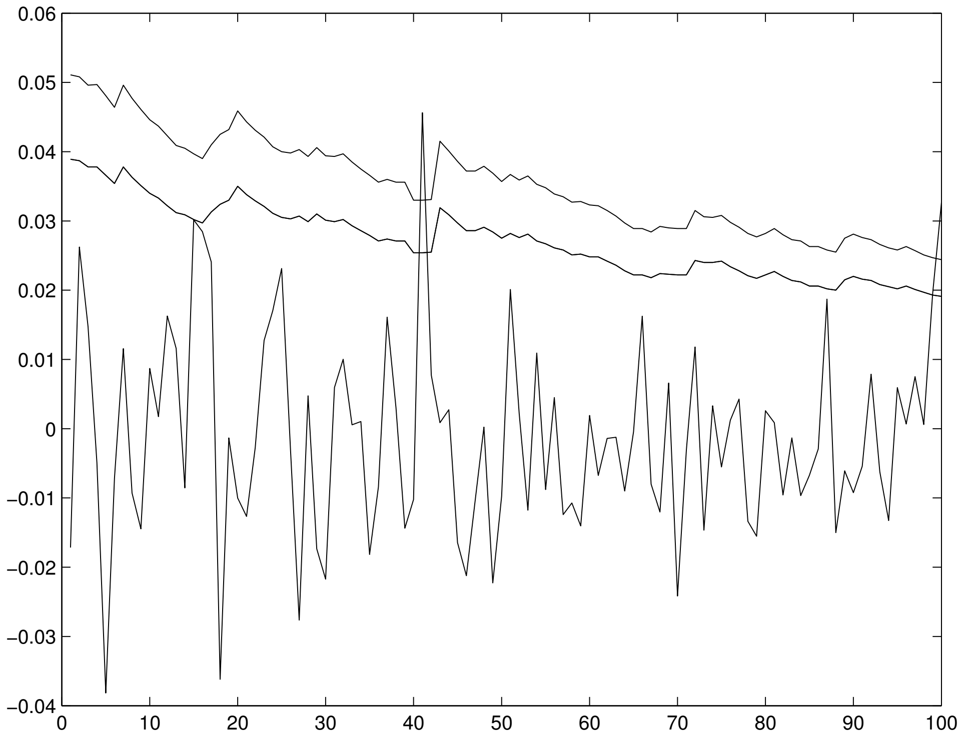

Figure 5.5: t-GARCH VaR, ES Estimates vs Observed Data

smaller number of hits. Although GPD seriously overestimates the risk in terms of both ES and VaR, it must be remembered that the tested period is quite stable and there is no significant movement in losses. Remembering whole past crisis, it is natural for GPD to offer a high risk capital. Of course optimizing risk capital is important but measuring the seriousness of underestimation is more important. There are various backtests dealing with this issue. For instance, Kupeic backtest looks to the frequency of underestimation of risk for the given confidence level (.95). According to these backtest results, the observation frequency of failures in WT is 0, in t-GARCH, 0.0300, in GPD, 0 and finally in MC 0 again for VaR values. When ES is considered, the Kupeic backtest gives 0 for all of the four measures. This result supports the superiority of ES to VaR.

Another backtest concerning the frequency of tail losses is the Lopez backtest. A loss function is defined as a function taking value 1 if observed loss

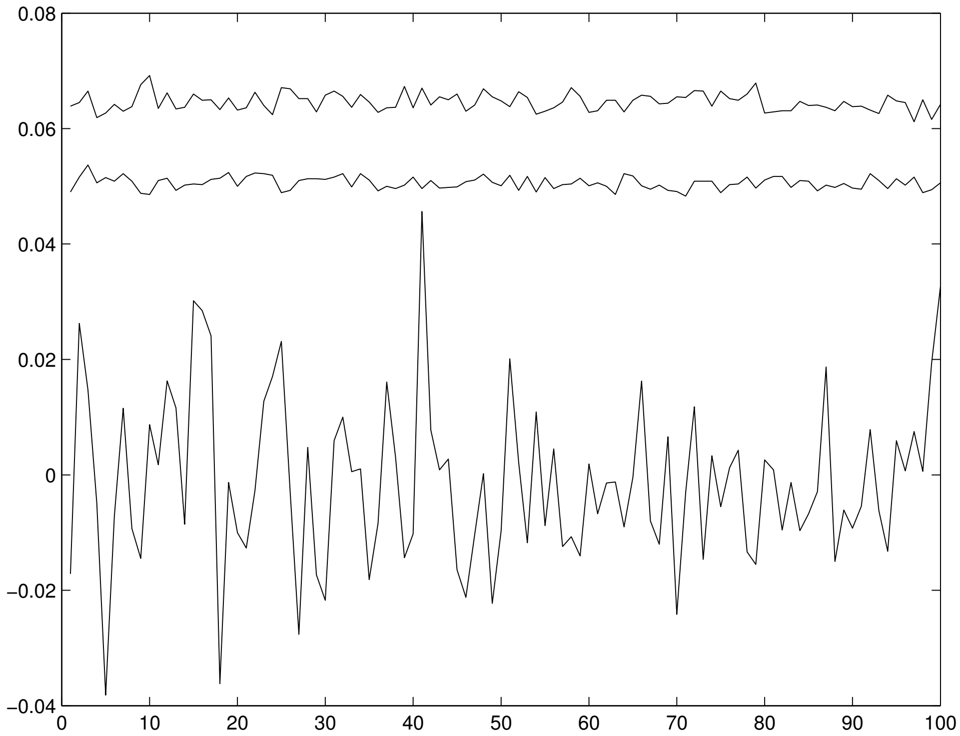

Figure 5.6: MCS VaR, ES Estimates vs Observed Data

exceeds the VaR value and 0 otherwise and then the expected value of this loss function is found. Then according to the expected value, risk measure is scored. The higher the expectation, the lower the score and the less reliable the corresponding risk measure. For WT model this score value is 0.5. If VaR is considered, GPD has the score of 0.32 whereas MC has 0.5. t-GARCH has the lowest score with the value 0.08. When we look at the ES values, GPD and MC methods are in the first rank with score 0.5. t-GARCH technique has again the lowest score, 0.18. Considering these results one can eliminate t-GARCH models from the comparison and be left with the other three methods. Among the three models WT has the advantage that it is easily applicable. From this point of view, GPD is a bit complicated and this complication may contain model risk. Although MC is not complicated it is a time consuming method. When we deal with these measures in terms of their consistency, WT and GPD estimates are more consistent and MC estimates are more volatile. That is to say, while EVT and WT methods suggest banks hold stable amounts, the MC model implies a drastically everyday changing capital. Finally, in the case of multivariate risk sources, EVT method might became painful which is a disadvantage. MC method is superior to others in the case of multivariate asset portfolios. To sum up, deciding which method to use among these three depends on the preferences of the agent. However, according to results, choosing ES rather than VaR would be a sensible decision in risk measurement.

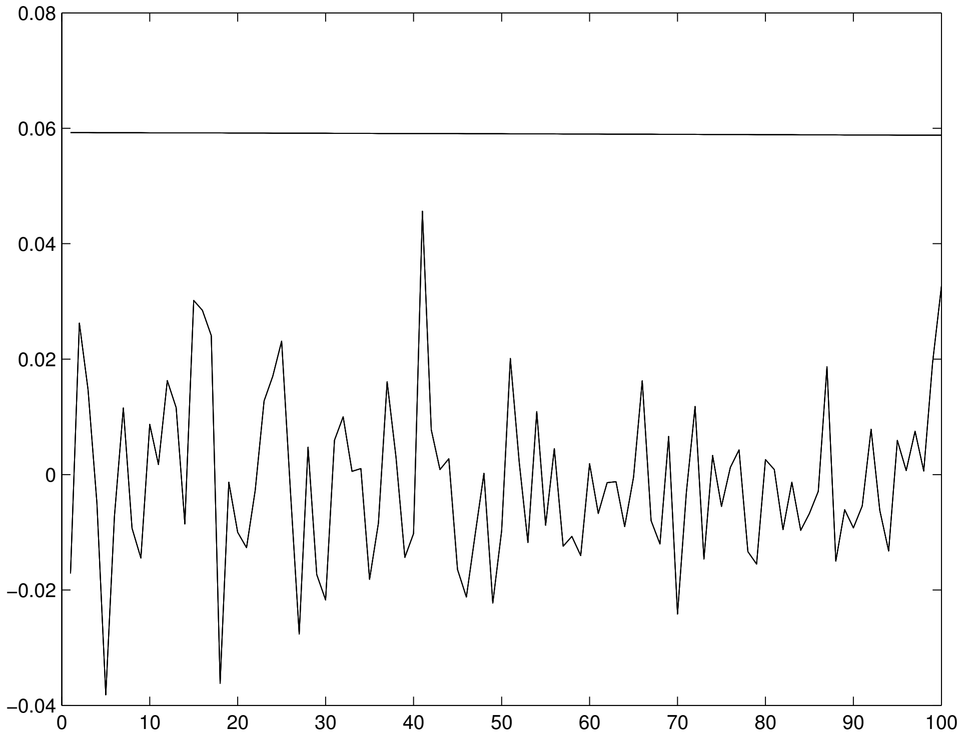

Figure 5.7: WT Estimates vs Observed Data

Chapter 6

CONCLUSION

Risk management has gained popularity in last the decade due to the increasing convulsion in financial markets. In the risk measurement process VaR, as a measure of market risk, is the most popular method in spite of its deficiencies. There is a common view in the market that alternative risk measures are useful in a theoretical setting since the underlying assumptions are unrealistic. However, using VaR may cause dangers. Thus, alternative risk measures should also be considered. In this study, risk measures that are introduced as consistent alternatives to VaR are compared in terms of their theoretical setting and applicability. It is concluded from the application part that Expected Shortfall, one alternative for VaR, is superior to VaR in terms of its capacity to capture risk accurately for the given data set. Besides, WT Measure also provides satisfactory results in risk estimation process. To sum up, in the future, VaR seems to be used only as a benchmark to evaluate the capacity of various consistent risk measures.

References

[AcNSi01] Acerbi C., Nordio C. and Sirtori C., *Expected Shortfall as a tool for Financial Risk Management*, Working Paper, AbaXBank (2001).

[AcT02a] Acerbi C. and Tasche D., *Expected Shortfall: a Natural Coherent Alternative to Value-at-Risk* , Economic Notes of Banca Monte dei Paschi di Siena SpA., 31(2), 1-10 (2002).

[AcT02b] Acerbi C. and Tasche D., *On the Coherence of Expected Shortfall*, Journal of Banking and Finance, 26(7) (2002).

[ADEH97] Artzner Ph., Delbean F., Eber J-M. and Heath D., *Thinking Coherently*, RISK 10, November, 68-71 (1997).

[ADEH99] Artzner Ph., Delbean F., Eber J-M. and Heath D., *Coherent Measures of Risk*, Mathematical Finance, 9, 203-228 (1999)

[ADEHKu02] Artzner Ph., Delbean F., Eber J-M., Heath D. and Ku H., *Multiperiod Risk and Coherent Multiperiod Risk Measurement*, Mimeo, ETH Zürich (2002).

[BaGV99] Barone-Adessi G., Giannopoulos K., *Don't Look Back*, Risk 11 (August) :100-103 (1998).

[BN04] Bion-Nadal J., *Conditional Risk Measure and Robust Representation of Convex Cnditional Risk Measures*, Ecole Polytechnique, Centre de Mathématics Appliquées, R.I. No 557 (2004).

[BRW98] Boudoukh J., Richardson M., Whitelaw R., *The Best of Both Words: A Hybrid Approach to Calculating Value at Risk*, Risk 11(May):64-67 (1998)

[D00] Delbean F., *Coherent Measures of Risk on General Probability Spaces*, Preprint, ETH Zürich (2000).

[D01] Delbean F., *Coherent Risk Measures*, Lecture Notes, Pisa (2001).

[DaM00] Danielsson J. and Morimoto J., *Forecasting Extreme Financial Risk: A Critical Analysis of Practical Methods for the Japanese Market*, Monetary Economic Studies, 12, 25-48 (2000).

[Do02] Dowd K., *An Introduction to Market Risk Measurement*, John Wiley & Sons (2002).

[DuSc58] Dunford N. and Schwartz J., *Linear Operators Vol 1*, Interscience Publishers, New York (1958).

[Dud89] Dudley R., *Real Analysis and Probability* , Wadsworth & Brooks/Cole, California (1989).

[EmKM97] Embrechts P., Klppelberg C., Mikosch T., *Extremal Events in Finance and Insurance*, Springer Verlag, Berlin (1997).

[EMS99] Embrechts P., McNeil A., Strausman D., *Correlation and Dependence in Risk Menagement: Properties*, www.math.ethz.ch/finance (1999).

[E00a] Embrechts P., *Extreme Value Theory: Potential and Limitations as an Integrated Risk Management Tool*, Working Paper (2000).

[E00b] Embrechts P., *Extremes and Integrated Risk Management*, Risk Books and UBS Warburg, London (2000).

[E01] Embrechts P., *Extremes in Economics and Economics of Extremes*, SemStat meeting on Extreme Value Theory and Applications, Gothenburg (2001).

[Fi01] Fischer T., *Examples of Coherent Risk Measures Depending on One-Sided Moments*, Working Paper, TU Darmstadt (2001).

[FrR02] Fritelli M. and Rosazza G., *Putting Order in Risk Measures*, Journal of Banking and Finance, 26, 1473-1486 (2002).

[FS02a] Föllmer H. and Schied A., *Convex Measures of Risk and Trading Constraints*, Finance and Stochastics, 6, 429-447 (2002).

[FS02b] Föllmer H. and Schied A., *Stochastic Finance: An Introduction in Discrete Time*, de Gruyter Studies in Mathematics, Berlin (2002).

[FS02c] Föllmer H. and Schied A., *Robust Representation of Convex Measures of Risk*, Advances in Finance and Stochastics, Essay in Honour of Dieter Sondermann, Springer-Verlag, New York (2002).

[GeSe04] Gençay R. and Selçuk F., *Extreme Value Theory and Value-at-Risk: Relative Performance in Emerging Markets*, International Journal of Forecasting, 20, 287-303 (2004).

[GeSeUlu03] Gençay R., Selçuk F. and Ulugülyağci A., *High Volatility, Thick Tails and Extreme Value Theory in Value-at-Risk Estimation*, Insurance: Mathematics and Economics, 33, 337-356 (2003).

[GHSh00] Glasserman P., Heidlberger P., Shahabuddin P., *Variance Reduction Techniques for Estimating Value at Risk*, Management Science 46: 1349-1364 (2000).

[Hol03] Holton G., *Value-at-Risk: Theory and Practice*, Academic Press (2003).

[Hu81] Huber P., *Robust Statistics*, Wiley, New York (1981).

[HW98] Hull J., White A., *Incorporating volatility updating into the historical simulation method for VaR*, Journal of Risk 1 (Fall):5-19 (1998)

[JP02] Jarrow R. and Purnanandam A., *Generalized Coherent Risk Measures: The Firm's Perspective* (2002).

[K04] Körezlioğlu H., *on Financial Risk Measures (A Quick Survey)*, Working Notes (2004).

[KHa01] Körezlioğlu H. and Hayfavi A., *Elements of Probability Theory*, METU, (2001).

[P00] Pflug G., *Some Remarks on the Value-at-Risk and the Conditional Value-at-Risk in Uryasev, S.*, Probabilistic Constrained Optimization: Methodology and Application, Kluwer Academic Pub. (2000).

[Pe03] Peter A., *Risk Measures*, SonderForschungsBereich 504, No 03-01 (2003).

[Ro70] Rockafeller R. T., *Convex Analysis*, Princeton University Press, New Jersey (1970)

[RoU00] Rockafeller R. T. and Uryasev S., *Optimization of Conditional Value-at-Risk*, The Journal of Risk, 2(3), 21-41 (2000).

[RoU01] Rockafeller R. T. and Uryasev S., *Conditional Value-at-Risk for General Loss Distributions*, Research Report 2001-5, ISE Dept. University of Florida (2001).

[W02] Wang S. S., *A Risk Measure That Goes Beyond Coherence*, 12th AFIR International Colloquium, Mexico (2002).

[We02] West G., *Coherent VaR-Type Measures*, Financial Modelling Agency (2002).