Convex risk measures and duality

October 31, 2018

The purpose of this chapter is to present a possible way to quantify monetary risk. In practice, several different risk measures are used for this purpose. However, we will focus on a class of such measures, called convex risk measures, which satisfy certain economically reasonable properties. In Section 1, we define convex risk measures and some corresponding notions, such as coherent risk measures and acceptance sets. Then, we derive some properties of convex risk measures. In Section 2 we present a brief introduction to convex duality theory. This theory is needed in Section 3, where we prove dual representation theorems for convex risk measures. These theorems provide an alternative characterisation of these measures. In Section 4, we give some examples of measures of risk commonly used in finance.

1 Convex and coherent risk measures

In the literature, there are many different methods for quantifying risk depending on the context and purpose. In this section, we focus on monetary measures of financial risk. We will introduce measures for the risk of a financial position, \( X \), which takes a random value at some set terminal time. This value depends on the current world scenario. An intuitive approach for quantifying risk is the variance. However, the variance does not separate between negative and positive deviations. Hence, it is only suitable as a measure of risk in cases where any kind of deviation from the target is a problem. In situations where deviations in one direction is OK, or even good, while deviations is the other direction is bad, the variance is unsuitable. Clearly, the variance is an unsuitable measure of financial risk. In finance, positive deviations are good (earning more money), but negative deviations are bad (earning less money). In order to resolve this, Artzner et al. [1] set up some economically reasonable axioms that a measure of risk should satisfy and thereby introduced coherent risk measures. This notion has later been modified to so-called convex risk measures.

In order to understand convex risk measures properly, we need some essential concepts from measure theory. We recall these definitions here for completeness, and refer the reader to Shilling [13] for a detailed introduction to measure- and integration theory. Consider a given scenario space \( \Omega \). This may be a finite set \( \Omega = \{ \omega_1, \omega_2, \ldots, \omega_n \} \) or an infinite set. On this space, we can define a \( \sigma \)-algebra \( \mathcal{F} \), i.e. a family of subsets of \( \Omega \) that contains the empty set \( \emptyset \) and is closed under complements and countable unions. The elements in the \( \sigma \)-algebra \( \mathcal{F} \) are called measurable sets. \( (\Omega, \mathcal{F}) \) is then called a measurable space. A measurable function is a function \( f : (\Omega, \mathcal{F}) \rightarrow (\Omega', \mathcal{F}') \) (where \( (\Omega', \mathcal{F}') \) is another measurable space)
such that for any measurable set, $E \in \mathcal{F}'$, the inverse image (preimage), $f^{-1}(E)$, is a measurable set, i.e.,

$$f^{-1}(E) := \{\omega \in \Omega : f(\omega) \in E\} \in \mathcal{F}.$$

A random variable is a real-valued measurable function. On a measurable space $(\Omega, \mathcal{F})$ one can define a measure, i.e., a non-negative countably additive function $\mu : \Omega \to \mathbb{R}$ such that $\mu(\emptyset) = 0$. Then, $(\Omega, \mathcal{F}, \mu)$ is called a measure space. A signed measure is the same as a measure, but without the non-negativity requirement. A probability measure is a measure $P$ such that $P(\Omega) = 1$. Let $\mathcal{P}$ denote the set of all probability measures on $(\Omega, \mathcal{F})$, and $\mathcal{V}$ the set of all measures on $(\Omega, \mathcal{F})$. Then $\mathcal{V}$ is a vector space (also called linear space), and $\mathcal{P} \subseteq \mathcal{V}$ is a convex set (check this yourself as an exercise!).

In the following, let $\Omega$ be a fixed set of scenarios, or possible states of the world. Note that we make no further assumptions on $\Omega$, so in particular, it may be infinite. Consider the measure space $(\Omega, \mathcal{F}, P)$, where $\mathcal{F}$ is a given $\sigma$-algebra on $\Omega$ and $P$ is a given probability measure on $(\Omega, \mathcal{F})$. A financial position (such as a portfolio of stocks) can be described by a mapping $X : \Omega \to \mathbb{R}$, where $X(\omega)$ is the value of the position at the end of the trading period if the state $\omega$ occurs. More formally, $X$ is a random variable. Hence, the dependency of $X$ on $\omega$ describes the uncertainty of the value of the portfolio. Let $\mathcal{X}$ be a given vector space of such random variables $X : \Omega \to \mathbb{R}$, which contains the constant functions. For $c \in \mathbb{R}$, let $c1 \in \mathcal{X}$ denote the constant function $c1(\omega) = c$ for all $\omega \in \Omega$. An example of such a space is

$$L^p(\Omega, \mathcal{F}, P) := \{f : f \text{ is measurable and } \left( \int_{\Omega} |f(\omega)|^p dP(\omega) \right)^{1/p} < \infty \}, \ 1 \leq p \leq \infty.$$  

A convex risk measure is defined as follows:

**Definition 1.1 (Convex risk measure)** A convex risk measure is a function $\rho : \mathcal{X} \to \mathbb{R}$ which satisfies the following for each $X, Y \in \mathcal{X}$:

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

(ii) **(Monotonicity)** If $X \leq Y$, then $\rho(X) \geq \rho(Y)$.

(iii) **(Translation invariance)** If $m \in \mathbb{R}$, then $\rho(X + m1) = \rho(X) - m$.

If $\rho(X) \leq 0$, $X$ is acceptable since it does not have a positive risk. On the other hand, if $\rho(X) > 0$, $X$ is unacceptable.

If a convex risk measure also satisfies positive homogeneity, that is if

$$\lambda \geq 0 \Rightarrow \rho(\lambda X) = \lambda \rho(X)$$

then $\rho$ is called a coherent risk measure. The original definition of a coherent risk measure, did not involve convexity directly, but instead required subadditivity:

**Definition 1.2 (Coherent risk measure)** A coherent risk measure is a function $\pi : \mathcal{X} \to \mathbb{R}$ which satisfies the following for each $X, Y \in \mathcal{X}$:

(i) **(Positive homogeneity)** $\pi(\lambda X) = \lambda \pi(X)$ for $\lambda \geq 0$. 

2
(ii) (Subadditivity) \( \pi(X + Y) \leq \pi(X) + \pi(Y) \).

(iii) (Monotonicity) If \( X \leq Y \), then \( \pi(X) \geq \pi(Y) \).

(iv) (Translation invariance) If \( m \in \mathbb{R} \), then \( \pi(X + m1) = \pi(X) - m \).

We can interpret \( \rho \) as a capital requirement, that is: \( \rho(X) \) is the extra amount of money which should be added to the portfolio in a risk free way to make the position acceptable for an agent.

The conditions in Definition 1.1 are quite natural. The convexity reflects that diversification reduces risk. The total risk of loss in two portfolios should be reduced when the two are weighed into a mixed portfolio. Roughly speaking, spreading your eggs in several baskets should reduce the risk of broken eggs.

Monotonicity says that the downside risk, the risk of loss, is reduced by choosing a portfolio that has a higher value in every possible state of the world.

Finally, translation invariance can be interpreted in the following way: \( \rho \) is the amount of money one needs to add to the portfolio in order to make it acceptable for an agent. Hence, if one adds a risk free amount \( m \) to the portfolio, the capital requirement should be reduced by the same amount.

As mentioned, Artzner et al. [1] originally defined coherent risk measures, that is, they required positive homogeneity. The reason for skipping this requirement in the definition of a convex risk measure is that positive homogeneity means that risk grows linearly with \( X \), and this may not always be the case. In the following, we consider convex risk measures. However, the results can be proved for coherent risk measures as well.

Starting with \( n \) convex risk measures, one can derive more convex risk measures, as in the following theorem. This was proven by Rockafellar in [10], Theorem 3, for coherent risk measures.

**Theorem 1.3** Let \( \rho_1, \rho_2, \ldots, \rho_n \) be convex risk measures.

1. If \( \lambda_1, \lambda_2, \ldots, \lambda_n \geq 0 \) and \( \sum_{i=1}^{n} \lambda_i = 1 \), then \( \rho = \sum_{i=1}^{n} \lambda_i \rho_i \) is a convex risk measure as well.

2. \( \rho = \max\{\rho_1, \rho_2, \ldots, \rho_n\} \) is a convex risk measure.

**Proof.**

1. Let’s check the conditions of Definition 1.1. Obviously, \( \rho : \mathbb{R} \to \mathbb{R} \), so we check for any \( X, Y \in \mathbb{R}, 0 \leq \lambda \leq 1 \):

   (i) : This follows from that a sum of convex functions is also a convex function, and that a positive constant times a convex function is still convex.

   (ii) : If \( X \leq Y \), then \( \rho(X) = \sum_{i=1}^{n} \lambda_i \rho_i(X) \geq \sum_{i=1}^{n} \lambda_i \rho_i(Y) = \rho(Y) \).

   (iii) : If \( m \in \mathbb{R} \),
\[
\rho(X + m) = \sum_{i=1}^{n} \lambda_i \rho_i(X + m)
= \sum_{i=1}^{n} \lambda_i (\rho_i(X) - m)
= \sum_{i=1}^{n} \lambda_i \rho_i(X) - m \sum_{i=1}^{n} \lambda_i
= \rho(X) - m.
\]

2. Again, we check Definition 1.1 for any \(X, Y \in \mathbb{X}\):

(i) \(\lambda \in [0, 1]\),
\[
\rho(\lambda X + (1 - \lambda)Y) = \max\{\rho_1(\lambda X + (1 - \lambda)Y), \ldots, \rho_n(\lambda X + (1 - \lambda)Y)\}
\leq \max\{\lambda \rho_1(X) + (1 - \lambda) \rho_1(Y), \ldots, \lambda \rho_n(X) + (1 - \lambda) \rho_n(Y)\}
\leq \lambda \max\{\rho_1(X), \ldots, \rho_n(X)\} + (1 - \lambda) \max\{\rho_1(Y), \ldots, \rho_n(Y)\}
= \lambda \rho(X) + (1 - \lambda) \rho(Y).
\]

(ii) \(X \preceq Y\),
\[
\rho(X) = \max\{\rho_1(X), \ldots, \rho_n(X)\}
\geq \max\{\rho_1(Y), \ldots, \rho_n(Y)\}
= \rho(Y).
\]

(iii) \(m \in \mathbb{R}\),
\[
\rho(X + m) = \max\{\rho_1(X + m), \ldots, \rho_n(X + m)\}
= \max\{\rho_1(X) - m, \ldots, \rho_n(X) - m\}
= \max\{\rho_1(X), \ldots, \rho_n(X)\} - m
= \rho(X) - m.
\]

\[\square\]

Associated with every convex risk measure \(\rho\), there is a natural set of all acceptable portfolios, called the acceptance set, \(\mathcal{A}_\rho\), of \(\rho\).

**Definition 1.4** (The acceptance set of a convex risk measure, \(\mathcal{A}_\rho\)) A convex risk measure \(\rho\) induces a set
\[
\mathcal{A}_\rho = \{X \in \mathbb{X} : \rho(X) \leq 0\}
\]
The set \(\mathcal{A}_\rho\) is called the acceptance set of \(\rho\).

Conversely, given a class \(\mathcal{A} \subseteq \mathbb{X}\), one can associate a quantitative risk measure \(\rho_{\mathcal{A}}\) to it.
The set of acceptable portfolios

Figure 1: Illustration of the risk measure $\rho_A$ associated with a set $A$ of acceptable portfolios.

**Definition 1.5** Let $A \subseteq \mathcal{X}$ be a set of "acceptable" random variables. This set has an associated measure of risk $\rho_A$ defined as follows: For $X \in \mathcal{X}$, let

$$\rho_A(X) = \inf\{m \in \mathbb{R} : X + m \in A\}. \quad (1)$$

This means that $\rho_A(X)$ measures how much one must add to the portfolio $X$, in a risk free way, to get the portfolio into the set $A$ of acceptable portfolios. This is the same interpretation as for a convex risk measure.

The previous definitions show that one can either start with a risk measure, and derive an acceptance set, or one can start with a set of acceptable random variables, and derive a risk measure.

**Example 1.6** (Illustration of the risk measure $\rho_A$ associated with a set $A$ of acceptable portfolios)

Let $\Omega = \{\omega_1, \omega_2\}$, and let $X : \Omega \to \mathbb{R}$ be a portfolio. Let $x = (X(\omega_1), X(\omega_2))$. If the set of acceptable portfolios is as in Figure 1, the risk measure $\rho_A$ associated with the set $A$ can be illustrated as in the figure.

Based on this theory, a theorem on the relationship between risk measures and acceptable sets can be derived. The following theorem is a version of Proposition 2.2 in Föllmer and Schied [4].

**Theorem 1.7** Let $\rho$ be a convex risk measure with acceptance set $A_\rho$. Then:

(i) $\rho_{A_\rho} = \rho$

(ii) $A_\rho$ is a nonempty, convex set.

(iii) If $X \in A_\rho$ and $Y \in \mathcal{X}$ are such that $X \leq Y$, then $Y \in A_\rho$.

(iv) If $\rho$ is a coherent risk measure, then $A_\rho$ is a convex cone.
Conversely, let $A$ be a nonempty, convex subset of $\mathcal{X}$. Let $A$ be such that if $X \in A$ and $Y \in \mathcal{X}$ satisfy $X \leq Y$, then $Y \in A$. Then, the following holds:

(v) $\rho_A$ is a convex risk measure.

(vi) If $A$ is a convex cone, then $\rho_A$ is a coherent risk measure.

(vii) $A \subseteq A_{\rho_A}$.

Proof.

(i) For any $X \in A_{\rho}$

\[
\rho_{A_{\rho}}(X) = \inf \{ m \in \mathbb{R} : m + X \in A_{\rho} \} = \inf \{ m \in \mathbb{R} : m + X \in \{ Y \in \mathcal{X} : \rho(Y) \leq 0 \} \} = \inf \{ m \in \mathbb{R} : \rho(m + X) \leq 0 \} = \inf \{ m \in \mathbb{R} : \rho(X) - m \leq 0 \} = \inf \{ m \in \mathbb{R} : \rho(X) \leq m \} = \rho(X)
\]

where we have used the definition of a convex risk measure (Definition 1.1) and an acceptance set (Definition 1.4).

(ii) $A_{\rho} \neq \emptyset$ because $X = 0 \in A_{\rho}$. Since $\rho$ is a convex function, $A_{\rho}$ is a convex set.
(iii) Since $X \in \mathcal{A}_\rho$, $\rho(X) \leq 0$, but because $Y \in \mathcal{X}$ is such that $X \leq Y$, $\rho(Y) \leq \rho(X)$ (from the definition of a convex risk measure). Hence

$$\rho(Y) \leq \rho(X) \leq 0.$$  

So $Y \in \mathcal{A}_\rho$ (from the definition of an acceptance set).

(iv) Let $\rho$ be a coherent risk measure, and let $X, Y \in \mathcal{A}_\rho$ and $\alpha, \beta \geq 0$. Then, from the positive homogeneity and subadditivity of coherent risk measures (see Definition 1.2), in addition to the definition of $\mathcal{A}_\rho$

$$\rho(\alpha X + \beta Y) \leq \alpha \rho(X) + \beta \rho(Y) \leq \alpha \cdot 0 + \beta \cdot 0 = 0.$$  

Hence $\alpha X + \beta Y \in \mathcal{A}_\rho$, so $\mathcal{A}_\rho$ is a convex cone (from the definition of a convex cone).

(v) We check Definition 1.1: $\rho_{\mathcal{A}} : \mathcal{X} \rightarrow \mathbb{R}$. Also, for $0 \leq \lambda \leq 1, X, Y \in \mathcal{X}$

$$\rho_{\mathcal{A}}(\lambda X + (1 - \lambda)Y) = \inf\{m \in \mathbb{R} : m + \lambda X + (1 - \lambda)Y \in \mathcal{A}\}$$

$$\leq \lambda \inf\{m \in \mathbb{R} : m + X \in \mathcal{A}\} + (1 - \lambda) \inf\{m \in \mathbb{R} : m + Y \in \mathcal{A}\}$$

$$= \lambda \rho_{\mathcal{A}}(X) + (1 - \lambda) \rho_{\mathcal{A}}(Y)$$  

where the inequality follows because $\lambda \rho_{\mathcal{A}}(X) + (1 - \lambda) \rho_{\mathcal{A}}(Y) = K + L$, is a real number which will make the portfolio become acceptable since

$$(K + L) + (\lambda X + (1 - \lambda)Y) = (K + \lambda X) + (L + (1 - \lambda)Y)$$

$$= \lambda (\inf\{m \in \mathbb{R} : m + X \in \mathcal{A}\} + X) +$$

$$(1 - \lambda)(\inf\{m \in \mathbb{R} : m + Y \in \mathcal{A}\} + Y) \in \mathcal{A}$$

since $\mathcal{A}$ is convex (see Figure 2). In addition, if $X, Y \in \mathcal{X}, X \leq Y$

$$\rho_{\mathcal{A}}(X) = \inf\{m \in \mathbb{R} : m + X \in \mathcal{A}\}$$

$$\geq \inf\{m \in \mathbb{R} : m + Y \in \mathcal{A}\}$$

$$= \rho_{\mathcal{A}}(Y)$$

since $X \leq Y$. Finally, for $k \in \mathbb{R}$ and $X \in \mathcal{X}$

$$\rho_{\mathcal{A}}(X + k) = \inf\{m \in \mathbb{R} : m + X + k \in \mathcal{A}\}$$

$$= \inf\{s - k \in \mathbb{R} : s + X \in \mathcal{A}\}$$

$$= \inf\{s \in \mathbb{R} : s + X \in \mathcal{A}\} - k$$

$$= \rho_{\mathcal{A}}(X) - k.$$  

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

(vi) From (v), all that remains to show is positive homogeneity. For $\alpha > 0$

$$\rho_{\mathcal{A}}(\alpha X) = \inf\{m \in \mathbb{R} : m + \alpha X \in \mathcal{A}\}$$

$$= \inf\{m \in \mathbb{R} : \alpha \left(\frac{m}{\alpha} + X\right) \in \mathcal{A}\}$$

$$= \inf\{m \in \mathbb{R} : \frac{m}{\alpha} + X \in \mathcal{A}\}$$

$$= \inf\{ak \in \mathbb{R} : k + X \in \mathcal{A}\}$$

$$= \alpha \inf\{k \in \mathbb{R} : k + X \in \mathcal{A}\}$$

$$= \alpha \rho_{\mathcal{A}}(X)$$
where we have used that $\mathcal{A}$ is a convex cone in equality number three. Hence, $\rho_A$ is a coherent risk measure.

(vii) Note that $A_{\rho_A} = \{ X \in \mathbb{X} : \rho_A(X) \leq 0 \} = \{ X \in \mathbb{X} : \inf \{ m \in \mathbb{R} : m + X \in A \} \leq 0 \}$.

Let $X \in \mathcal{A}$, then $\inf \{ m \in \mathbb{R} : m + X \in A \} \leq 0$, since $m = 0$ will suffice (because $X \in \mathcal{A}$). Hence $X \in A_{\rho_A}$.

\[ \square \]

We would like to derive a alternative, dual characterisation of convex risk measures. However, in order to do so, we need convex duality theory (also called conjugate duality theory). This theory was first introduced by Rockafellar [9]:

\section{A short introduction to convexity theory and convex duality}

Let $\mathcal{X}$ be a linear space (also called a vector space). A pairing of two linear spaces $\mathcal{X}$ and $\mathcal{V}$ is a real-valued bilinear form $\langle \cdot, \cdot \rangle$ on $\mathcal{X} \times \mathcal{V}$. Assume there is a pairing between the spaces $\mathcal{X}$ and $\mathcal{V}$. A topology on $\mathcal{X}$ is compatible with the pairing if it is a locally convex topology such that the linear function $\langle \cdot, v \rangle$ is continuous, and any continuous linear function on $\mathcal{X}$ can be written in this form for some $v \in \mathcal{V}$. A compatible topology on $\mathcal{V}$ is defined similarly. The spaces $\mathcal{X}$ and $\mathcal{V}$ are paired spaces if there is a pairing between $\mathcal{X}$ and $\mathcal{V}$ and the two spaces have compatible topologies with respect to the pairing. An example is the spaces $\mathcal{X} = L^p(\Omega, F, P)$ and $\mathcal{V} = L^q(\Omega, F, P)$, where $\frac{1}{p} + \frac{1}{q} = 1$. These spaces are paired via the bilinear form $\langle x, v \rangle = \int_{\Omega} x(s)v(s)dP(s)$.

In the following, we will sometimes consider the extended real numbers, $\mathbb{R} := \mathbb{R} \cup \{-\infty, +\infty\}$. When working with the extended real numbers the following computational rules apply: $a - \infty = -\infty$, $a + \infty = \infty$, $\infty + \infty = \infty$, $-\infty - \infty = -\infty$ and $\infty - \infty$ is not defined.

We need some fundamental concepts from convexity theory in order to derive our duality result.

\begin{definition}[Convex function] Let $C \subseteq \mathcal{X}$ be a convex set. A function $f : C \to \mathbb{R}$ is called convex if the inequality

$$ f(\lambda x + (1 - \lambda)y) \leq \lambda f(x) + (1 - \lambda)f(y) \quad (3) $$

holds for all $x, y \in C$ and every $0 \leq \lambda \leq 1$. \end{definition}

See Figure 3 for an illustration of a convex function.

There is an alternative way of defining convex functions, which is based on the notion of epigraph.

\begin{definition}[Epigraph, $\text{epi}(\cdot)$] Let $f : \mathcal{X} \to \mathbb{R}$ be a function. Then the epigraph of $f$ is defined as $\text{epi}(f) = \{(x, \alpha) : x \in \mathcal{X}, \alpha \in \mathbb{R}, \alpha \geq f(x)\}$. \end{definition}
Definition 2.3  (Convex function) Let $A \subseteq \bar{X}$. A function $f : A \to \bar{\mathbb{R}}$ is called convex if the epigraph of $f$ is convex (as a subset of the vector space $\bar{X} \times \mathbb{R}$).

Definitions 2.1 and 2.3 are equivalent if the set $A$ in Definition 2.3 is convex (prove this as an exercise!).
Definition 2.4 (Lower semi-continuity, lsc) Let $A \subseteq \hat{X}$ be a set, and let $f : A \rightarrow \bar{\mathbb{R}}$ be a function. $f$ is called lower semi-continuous, lsc, at a point $x_0 \in A$ if for each $k \in \mathbb{R}$ such that $k < f(x_0)$ there exists a neighborhood $U$ of $x_0$ such that $f(u) > k$ for all $u \in U$. Equivalently: $f$ is lower semi-continuous at $x_0$ if and only if $\liminf_{x \rightarrow x_0} f(x) \geq f(x_0)$.

See Figure 5 for an illustration of a lower semi-continuous function.

Definition 2.5 (Convex hull of a function, co$(f)$) Let $A \subseteq \hat{X}$ be a set, and let $f : A \rightarrow \bar{\mathbb{R}}$ be a function. Then the convex hull of $f$ is the (pointwise) largest convex function $h$ such that $h(x) \leq f(x)$ for all $x \in A$.

Clearly, if $f$ is a convex function co$(f) = f$. One can define the lower semi-continuous hull, lsc$(f)$, of a function $f$ in a similar way.

Definition 2.6 (Closure of a function, cl$(f)$) Let $A \subseteq \hat{X}$ be a set, and let $f : A \rightarrow \bar{\mathbb{R}}$ be a function. We define:

$$\text{cl}(f)(x) = \text{lsc}(f)(x) \quad \text{for all} \quad x \in A \quad \text{if} \quad \text{lsc}(f)(x) > -\infty \quad \forall \quad x \in A.\quad \text{and}$$

$$\text{cl}(f)(x) = -\infty \quad \text{for all} \quad x \in A \quad \text{if} \quad \text{lsc}(f)(x') = -\infty \quad \text{for some} \quad x' \in A.$$  

We say that a function $f$ is closed if $\text{cl}(f) = f$. Hence, $f$ is closed if it is lower semi-continuous and $f(x) > -\infty$ for all $x$ or if $f(x) = -\infty$ for all $x$.

When minimizing a function, the points where it is infinitely large are uninteresting, this motivates the following definitions.

Definition 2.7 (Effective domain, dom$(\cdot)$) Let $A \subseteq \hat{X}$ and let $f : A \rightarrow \bar{\mathbb{R}}$ be a function. The effective domain of $f$ is defined as $\text{dom}(f) = \{x \in A : f(x) < +\infty\}$.

Definition 2.8 (Proper function) Let $A \subseteq \hat{X}$ and let $f : A \rightarrow \bar{\mathbb{R}}$ be a function. $f$ is called proper if $\text{dom}(f) \neq \emptyset$ and $f(x) > -\infty$ for all $x \in A$.

Now, let $\hat{X}$ be paired with another linear space $V$, and $U$ paired with the linear space $\hat{Y}$. The choice of pairings may be important in applications.

Definition 2.9 (Convex conjugate of a function, $f^*$)

Let $\hat{X}$ and $V$ be paired spaces. For a function $f : \hat{X} \rightarrow \bar{\mathbb{R}}$, define the conjugate of $f$, denoted by $f^* : V \rightarrow \bar{\mathbb{R}}$, by

$$f^*(v) = \sup\{\langle x, v \rangle - f(x) : x \in \hat{X}\}. \quad (4)$$

Finding $f^*$ is called taking the conjugate of $f$ in the convex sense. One may also define the conjugate $g^*$ of a function $g : V \rightarrow \bar{\mathbb{R}}$ correspondingly. Similarly, define

Definition 2.10 (Biconjugate of a function, $f^{**}$) Let $\hat{X}$ and $V$ be paired spaces. For a function $f : \hat{X} \rightarrow \bar{\mathbb{R}}$, define the biconjugate of $f$, $f^{**}$, to be the conjugate of $f^*$, so $f^{**}(x) = \sup\{\langle x, v \rangle - f^*(v) : v \in V\}$.  

10
To understand why the conjugate function $f^*$ is important, it is useful to consider it via the epigraph. This is most easily done in $\mathbb{R}^n$, so let $f : \mathbb{R}^n \to \mathbb{R}$ and consider $X = \mathbb{R}^n = V$. From equation (4), it is not difficult to show that

$$(v, b) \in \text{epi}(f^*) \iff b \geq \langle v, x \rangle - a \text{ for all } (x, a) \in \text{epi}(f)$$  \hspace{1cm} (5)
This can also be expressed as

\[(v, b) \in \text{epi}(f^*) \iff l_{v, b} \leq f \]

(6)

where \(l_{v, b}(x) := \langle v, x \rangle - b\). So, since specifying a function on \(\mathbb{R}^n\) is equivalent to specifying its epigraph, equation (6) shows that \(f^*\) describes the family of all affine functions that are majorized by \(f\) (since all affine functions on \(\mathbb{R}^n\) are of the form \(\langle v, x \rangle - b\) for fixed \(v, b\)). But from equation (5)

\[b \geq f^*(v) \iff b \geq l_{x, a}(v) \text{ for all } (x, a) \in \text{epi}(f).\]

This means that \(f^*\) is the pointwise supremum of all affine functions \(l_{x, a}\) for \((x, a) \in \text{epi}(f)\). This is illustrated in Figure 6 and Figure 7. We then have the following important theorem, which we will use later on to prove an alternative characterisation of convex risk measures:

**Theorem 2.11** Let \(f : \bar{X} \to \mathbb{R}\) be arbitrary. Then the conjugate \(f^*\) is a closed, convex function on \(V\) and \(f^{**} = \text{cl(co(f))}\). Similarly if one starts with a function in \(V\). In particular, \(f \mapsto h, h = f^*\) is a one-to-one correspondence between the closed, convex functions on \(\bar{X}\) and the closed, convex functions on \(V\).

**Proof.** By definition \(f^*\) is the pointwise supremum of the continuous, affine functions \(V \mapsto \langle x, v \rangle - \alpha\), where \((x, \alpha) \in \text{epi}(f)\). Therefore, \(f^*\) is convex and lsc, hence it is closed. \((v, \beta) \in \text{epi}(f^*)\) if and only if the continuous affine function \(x \mapsto \langle x, v \rangle - \beta\) satisfies \(f(x) \geq \langle x, v \rangle - \beta\) for all \(x \in \bar{X}\), that is if the epigraph of this affine function contains the epigraph of \(f\). Thus, \(\text{epi}(f^{**})\) is the intersection of all the nonvertical, closed halfspaces in \(\bar{X} \times \mathbb{R}\) containing \(\text{epi}(f)\). This implies, using what a closed, convex set is, that \(f^{**} = \text{cl(co(f))}\). \(\square\)

Theorem 2.11 implies that if \(f\) is convex and closed, then \(f = f^{**}\). This gives a one-to-one correspondence between the closed convex functions on \(\bar{X}\), and the same type of functions on \(V\). Hence, all properties and operations on such functions must have conjugate counterparts, see Rockafellar and Wets [11]).

### 3 A dual characterisation of convex risk measures

Now, we are ready to derive the dual characterisation of a convex risk measure \(\rho\). Therefore, let \(V\) be a vector space that is paired with the vector space \(\mathbb{X}\) of financial positions. For instance, if \(\mathbb{X}\) is given a Hausdorff topology, so that it becomes a topological vector space (for definitions of these terms, see Pedersen [8]), \(V\) can be the set of all continuous linear functionals from \(\mathbb{X}\) into \(\mathbb{R}\), as in Frittelli and Gianin [6]. Using the theory presented in Section 2, a dual characterisation of a convex risk measure \(\rho\) can be derived. The following Theorem 3.1 was originally proved by Frittelli and Gianin [6]. In the following, \(\rho^*\) denotes the conjugate of \(\rho\) in the sense of Definition 2.9, \(\rho^{**}\) is the biconjugate of \(\rho\) as in Definition 2.10, and \(\langle \cdot, \cdot \rangle\) is a pairing.
Theorem 3.1 Let $\rho : \mathbb{X} \to \mathbb{R}$ be a convex risk measure. Assume in addition that $\rho$ is lower semi-continuous. Then $\rho = \rho^{**}$. Hence for each $X \in \mathbb{X}$,

$$\rho(X) = \sup \{ \langle X, v \rangle - \rho^*(v) : v \in V \}$$

where $\langle \cdot, \cdot \rangle$ is a pairing between $\mathbb{X}$ and $V$.

Proof. Since $\rho$ is a convex risk measure, it is a convex function (see Definition 1.1). Hence, the convex hull of $\rho$ is equal to $\rho$, i.e., $\text{co}(\rho) = \rho$ (see Definition 2.5). In addition, since $\rho$ is lower semi-continuous and always greater than $-\infty$, $\rho$ is closed (see comment after Definition 2.6), so $\text{cl}(\rho) = \rho$. Therefore

$$\text{cl}(\text{co}(\rho)) = \text{cl}(\rho) = \rho.$$ 

But Theorem 2.11 says that $\rho^{**} = \text{cl}(\text{co}(\rho))$, hence $\rho = \rho^{**}$.

The second to last equation in the theorem follows directly from the definition of $\rho^{**}$ (Definition 2.10), while the last equation follows because the supremum cannot be achieved when $\rho^* = +\infty$. \qed

3.1 The finite dimensional case

Theorem 3.1 is quite abstract, but by choosing a specific set of paired spaces, $\mathbb{X}$ and $V$, some nice results can be derived. The next theorem is due to Föllmer and Schied [3]. Consider the paired spaces $\mathbb{X} = \mathbb{R}^n, V = \mathbb{R}^n$ with the standard Euclidean inner product, denoted $\cdot$, as pairing. In the following, let $(\Omega, \mathcal{F})$ be a measurable space and let $\mathcal{P}$ denote the set of all probability measures over $\Omega$. 


Figure 7: Affine functions majorized by $f^*$. 

![Figure 7: Affine functions majorized by $f^*$](image-url)
Theorem 3.2 Assume that \( \Omega \) is finite. Then, any convex risk measure \( \rho : X \to \mathbb{R} \) can be represented in the form

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

(7)

where \( E_Q[\cdot] \) denotes the expectation with respect to \( Q \) and \( \alpha : \mathcal{P} \to (-\infty, \infty] \) is a "penalty function" which is convex and closed. Actually, \( \alpha(Q) = \rho^*(-Q) \) for all \( Q \in \mathcal{P} \).

Proof. (Luthi and Doege [7]) To show that \( \rho : X \to \mathbb{R} \) (as in Theorem 3.2) is a convex risk measure we check Definition 1.1: Let \( \lambda \in [0, 1], m \in \mathbb{R}, X, Y \in X \).

(i) :

\[
\rho(\lambda X + (1 - \lambda)Y) = \sup_{Q \in \mathcal{P}} \{ E_Q[-(\lambda X + (1 - \lambda)Y)] - \alpha(Q) \}
\]

\[
= \sup_{Q \in \mathcal{P}} \{ \lambda E_Q[-X] + (1 - \lambda)E_Q[-Y] - \alpha(Q) \}
\]

\[
\leq \lambda \sup_{Q \in \mathcal{P}} \{ E_Q[-X] - \alpha(Q) \}
\]

\[
+ (1 - \lambda) \sup_{Q \in \mathcal{P}} \{ E_Q[-Y] - \alpha(Q) \}
\]

\[
= \lambda \rho(X) + (1 - \lambda) \rho(Y).
\]

(ii) : Assume \( X \leq Y \). Then \(-X \geq -Y\), so

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

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

\[
= \rho(Y).
\]

(iii) :

\[
\rho(X + m1) = \sup_{Q \in \mathcal{P}} \{ E_Q[-(X + m1)] - \alpha(Q) \}
\]

\[
= \sup_{Q \in \mathcal{P}} \{ E_Q[-X] - mE_Q[1] - \alpha(Q) \}
\]

\[
= \sup_{Q \in \mathcal{P}} \{ E_Q[-X] - m - \alpha(Q) \}
\]

\[
= \sup_{Q \in \mathcal{P}} \{ E_Q[-X] - \alpha(Q) \} - m
\]

\[
= \rho(X) - m.
\]

Hence, \( \rho \) is a convex risk measure.

So, assume that \( \rho \) is a convex risk measure. The conjugate function of \( \rho \), denoted \( \rho^* \), is then defined as \( \rho^*(v) = \sup_{X \in X} \{ v \cdot (X - \rho(X)) \} \) (where \( \cdot \) denotes Euclidean inner product) for all \( v \in V = \mathbb{R}^n \). Fix an \( X \in X \) and consider \( Y_m := X + m1 \in X \) for an arbitrary \( m \in \mathbb{R} \). Then

\[
\rho^*(v) \geq \sup_{m \in \mathbb{R}} \{ v \cdot Y_m - \rho(Y_m) \}
\]

14
because \( \{Y_m\}_{m \in \mathbb{R}} \subset \mathcal{X} \). This means that
\[
\rho^*(v) \geq \sup_{m \in \mathbb{R}} \{v \cdot (X + m1) - \rho(X + m1)\}
\]
\[
= \sup_{m \in \mathbb{R}} \{m(v \cdot 1 + 1) + v \cdot X - \rho(X)\}
\]
where the equality follows from the translation invariance of \( \rho \) (see Definition 1.1). The first term in the last expression is only finite if \( v \cdot 1 + 1 = 0 \), (where \( 1 = (1,1, \ldots ,1) \in \mathbb{R}^n \)) i.e. if \( \sum_{i=1}^n v_i = -1 \) (if not, one can make the first term go towards \(+\infty\) by letting \( m \) go towards either \(+\infty\) or \(-\infty\)). It is now proved that in order for \( \rho^*(v) < +\infty \), \( \sum_{i=1}^n v_i = -1 \) must hold.

Again, consider an arbitrary, but fixed \( X \in \mathcal{X}, X \geq 0 \) (here, \( X \geq 0 \) means component-wise). Then, for all \( \rho \in \mathcal{P} \) by the same type of arguments as above

\[
\rho^*(v) = \begin{cases} 
\sup_{X \in \mathcal{X}} \{v \cdot X - \rho(X)\} & \text{where } \rho \cdot 1 = -1 \text{ and } v \leq 0 \\
+\infty & \text{otherwise}.
\end{cases}
\]

Now, define \( \alpha(Q) = \rho^*(-Q) \) for all \( Q \in \mathcal{P} \). From Theorem 3.1, \( \rho = \rho^{**} \). But
\[
\rho^{**}(X) = \sup_{v \in \mathcal{V}} \{v \cdot X - \rho^*(v)\}
\]
\[
= \sup_{Q \in \mathcal{P}} \{(-Q) \cdot X - \alpha(Q)\}
\]
\[
= \sup_{Q \in \mathcal{P}} \left( \sum_{i=1}^n Q_i(-X_i) - \alpha(Q) \right)
\]
\[
= \sup_{Q \in \mathcal{P}} \{E_Q[-X] - \alpha(Q)\}
\]
where \( Q_i, X_i \) denote the \( i \)'th components of the vectors \( Q, X \) respectively. Hence
\[
\rho(X) = \rho^{**}(X) = \sup_{Q \in \mathcal{P}} \{E_Q[-X] - \alpha(Q)\}. \tag*{\Box}
\]

Theorem 3.2 says that any convex risk measure \( \rho : \mathbb{R}^n \to \mathbb{R} \) is the expected value of the negative of a contingent claim, \(-X\), minus a penalty function, \( \alpha(\cdot) \), under the worst case probability. Note that we consider the expectation of \(-X\), not \( X \), since losses are negative in our context.

We already know that the penalty function \( \alpha \) in Theorem 3.2 is of the form \( \alpha(Q) = \rho^*(-Q) \). Actually, Luthi and Doege [7] proved that it is possible to derive a more intuitive representation of \( \alpha \) (see Corollary 2.5 in [7]).

**Theorem 3.3** Let \( \rho : \mathbb{R}^n \to \mathbb{R} \) be a convex risk measure, and let \( \mathcal{A}_\rho \) be its acceptance set (in the sense of Definition 1.4). Then, Theorem 3.2 implies that \( \rho(X) = \sup_{Q \in \mathcal{P}} \{E_Q[-X] - \alpha(Q)\} \), where \( \alpha : \mathcal{P} \to \mathbb{R} \) is a penalty function. Then, \( \alpha \) is of the form
\[
\alpha(Q) = \sup_{X \in \mathcal{A}_\rho} \{E_Q[-X]\}. \tag{15}
\]
Proof. It suffices to prove that for all $Q \in \mathcal{P}$,
$$\rho^*(-Q) = \sup_{X \in X} \{E_Q[-X] - \rho(X)\} = \sup_{X \in A_\rho} \{E_Q[-X]\} \quad (8)$$
since we know that $\alpha(Q) = \rho^*(-Q)$. For all $X \in A_\rho$, $\rho(X) \leq 0$ (see Definition 1.4), so $E_Q[-X] - \rho(X) \geq E_Q[-X]$. Hence, since $A_\rho \subseteq X$
$$\rho^*(-Q) \geq \sup_{X \in A_\rho} \{E_Q[X] - \rho(X)\} \geq \sup_{X \in A_\rho} \{E_Q[-X]\}.$$

To prove the opposite inequality, and hence to prove equation (8), assume for contradiction that there exists $Q \in \mathcal{P}$ such that $\rho^*(-Q) > \sup_{X \in A_\rho} \{E_Q[-X]\}$. From the definition of supremum, there exists a $Y \in X$ such that $E_Q[-Y] - \rho(Y) > E_Q[-X]$ for all $X \in A_\rho$. Note that $Y + \rho(Y) \mathbf{1} \in A_\rho$ since $\rho(Y + \rho(Y) \mathbf{1}) = \rho(Y) - \rho(Y) = 0$. Therefore $E_Q[-Y] - \rho(Y) > E_Q[-(Y + \rho(Y) \mathbf{1})] = E_Q(-Y) + \rho(Y)E_Q[-\mathbf{1}] = E_Q(-Y) - \rho(Y)$, which is a contradiction. Hence, the result is proved. □

Together, Theorem 3.2 and Theorem 3.3 provide a good understanding of convex risk measures in $\mathbb{R}^n$: Any convex risk measure $\rho : \mathbb{R}^n \to \mathbb{R}$ can be written in the form $\rho(X) = \sup_{Q \in \mathcal{P}} \{E_Q[-X] - \alpha(Q)\}$, where $\alpha(Q) = \sup_{X \in A_\rho} \{E_Q[-X]\}$ and $A_\rho$ is the acceptance set of $\rho$.

### 3.2 A little measure theory

We would like to prove a dual representation theorem for convex risk measures in the case where $\Omega$ is infinite as well. In order to do so, we need some concepts from measure theory. For completeness, we include the required definitions here. However, we refer the reader to Shilling [13] for a thorough introduction.

First, we consider the Jordan decomposition theorem, which says that any signed measure can be uniquely decomposed into a positive and a negative part:

**Theorem 3.4 (Jordan decomposition theorem)** Every signed measure $\mu$ has a unique decomposition into a difference:
$$\mu = \mu_+ - \mu_-$$
of two positive measures, $\mu_+$ and $\mu_-$ where at least one of these two measures is finite. We say that $\mu_+$ is the positive part, and $\mu_-$ is the negative part, of $\mu$, respectively.

We also need to define absolutely continuous measures.

**Definition 3.5 (Absolutely continuous measure)** Let $\mu, \nu$ be two measures on the measurable space $(\Omega, \mathcal{F})$. If
$$F \in \mathcal{F}, \mu(F) = 0 \implies \nu(F) = 0,$$
we say that $\nu$ is absolutely continuous w.r.t. $\mu$ and write $\nu \ll \mu$. 

16
We also say that $\mu$ dominates $\nu$. There is a very useful way of representing an absolutely continuous measure via its dominating measure. This representation is called the Radon-Nikodým theorem. In order to state this theorem, we need the concept of $\sigma$-finite measures.

**Definition 3.6 ($\sigma$-finite measure)** Let $\mu$ be a measure on the measurable space $(\Omega, \mathcal{F})$. Then, $\mu$ is called $\sigma$-finite if the set $\Omega$ can be covered with at most countably many measurable sets with finite measure, i.e. there exists $F_1, F_2, \ldots \in \mathcal{F}$ with $\mu(A_i) < \infty$ for all $i \in \mathbb{N}$ such that $\bigcup_{i \in \mathbb{N}} F_i = \Omega$.

Now, we are ready to state the important Radon-Nikodým theorem. For a proof of this theorem, we refer to Shilling [13].

**Theorem 3.7 (Radon-Nikodým theorem)** Let $\mu, \nu$ be two measures on the measurable space $(\Omega, \mathcal{F})$. If $\mu$ is $\sigma$-finite, then the following are equivalent:

- $\nu(F) = \int_F f(x) \mu(dx)$ for some almost everywhere unique, non-negative measurable function $f$.
- $\nu \ll \mu$.

The unique function $f$ is called the Radon-Nikodým derivative and is often denoted by $f = \frac{d\nu}{d\mu}$.

With these results from measure theory at hand, we are ready to generalize the dual representation of convex risk measures to the infinite dimensional case.

### 3.3 The infinite dimensional case

How about infinite-dimensional spaces? Can a similar representation of $\rho$ be derived? This is partially answered in the following Theorem 3.8, which is Theorem 2.2 in Ruszczynski and Shapiro [12], modified slightly to our setting.

First, let’s introduce the setting. Let $(\Omega, \mathcal{F})$ be a measurable space and let $\mathcal{V}$ be the vector space of all finite signed measures on $(\Omega, \mathcal{F})$. For each $v \in \mathcal{V}$ we know that there exists a Jordan decomposition of $v$, so $v = v^+ + v^-$. Let $X : \Omega \to \mathbb{R}$. Also, let $X_+ = \{X \in X : X(\omega) \geq 0 \ \forall \ \omega \in \Omega\}$. This gives a partial order relation on $X$, so for $X, Y \in X$, $X \leq Y$ means that $Y - X \in X_+$.

Let $V \subseteq \mathcal{V}$ be the measures $v \in \mathcal{V}$ such that $\int_{\Omega} |X(\omega)||dv| < +\infty$ for all $X \in X$. $V$ is a vector space because of uniqueness of the Jordan decomposition and linearity of integrals. For example: If $v, w \in V$ then $|v + w| = (v+w)^+ + (v+w)^- = (v^+ + w^+) + (v^- + w^-) = |v| + |w|$ by uniqueness of the decomposition, hence $\int_{\Omega} |X||v + w| = \int_{\Omega} |X||v| + \int_{\Omega} |X||w| < +\infty$. Define the pairing $\langle X, v \rangle = \int_{\Omega} X(\omega)dv(\omega)$. Let $V_- \subseteq V$ be the non-positive measures in $V$ and let $\mathcal{P}$ be the set of probability measures in $V$.

Assume the following:

(A): If $v \notin V_- = \{v \in V : v \leq 0\}$, then there exists an $X' \in X_+$ such that $\langle X', v \rangle > 0$. 

17
Now, let $X$ and $V$ have topologies so that they become paired spaces under the pairing $\langle \cdot, \cdot \rangle$.

For example, let $X = L^p(\Omega, F, P)$ where $P$ is a measure, and let $V$ be as above. Each signed measure $v \in V$ can be decomposed so that $v = v_P + v'$, where $v_P$ is absolutely continuous with respect to $P$ (i.e. $P(A) = 0 \Rightarrow v_P(A) = 0$). Then, $dv_P = MdP$, where $M : \Omega \to \mathbb{R}$ is the Radon-Nikodym density of $v$ w.r.t. $P$. Look at $V' := \{v \in V : \int_\Omega |X(\omega)|d|v_P| < +\infty \} \subseteq V$. This is a vector space for the same reasons that $V$ is a vector space. Then, any signed measure $v \in V'$ can be identified by the Radon-Nikodym derivative of $v_P$ w.r.t. $P$, that is by $M$. Actually, $M \in L^q(\Omega, F, P)$, where $\frac{1}{p} + \frac{1}{q} = 1$, because $\int_\Omega |M|^q dP = \int_\Omega |M|^{q-1}d|v_P| < +\infty$. Hence, each signed measure $v \in V'$ is identified in $L^q$ by its Radon-Nikodym density with respect to $P$.

Note that the pairing defined above actually is the usual bilinear form between $L^p$ and $L^q$ since for $\bar{p} \in L^p$, $\bar{q} \in L^q$

$$\langle \bar{p}, \bar{q} \rangle = \int_\Omega \bar{p}(\omega)\bar{q}(\omega)dP(\omega) = \int_\Omega \bar{p}(\omega)M(\omega)dP(\omega) = \int_\Omega \bar{p}(\omega)d\bar{v}(\omega)$$

(9)

where the second equality follows from that any $\bar{q} \in L^q$ can be viewed as a Radon-Nikodym derivative w.r.t. $P$ for some signed measure $v \in V'$ and the third equality from the definition of a Radon-Nikodym derivative.

In the following theorem monotonicity and translation invariance mean the same as in Definition 1.1.

**Theorem 3.8** Let $X$ be a vector space paired with the space $V$, both of the form above. Let $\rho : X \to \mathbb{R}$ be a proper, lower semi-continuous, convex function. From Theorem 2.11 the following holds: $\rho(X) = \sup\{\langle X, v \rangle - \rho^*(v) : v \in \text{dom}(\rho^*)\}$. Then,

(i) $\rho$ is monotone $\iff$ All $v \in \text{dom}(\rho^*)$ are such that $v \leq 0$

(ii) $\rho$ is translation invariant $\iff$ $v(\Omega) = -1$ for all $v \in \text{dom}(\rho^*)$.

Hence, if $\rho$ is a convex risk measure (so monotonicity and translation invariance hold), then $v \in \text{dom}(\rho^*)$ implies that $Q := -v \in \mathcal{P}$ and

$$\rho(X) = \sup_{Q \in \mathcal{P}} \{\langle X, -Q \rangle - \rho^*(-Q)\} = \sup_{Q \in \mathcal{P}} \{\langle -X, Q \rangle - \rho^*(-Q)\} = \sup_{Q \in \mathcal{P}} \{E_Q[-X] - \alpha(Q)\}$$

where $\alpha(Q) := \rho^*(-Q)$ is a penalty function and the pairing, i.e. the integral, is viewed as an expectation.

**Proof.**
(i) : Assume monotonicity of \( \rho \). We want to show that \( \rho^*(v) = +\infty \) for all \( v \notin V_- \). From assumption (A), \( v \notin V_- \Rightarrow \) there exists \( X' \in X_+ \) such that \( \langle X', v \rangle > 0 \). Take \( X \in \text{dom}(\rho) \), so that \( \rho(X) < +\infty \) and consider \( Y_m := X + mX' \). For \( m \geq 0 \), monotonicity implies that \( \rho(X) \geq \rho(Y_m) \) (since \( Y_m = X + mX' \geq X \) because \( X' \geq 0 \)). Hence

\[
\rho^*(v) \geq \sup_{m \in \mathbb{R}^+} \{\langle Y_m, v \rangle - \rho(Y_m)\}
\]

\[
= \sup_{m \in \mathbb{R}^+} \{\langle X, v \rangle + m\langle X', v \rangle - \rho(X + mX')\}
\]

\[
\geq \sup_{m \in \mathbb{R}^+} \{\langle X, v \rangle + m\langle X', v \rangle - \rho(X)\}
\]

where the last inequality uses the monotonicity. But since \( \langle X', v \rangle > 0 \), by letting \( m \to +\infty \), one gets \( \rho^*(v) = +\infty \) (since \( X \in \text{dom}(\rho) \), so \( \rho(X) < +\infty \), and \( \langle X, v \rangle \) is bounded since \( \langle X, \cdot \rangle \) and \( \langle \cdot, v \rangle \) are bounded linear functionals).

Hence, monotonicity implies that \( \rho^*(v) = +\infty \), unless \( v \leq 0 \), so all \( v \in \text{dom}(\rho^*) \) are such that \( v \geq 0 \).

Conversely, assume that all \( v \in \text{dom}(\rho^*) \) are such that \( v \leq 0 \). Take \( X, Y \in \mathbb{X} \) such that \( Y \leq X \) (i.e. \( X - Y \geq 0 \)). Then \( \langle Y, v \rangle \geq \langle X, v \rangle \) (from the linearity of the pairing). Since \( \rho(X) = \sup_{v \in \text{dom}(\rho^*)} \{\langle X, v \rangle - \rho^*(v)\} \), it follows that \( \rho(X) \leq \rho(Y) \). Hence (i) is proved.

(ii) : Assume translation invariance. Let \( 1 : \Omega \to \mathbb{R} \) denote the random variable constantly equal to 1, so \( 1(\omega) = 1 \) \( \forall \omega \in \Omega \). This random variable is clearly measurable, so \( 1 \in \mathbb{X} \). For \( X \in \text{dom}(\rho) \)

\[
\rho^*(v) \geq \sup_{m \in \mathbb{R}} \{\langle X + m1, v \rangle - \rho(X + m)\}
\]

\[
= \sup_{m \in \mathbb{R}} \{m\langle 1, v \rangle + \langle X, v \rangle - \rho(X) + m\}
\]

\[
= \sup_{m \in \mathbb{R}} \{m(v(\Omega) + 1) + \langle X, v \rangle - \rho(X)\}.
\]

Hence, \( \rho^*(v) = +\infty \), unless \( v(\Omega) = \langle 1, v \rangle = -1 \).

Conversely, if \( v(\Omega) = -1 \), then \( \langle X + m1, v \rangle = \langle X, v \rangle + v(\Omega)m = \langle X, v \rangle - m \).

(where the first equality follows from linearity of the pairing). Hence, translation invariance follows from \( \rho(X) = \sup_{v \in \text{dom}(\rho^*)} \{\langle X, v \rangle - \rho^*(v)\} \).

\[\square\]

Föllmer and Schied [3] proved a version of Theorem 3.8 for \( \mathbb{X} = L^\infty(\Omega, \mathcal{F}, P) \), \( V = L^1(\Omega, \mathcal{F}, P) \). In this case, it is sufficient to assume that the acceptance set \( \mathcal{A}_\rho \) of \( \rho \) is weak*-closed (i.e., closed with respect to the coarsest topology that makes all the linear functionals originating from the inner product, \( \langle \cdot, \cdot \rangle \) continuous) in order to derive a representation of \( \rho \) as above.
4 Two commonly used measures of financial risk

In this section, we present two measures of monetary risk which are frequently used in practice. One of these measures, called value at risk, is not coherent, or even convex in general. One may (rightfully so!) wonder why risk professionals use measures not satisfying the economically reasonable conditions of Definition 1.1 and 1.2. There are several reasons for this:

- **Old habits die hard:** These measures were used before the concepts coherent and convex risk measures were introduced. Hence, people are so used to the old measures that they are hesitant to implement others.

- **Simplicity:** As we will see in Section 4.2, value at risk is a very intuitive concept.

- **Good enough:** In many practical situations, the results attained are sufficient, though the measures in question are economically unreasonable in theory.

In the next subsection, we define value at risk and look at its limitations as a measure of financial risk.

4.1 Value at risk

*Value at risk.* VaR, is the most commonly used risk measure in practice. It is defined as follows: Fix some level \( \lambda \in (0, 1) \). Then,

\[
\text{VaR}_\lambda(X) := \inf\{m \in \mathbb{R} | P(X + m < 0) \leq \lambda\},
\]

i.e., \( \text{VaR}_\lambda(X) \) is the smallest amount of money that needs to be added to \( X \) in order for the probability of a loss to be less than \( \lambda \). From this definition, we see that \( \text{VaR}_\lambda(X) \) is decreasing in \( \lambda \) (if you allow a higher probability of loss, you can add less money to \( X \)).

Note that we can rewrite VaR as

\[
\text{VaR}_\lambda(X) = -\sup\{c \in \mathbb{R} | P(X < c) \leq \lambda\},
\]

i.e., \( -\text{VaR}_\lambda \) is the upper \( \lambda \)-quantile of \( X \) (do this rewriting yourself to make sure you understand it). Note also that from equation (10),

\[
\text{VaR}_\lambda(X) = \inf\{m \in \mathbb{R} | P(\neg X > m) \leq \lambda\},
\]

so VaR can be interpreted as the minimum loss occurring in a presupposed percentage of worst cases.

Value at risk has some major drawbacks as a risk measure:

- In general, \( \text{VaR}_\lambda \) is not convex, see Föllmer and Knispel [2]. This means that diversification may increase the risk w.r.t. VaR, which is economically unreasonable.

- In addition, we see from equation (11) that \( \text{VaR}_\lambda \) ignores extreme losses which occur with small probability. This tail insensitivity makes it an unsuitable measure of risk in situations where the consequences of large losses are very bad.
These drawbacks of VaR as a risk measure is what lead to the development of the theory of convex and coherent risk measures. Nevertheless, value at risk is still widely used in practice, despite its deficiencies.

4.2 Average value at risk

Average value at risk (AVaR), also called expected shortfall (ES) or conditional value at risk (CVaR), was introduced to mend the deficiencies of value at risk. For \( \lambda \in (0, 1] \), the average value at risk is defined as

\[
\text{AVaR}_\lambda(X) := \frac{1}{\lambda} \int_0^\lambda \text{VaR}_\alpha(X)d\alpha.
\]  

(13)

Hence, average value at risk can be interpreted as the expected loss in a presupposed percentage of worst cases. Note that

\[
\text{AVaR}_\lambda(X) \geq \frac{1}{\lambda} \int_0^\lambda \text{VaR}_\lambda(X)d\alpha = \text{VaR}_\lambda(X),
\]

where the first equality follows because \( \text{VaR}_\lambda \) is decreasing in \( \lambda \). So, when considering the same level \( \lambda \), the average value at risk is always greater than or equal the value at risk.

Föllmer and Schied [5] prove that \( \text{AVaR}_\lambda \) is a coherent risk measure, with a dual representation

\[
\text{AVaR}_\lambda(X) = \max_{Q \in \mathcal{Q}_\lambda} E_Q[-X]
\]

where \( \mathcal{Q}_\lambda := \{Q << P|\frac{dQ}{dP} \leq \lambda\} \). That is, \( \mathcal{Q}_\lambda \), is the set of all measures \( Q \) that are absolutely continuous w.r.t. the measure \( P \) given that the Radon-Nikodym derivative of \( Q \) w.r.t. \( P \) is less than or equal \( \lambda \) (see Shilling [13] for more on these measure theoretical concepts). Note also that for \( \lambda = 1 \), average value at risk reduces to \( E_P[-X] \), i.e., the expected loss.

Other examples of convex risk measures are shortfall risk and divergence risk measures, but are beyond the scope of this chapter. We refer the interested reader to Föllmer and Knispel [2].

References


