

Risk measures and robust optimization problems

Lecture notes of a minicourse held at the
8th Symposium on Probability and Stochastic Processes

Universidad de las Américas, Puebla

June 20-25, 2004

Alexander Schied*

Institut für Mathematik, MA 7-4

TU Berlin, Strasse des 17. Juni 136

10623 Berlin, Germany

e-mail: schied@math.tu-berlin.de

First version: November 23, 2004; this version: August 8, 2006

Abstract: These lecture notes give a survey on recent developments in the theory of risk measures. The first part outlines the general representation theory of risk measures in a static one-period setting. In particular, it provides structure theorems for law-invariant risk measures. Examples include Value at Risk, Average Value at Risk, distortion risk measures, and risk measures arising from robust preferences. The second part analyzes risk measures and associated robust optimization problems in the framework of dynamic financial market models. The concept of efficient hedging, as introduced by Föllmer and Leukert (2000), is discussed in terms of the more general framework of convex risk measures. The last two sections are devoted to the construction of optimal investment strategies under Knightian uncertainty.

Contents

Introduction	2
1 Measures of risk: axioms and structure theorems	2
1.1 Risk measures and their acceptance sets	3
1.2 Structure theorems for risk measures on L^∞	6
1.3 Value at Risk	12
1.4 Law-invariant risk measures	17
1.5 Comonotonic law-invariant risk measures	25
1.6 Risk measures arising from robust preferences	27
2 Risk measures and robust optimization in financial markets	31
2.1 Measures of risk in a financial market with convex constraints	31
2.2 Efficient hedging with $AV@R_\lambda$ and other convex risk measures	37
2.3 Optimal investment under Knightian uncertainty	41
2.4 Examples of least favorable measures	47
References	51

*Supported by the DFG Research Center MATHEON “Mathematics for key technologies” (FZT 86).

Introduction

These lecture notes reflect the material presented in a minicourse on risk measures given at the *8th Symposium on Probability and Stochastic Processes* in Puebla. They were also used in a *Cours Bachelier* at the Institut Henri Poincaré, Paris, in 2005. In the first chapter, we discuss the axiomatic structure theory for monetary measures of risk. This theory was initiated by P. Artzner, F. Delbaen, J. Eber, and D. Heath in their seminal paper [3] and further developed by Carlier and Dana [11], Delbaen [20, 21], Föllmer and the author [33, 34, 36, 35], Frittelli and Rosazza Gianin [37, 38], Heath [43], Heath and Ku [44], and Kusuoka [53], to mention only a few. The structure of coherent or, more generally, convex risk measures is also closely connected to the numerical representation of risk-averse preferences under Knightian uncertainty, a topic discussed in Section 1.6 along the lines of Schmeidler [65] and Gilboa and Schmeidler [40]. Most of the material presented in this first chapter is based on the second edition of the author's joint book [35] with Hans Föllmer, and the reader is referred to its bibliographic notes for a more detailed historical account.

In Chapter 2, we discuss risk measures and associated robust optimization problems in the framework of dynamic financial market models. In Section 2.1, we study the effect that hedging has on the acceptability of a financial position. This part is based on [33]. In Section 2.2, we discuss the concept of efficient hedging, as introduced by Föllmer and Leukert [32], in terms of the more general framework of convex risk measures. In particular, we will give short proofs of recent results due to Sekine [66] and the author [61]. In the last two sections of Chapter 2, we will get back to the setting of Section 1.6 and discuss the construction of optimal investment strategies under Knightian uncertainty as considered in [62].

It is a great pleasure to thank the organizers of the 8th SPSP: Mogens Bladt, Antonio González, José Alfredo López-Mimbela, Reyla Navarro, and Juan Ruiz de Chávez. I am also grateful to my co-author Hans Föllmer and our publisher Walter de Gruyter Verlag, Berlin, for the permission to use material from [35] in the preparation of these notes. I furthermore thank the participants of my lecture series in Paris and Puebla for their comments, which helped to improve the first draft of these notes.

1 Measures of risk: axioms and structure theorems

In this chapter, we discuss the problem of quantifying the risk of a financial position, whose profits and losses (P&L) are described by a real-valued function X on some set Ω of possible scenarios ω . The basic asymmetry in the P&L interpretation of X will be taken into account by requiring a property of monotonicity. Convexity of the risk measure will make sure that diversification of portfolios will decrease the overall risk. From the point of view of a supervising agency it is natural to use risk measures in quantifying a capital requirement, i.e., the minimal amount of capital that, if added to the position and invested in a risk-free manner, makes the position acceptable. This monetary interpretation is captured by an axiom of cash invariance. Together with convexity and monotonicity, it singles out the class of convex measures of risk. Under the additional condition of positive homogeneity, we obtain the class of coherent risk measures. This axiomatic approach to monetary risk measures was initiated by P. Artzner, F. Delbaen, J. Eber, and D. Heath [3].

In this chapter, we will develop the structure theory of such risk measures in a static situation without reference to the possible elimination of risk via hedging strategies. In the first two sections, we present the general representation theory on L^∞ . In Section 1.3 we discuss some coherent risk measures related to Value at Risk. These risk measures only involve the P&L distribution with respect to a given probability measure. In Section 1.4, we characterize the class of convex risk measures which share this property of law-invariance. In Section 1.5, the resulting risk measures are characterized by a property of comonotonicity.

1.1 Risk measures and their acceptance sets

Let Ω be a fixed set of scenarios. The P&L of a financial position is described by a mapping $X : \Omega \rightarrow \mathbb{R}$ where $X(\omega)$ is the discounted net worth of the position at the end of the trading period if the scenario $\omega \in \Omega$ is realized. Our aim is to quantify the risk of X by some number $\rho(X)$, where X belongs to a given class \mathcal{X} of financial positions. Throughout this section, \mathcal{X} will be a linear space of bounded functions containing the constants.

Definition 1.1. A mapping $\rho : \mathcal{X} \rightarrow \mathbb{R}$ is called a *monetary measure of risk* if it satisfies the following conditions for all $X, Y \in \mathcal{X}$.

- *Monotonicity:* If $X \leq Y$, then $\rho(X) \geq \rho(Y)$.
- *Cash invariance:* If $m \in \mathbb{R}$, then $\rho(X + m) = \rho(X) - m$.

The financial meaning of monotonicity is clear. Cash invariance is also called *translation invariance*. It is motivated by the interpretation of $\rho(X)$ as a capital requirement: if the amount m is added to the position and invested in a risk-free manner, the capital requirement is reduced by the same amount. In particular, cash invariance implies

$$\rho(X + \rho(X)) = 0, \quad (1)$$

and

$$\rho(m) = \rho(0) - m \quad \text{for all } m \in \mathbb{R}.$$

For most purposes it would be no loss of generality to assume that a given monetary risk measure satisfies the condition of

- *Normalization:* $\rho(0) = 0$.

In some situations, however, it will be convenient not to insist on normalization.

Lemma 1.2. Any monetary measure of risk ρ is Lipschitz continuous with respect to the supremum norm $\|\cdot\|$:

$$|\rho(X) - \rho(Y)| \leq \|X - Y\|.$$

Proof. Clearly, $X \leq Y + \|X - Y\|$, and so $\rho(Y) - \|X - Y\| \leq \rho(X)$ by monotonicity and cash invariance. Reversing the roles of X and Y yields the assertion. \square

From now on our main interest will be on monetary measures of risk that have an additional convexity property.

Definition 1.3. A monetary risk measure $\rho : \mathcal{X} \rightarrow \mathbb{R}$ is called a *convex measure of risk* if it satisfies the property of

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

Consider the collection of possible future outcomes that can be generated with the resources available to an investor: One investment strategy leads to X , while a second strategy leads to Y . If one *diversifies*, spending only the fraction λ of the resources on the first possibility and using the remaining part for the second alternative, one obtains $\lambda X + (1 - \lambda)Y$. Thus, the axiom of convexity gives a precise meaning to the idea that diversification should not increase the risk. If ρ is convex and normalized, then

$$\begin{aligned} \rho(\lambda X) &\leq \lambda\rho(X) \quad \text{for } 0 \leq \lambda \leq 1, \\ \rho(\lambda X) &\geq \lambda\rho(X) \quad \text{for } \lambda \geq 1. \end{aligned}$$

Definition 1.4. A convex measure of risk ρ is called a *coherent measure of risk* if it satisfies the condition of

- *Positive Homogeneity:* If $\lambda \geq 0$, then $\rho(\lambda X) = \lambda\rho(X)$.

If a monetary measure of risk ρ is positively homogeneous, then it is normalized, i.e., $\rho(0) = 0$. Under the assumption of positive homogeneity, convexity is equivalent to

- *Subadditivity:* $\rho(X + Y) \leq \rho(X) + \rho(Y)$.

This property allows to decentralize the management of risk arising from a collection of different positions: If separate risk limits are given to different “desks”, then the risk of the aggregate position is bounded by the sum of the individual risk limits. In many situations, however, risk may grow in a *non-linear* way as the size of the position increases. Therefore, we will not insist on positive homogeneity.

A monetary measure of risk ρ induces an *acceptance set*

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

consisting of positions which are acceptable in the sense that they do not require additional capital. The following two propositions summarize the relations between monetary measures of risk and their acceptance sets. The author is grateful to Patrick Cheridito for remarks on the closure condition, which helped to improve the statements of these results.

Proposition 1.5. *Suppose that ρ is a monetary measure of risk with acceptance set $\mathcal{A} := \mathcal{A}_\rho$.*

- (a) \mathcal{A} is non-empty, closed with respect to the supremum norm $\|\cdot\|$, and satisfies the following two conditions:

$$\inf\{ m \in \mathbb{R} \mid m \in \mathcal{A} \} > -\infty. \quad (2)$$

$$X \in \mathcal{A}, Y \in \mathcal{X}, Y \geq X \implies Y \in \mathcal{A}. \quad (3)$$

- (b) ρ can be recovered from \mathcal{A} :

$$\rho(X) = \inf\{ m \in \mathbb{R} \mid m + X \in \mathcal{A} \}. \quad (4)$$

- (c) ρ is a convex risk measure if and only if \mathcal{A} is convex.

- (d) ρ is positively homogeneous if and only if \mathcal{A} is a cone. In particular, ρ is coherent if and only if \mathcal{A} is a convex cone.

Proof. (a): Closedness follows from Lemma 1.2, the remaining properties in (a) are straightforward.

- (b): Cash invariance implies that for $X \in \mathcal{X}$,

$$\begin{aligned} \inf\{ m \in \mathbb{R} \mid m + X \in \mathcal{A}_\rho \} &= \inf\{ m \in \mathbb{R} \mid \rho(m + X) \leq 0 \} \\ &= \inf\{ m \in \mathbb{R} \mid \rho(X) \leq m \} \\ &= \rho(X). \end{aligned}$$

(c): \mathcal{A} is clearly convex if ρ is a convex measure of risk. The converse will follow from Proposition 1.6 together with (6).

- (d): Clearly, positive homogeneity of ρ implies that \mathcal{A} is a cone. The converse follows as in (c). \square

Conversely, one can take a given set $\mathcal{A} \subset \mathcal{X}$ of acceptable positions as the initial object. For $X \in \mathcal{X}$, we can then define the capital requirement as the minimal amount m for which $m + X$ becomes acceptable:

$$\rho_{\mathcal{A}}(X) := \inf\{ m \in \mathbb{R} \mid m + X \in \mathcal{A} \}. \quad (5)$$

Note that, with this notation, (4) takes the form

$$\rho_{\mathcal{A}_\rho} = \rho. \quad (6)$$

Proposition 1.6. *Assume that \mathcal{A} is a non-empty subset of \mathcal{X} satisfying (2) and (3). Then the functional $\rho_{\mathcal{A}}$ has the following properties:*

- (a) $\rho_{\mathcal{A}}$ is a monetary measure of risk.

(b) If \mathcal{A} is a convex set, then $\rho_{\mathcal{A}}$ is a convex measure of risk.

(c) If \mathcal{A} is a cone, then $\rho_{\mathcal{A}}$ is positively homogeneous. In particular, $\rho_{\mathcal{A}}$ is a coherent measure of risk if \mathcal{A} is a convex cone.

(d) \mathcal{A} is a subset of $\mathcal{A}_{\rho_{\mathcal{A}}}$, and $\mathcal{A} = \mathcal{A}_{\rho_{\mathcal{A}}}$ holds if and only if \mathcal{A} is $\|\cdot\|$ -closed.

Proof. (a): It is straightforward to verify that $\rho_{\mathcal{A}}$ satisfies cash invariance and monotonicity. We show next that $\rho_{\mathcal{A}}$ takes only finite values. To this end, fix some Y in the non-empty set \mathcal{A} . For $X \in \mathcal{X}$ given, there exists a finite number m with $m + X > Y$, for X and Y are both bounded. Then

$$\rho_{\mathcal{A}}(X) - m = \rho_{\mathcal{A}}(m + X) \leq \rho_{\mathcal{A}}(Y) \leq 0,$$

and hence $\rho_{\mathcal{A}}(X) \leq m < \infty$. Note that (2) is equivalent to $\rho_{\mathcal{A}}(0) > -\infty$. To show that $\rho_{\mathcal{A}}(X) > -\infty$ for arbitrary $X \in \mathcal{X}$, we take m' such that $X + m' \leq 0$ and conclude by monotonicity and cash invariance that $\rho_{\mathcal{A}}(X) \geq \rho_{\mathcal{A}}(0) + m' > -\infty$.

(b): Suppose that $X_1, X_2 \in \mathcal{X}$ and that $m_1, m_2 \in \mathbb{R}$ are such that $m_i + X_i \in \mathcal{A}$. If $\lambda \in [0, 1]$, then the convexity of \mathcal{A} implies that $\lambda(m_1 + X_1) + (1 - \lambda)(m_2 + X_2) \in \mathcal{A}$. Thus, by the cash invariance of $\rho_{\mathcal{A}}$,

$$\begin{aligned} 0 &\geq \rho_{\mathcal{A}}(\lambda(m_1 + X_1) + (1 - \lambda)(m_2 + X_2)) \\ &= \rho_{\mathcal{A}}(\lambda X_1 + (1 - \lambda)X_2) - (\lambda m_1 + (1 - \lambda)m_2), \end{aligned}$$

and the convexity of $\rho_{\mathcal{A}}$ follows.

(c): As in the proof of convexity, we obtain that $\rho_{\mathcal{A}}(\lambda X) \leq \lambda \rho_{\mathcal{A}}(X)$ for $\lambda \geq 0$ if \mathcal{A} is a cone. To prove the converse inequality, let $m < \rho_{\mathcal{A}}(X)$. Then $m + X \notin \mathcal{A}$ and hence $\lambda m + \lambda X \notin \mathcal{A}$ for $\lambda \geq 0$. Thus $\lambda m < \rho_{\mathcal{A}}(\lambda X)$, and (c) follows.

(d): The inclusion $\mathcal{A} \subseteq \mathcal{A}_{\rho_{\mathcal{A}}}$ is obvious, and Proposition 1.5 implies that \mathcal{A} is $\|\cdot\|$ -closed as soon as $\mathcal{A} = \mathcal{A}_{\rho_{\mathcal{A}}}$. Conversely, assume that \mathcal{A} is $\|\cdot\|$ -closed. We have to show that $X \notin \mathcal{A}$ implies that $\rho_{\mathcal{A}}(X) > 0$. To this end, take $m > \|X\|$. Since \mathcal{A} is $\|\cdot\|$ -closed and $X \notin \mathcal{A}$, there is some $\lambda \in (0, 1)$ such that $\lambda m + (1 - \lambda)X \notin \mathcal{A}$. Thus,

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

Since $\rho_{\mathcal{A}}$ is a monetary measure of risk, Lemma 1.2 shows that

$$|\rho_{\mathcal{A}}((1 - \lambda)X) - \rho_{\mathcal{A}}(X)| \leq \lambda \|X\|.$$

Hence,

$$\rho_{\mathcal{A}}(X) \geq \rho_{\mathcal{A}}((1 - \lambda)X) - \lambda \|X\| \geq \lambda (m - \|X\|) > 0. \quad \square$$

In the following examples, we take \mathcal{X} as the linear space of all bounded measurable functions on some measurable space (Ω, \mathcal{F}) , and we denote by $\mathcal{M}_1 = \mathcal{M}_1(\Omega, \mathcal{F})$ the space of all probability measures on (Ω, \mathcal{F}) .

Example 1.7. Consider the *worst-case risk measure* ρ_{\max} defined by

$$\rho_{\max}(X) = - \inf_{\omega \in \Omega} X(\omega) \quad \text{for all } X \in \mathcal{X}.$$

The value $\rho_{\max}(X)$ is the least upper bound for the potential loss which can occur in any scenario. The corresponding acceptance set \mathcal{A} is given by the convex cone of all non-negative functions in \mathcal{X} . Thus, ρ_{\max} is a coherent measure of risk. It is the most conservative measure of risk in the sense that any normalized monetary risk measure ρ on \mathcal{X} satisfies

$$\rho(X) \leq \rho\left(\inf_{\omega \in \Omega} X(\omega)\right) = \rho_{\max}(X).$$

Note that ρ_{\max} can be represented in the form

$$\rho_{\max}(X) = \sup_{Q \in \mathcal{Q}} E_Q[-X], \quad (7)$$

where \mathcal{Q} is the class \mathcal{M}_1 of all probability measures on (Ω, \mathcal{F}) . ◇

Example 1.8. Consider a *utility function* $u : \mathbb{R} \rightarrow \mathbb{R}$, that is, u is concave and strictly increasing. If $Q \in \mathcal{M}_1$ is a probability measure, then we can consider the *expected utility* $E_Q[u(X)]$ of $X \in \mathcal{X}$ and define the corresponding *certainty equivalent* as the number $u^{-1}(E_Q[u(X)])$. This notion gives rise to a convex risk measure as follows: Fix some threshold $c \in \mathbb{R}$ and let us call a position X acceptable if its certainty equivalent is at least c , i.e., if its expected utility $E_Q[u(X)]$ is bounded from below by $u(c)$. Clearly, the set

$$\mathcal{A} := \{ X \in \mathcal{X} \mid E_Q[u(X)] \geq u(c) \}.$$

is non-empty, convex, and satisfies (2) and (3). Thus, $\rho_{\mathcal{A}}$ is a convex measure of risk. As an obvious robust extension, we can define acceptability in terms of a whole class \mathcal{Q} of probability measures on (Ω, \mathcal{F}) , i.e.,

$$\mathcal{A} := \bigcap_{Q \in \mathcal{Q}} \{ X \in \mathcal{X} \mid E_Q[u(X)] \geq u(c_Q) \},$$

with constants c_Q such that $\sup_{Q \in \mathcal{Q}} c_Q < \infty$. \diamond

Example 1.9. Suppose now that we have specified a probability measure P on (Ω, \mathcal{F}) . the distribution of $X \in \mathcal{X}$ under P is sometimes called the *profit-and-loss* or *P&L distribution*. In this context, X can be defined as being acceptable if the probability of a loss is bounded by a given level $\lambda \in (0, 1)$, i.e., if

$$P[X < 0] \leq \lambda.$$

The corresponding monetary risk measure $V@R_{\lambda}$, defined by

$$V@R_{\lambda}(X) = \inf \{ m \in \mathbb{R} \mid P[m + X < 0] \leq \lambda \},$$

is called *Value at Risk* at level λ . Note that it is well defined on the space $\mathcal{L}^0(\Omega, \mathcal{F}, P)$ of all random variables which are P -a.s. finite, and that

$$V@R_{\lambda}(X) = E[-X] + \Phi^{-1}(1 - \lambda)\sigma(X), \quad (8)$$

if X is a Gaussian random variable with variance $\sigma^2(X)$ and Φ^{-1} denotes the inverse of the distribution function Φ of $N(0, 1)$. Clearly, $V@R_{\lambda}$ is positively homogeneous, but in general it is not convex, as shown by Example 1.26 below. \diamond

1.2 Structure theorems for risk measures on L^{∞}

From now on we will fix a probability measure P on (Ω, \mathcal{F}) and consider risk measures ρ such that

$$\rho(X) = \rho(Y) \quad \text{if } X = Y \text{ } P\text{-a.s.}, \quad (9)$$

and we are interested in general structure theorems for such risk measures on $\mathcal{X} := L^{\infty} := L^{\infty}(\Omega, \mathcal{F}, P)$. This representation theory was developed by Delbaen [20, 21] in the coherent setting and extended to the convex case by Föllmer and Schied [33, 34, 35], Frittelli and Rosazza Gianin [37], and others.

To this end, let us introduce the notation $\mathcal{M}_1(P) := \mathcal{M}_1(\Omega, \mathcal{F}, P)$ for the set of all probability measures Q on (Ω, \mathcal{F}) which are absolutely continuous with respect to P . More generally, $\mathcal{M}_{1,f}(P) := \mathcal{M}_{1,f}(\Omega, \mathcal{F}, P)$ will denote the set of all *finitely additive* set functions $Q : \mathcal{F} \rightarrow [0, 1]$ which are normalized to $Q[\Omega] = 1$ and absolutely continuous with respect to P in the sense that $Q[A] = 0$ if $P[A] = 0$. By $E_Q[X]$ we denote the integral of $X \in L^{\infty}$ with respect to $Q \in \mathcal{M}_{1,f}(P)$. There are two equivalent ways of defining this integral. One can either define

$$E_Q[X] := \sum_{i=1}^n a_i Q[A_i]$$

for step functions $X = \sum_{i=1}^n a_i \mathbb{I}_{A_i}$, and then extend the integral to L^{∞} by using the density of the step functions. Alternatively, for any $X \in \mathcal{X}$, the expectation $E_Q[X]$ can be defined as the *Choquet integral*

$$\int X dQ = \int_{-\infty}^0 (Q[X > x] - 1) dx + \int_0^{\infty} Q[X > x] dx.$$

It is not difficult to check that both integral notions coincide on step functions and hence are equivalent.

Let $\alpha : \mathcal{M}_{1,f}(P) \rightarrow \mathbb{R} \cup \{+\infty\}$ be any functional such that

$$\inf_{Q \in \mathcal{M}_{1,f}(P)} \alpha(Q) \in \mathbb{R}.$$

For each $Q \in \mathcal{M}_{1,f}(P)$ with $\alpha(Q) < \infty$, the functional $X \mapsto E_Q[-X] - \alpha(Q)$ is convex, monotone, and cash invariant on L^∞ , and these three properties are preserved when taking the supremum over $Q \in \mathcal{M}_{1,f}(P)$. Hence,

$$\rho(X) := \sup_{Q \in \mathcal{M}_{1,f}(P)} (E_Q[-X] - \alpha(Q)) \quad (10)$$

defines a convex measure of risk on L^∞ such that

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

The functional α will be called a *penalty function* for ρ on $\mathcal{M}_{1,f}(P)$, and we will say that ρ is *represented* by α on $\mathcal{M}_{1,f}(P)$. The following representation theorem states that any convex risk measure on L^∞ is of this form and it also gives a variational formula for the minimal penalty function α_{\min} .

Theorem 1.10. *Any convex measure of risk ρ on L^∞ is of the form*

$$\rho(X) = \max_{Q \in \mathcal{M}_{1,f}(P)} (E_Q[-X] - \alpha_{\min}(Q)), \quad (11)$$

where the penalty function α_{\min} is given by

$$\alpha_{\min}(Q) := \sup_{X \in \mathcal{A}_\rho} E_Q[-X] \quad \text{for } Q \in \mathcal{M}_{1,f}(P). \quad (12)$$

Moreover, α_{\min} is the minimal penalty function which represents ρ , i.e., any penalty function α , for which (10) holds, satisfies $\alpha(Q) \geq \alpha_{\min}(Q)$ for all $Q \in \mathcal{M}_{1,f}(P)$.

Proof. Recall that $X' := \rho(X) + X \in \mathcal{A}_\rho$ by (1). Thus, for all $Q \in \mathcal{M}_{1,f}(P)$,

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

and we get that α_{\min} defined by (12) is also given by

$$\alpha_{\min}(Q) = \sup_{X \in L^\infty} (E_Q[-X] - \rho(X)). \quad (13)$$

Thus, α_{\min} corresponds to the Fenchel-Legendre transform of the convex function ρ on the Banach space L^∞ . More precisely,

$$\alpha_{\min}(Q) = \rho^*(\ell_Q), \quad (14)$$

where $\rho^* : (L^\infty)' \rightarrow \mathbb{R} \cup \{+\infty\}$ is defined on the dual $(L^\infty)'$ of L^∞ by

$$\rho^*(\ell) = \sup_{X \in L^\infty} (\ell(X) - \rho(X)),$$

and where $\ell_Q \in (L^\infty)'$ is given by $\ell_Q(X) = E_Q[-X]$.

The functional ρ is lower semicontinuous with respect to the weak topology on L^∞ , since any set $\{\rho \leq c\}$ is convex, strongly closed due to Lemma 1.2, and hence weakly closed (see, e.g., Theorem V.3.13 in [24]). Thus, the general biduality theorem for conjugate functions as stated in [26] or in [35, Theorem A.61] yields

$$\rho^{**} = \rho,$$

where ρ^{**} denotes the conjugate function of ρ^* , i.e.,

$$\rho(X) = \rho^{**}(X) = \sup_{\ell \in (L^\infty)'} (\ell(X) - \rho^*(\ell)). \quad (15)$$

Recall next that $(L^\infty)'$ can be identified with the space $ba(P) := ba(\Omega, \mathcal{F}, P)$ of finitely additive set functions with finite total variation that are absolutely continuous with respect to P (see, e.g., [35, Theorem A.50] or Theorem IV.8.16 in [24]). Moreover, $\rho^*(\ell) < \infty$ implies that $-\ell$ can be identified with some $Q \in \mathcal{M}_{1,f}(P)$, as we will show now. First, the cash invariance of ρ implies that

$$\rho^*(\ell) = \sup_{X \in L^\infty} (\ell(X+1) - \rho(X+1)) = \rho^*(\ell) + \ell(1) + 1$$

and hence that $\ell(1) = -1$. Second, we have $\ell(X) \leq 0$ for $X \geq 0$, since

$$\rho^*(\ell) \geq \ell(cX) - \rho(cX) \geq c \cdot \ell(X) - \rho(0)$$

for all $c \geq 0$.

Thus, we see that (15) reduces to the representation

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

Moreover, the supremum is actually attained: $\mathcal{M}_{1,f}(P)$ is weak* compact in $(L^\infty)' = ba(P)$ due to the Banach-Alaoglu theorem, and so the upper semicontinuous functional $Q \mapsto E_Q[-X] - \alpha_{\min}(Q)$ attains its maximum on $\mathcal{M}_{1,f}(P)$.

Finally, let α be any penalty function for ρ . Then, for all $Q \in \mathcal{M}_{1,f}$ and $X \in L^\infty$,

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

and hence

$$\alpha(Q) \geq \sup_{X \in L^\infty} (E_Q[-X] - \rho(X)) \geq \sup_{X \in \mathcal{A}_\rho} (E_Q[-X] - \rho(X)) \geq \alpha_{\min}(Q). \quad (16)$$

Thus, α dominates α_{\min} . □

Corollary 1.11. *The minimal penalty function α_{\min} of a coherent measure of risk ρ takes only the values 0 and $+\infty$. In particular,*

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

for the convex set

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

and \mathcal{Q}_{\max} is the largest set \mathcal{Q} for which a representation of the form $\rho(X) = \sup_{Q \in \mathcal{Q}} E_Q[-X]$ holds.

Proof. Due to the positive homogeneity of ρ , its minimal penalty function satisfies

$$\alpha_{\min}(Q) = \sup_{X \in L^\infty} (E_Q[-X] - \rho(X)) = \sup_{X \in L^\infty} (E_Q[-\lambda X] - \rho(\lambda X)) = \lambda \alpha_{\min}(Q)$$

for all $Q \in \mathcal{M}_{1,f}(P)$ and $\lambda > 0$. Hence, α_{\min} can take only the values 0 and $+\infty$. □

The penalty function α arising in the representation (10) is not unique, and it is often convenient to represent a convex measure of risk by a penalty function that is not the minimal one. For instance, the minimal penalty function may be finite for certain finitely additive set functions while another α is concentrated only on probability measures as in the case of Example 1.7. Another situation of this type occurs for risk measures that are constructed as the supremum of a family of convex measures of risk:

Proposition 1.12. *Suppose that for every i in some index set I we are given a convex measure of risk ρ_i on L^∞ with associated penalty function α_i . If $\sup_{i \in I} \rho_i(0) < \infty$ then*

$$\rho(X) := \sup_{i \in I} \rho_i(X)$$

is a convex measure of risk that can be represented with the penalty function

$$\alpha(Q) := \inf_{i \in I} \alpha_i(Q).$$

Proof. The condition $\rho(0) = \sup_{i \in I} \rho_i(0) < \infty$ implies that ρ takes only finite values. Moreover,

$$\rho(X) = \sup_{i \in I} \sup_{Q \in \mathcal{M}_{1,f}(P)} (E_Q[-X] - \alpha_i(Q)) = \sup_{Q \in \mathcal{M}_{1,f}(P)} (E_Q[-X] - \inf_{i \in I} \alpha_i(Q)),$$

and the assertion follows. \square

In the sequel, we are particularly interested in convex measures of risk, which admit a representation in terms of σ -additive probability measures. Such a risk measure ρ can be represented by a penalty function α , which is infinite outside the set $\mathcal{M}_1(P) = \mathcal{M}_1(\Omega, \mathcal{F}, P)$:

$$\rho(X) = \sup_{Q \in \mathcal{M}_1(P)} (E_Q[-X] - \alpha(Q)). \quad (17)$$

In this case, one can no longer expect that the supremum above is attained. This will also be illustrated by Example 1.18 below.

Theorem 1.13. *Suppose $\rho : L^\infty \rightarrow \mathbb{R}$ is a convex measure of risk. Then the following conditions are equivalent.*

(a) ρ can be represented by some penalty function on $\mathcal{M}_1(P)$.

(b) ρ can be represented by the restriction of the minimal penalty function α_{\min} to $\mathcal{M}_1(P)$:

$$\rho(X) = \sup_{Q \in \mathcal{M}_1(P)} (E_Q[-X] - \alpha_{\min}(Q)), \quad X \in L^\infty. \quad (18)$$

(c) ρ is continuous from above: If $X_n \searrow X$ P -a.s. then $\rho(X_n) \nearrow \rho(X)$.

(d) ρ has the ‘‘Fatou property’’: For any bounded sequence (X_n) which converges P -a.s. to some X ,

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

(e) ρ is lower semicontinuous for the weak* topology $\sigma(L^\infty, L^1)$.

(f) The acceptance set \mathcal{A}_ρ of ρ is weak* closed.

Proof. (b) \Rightarrow (a) is obvious.

(a) \Rightarrow (d): Dominated convergence implies that $E_Q[X_n] \rightarrow E_Q[X]$ for each $Q \in \mathcal{M}_1(P)$. Hence,

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

(d) \Rightarrow (c): By monotonicity, $\rho(X_n) \leq \rho(X)$ for each n if $X_n \searrow X$, and so $\rho(X_n) \nearrow \rho(X)$ follows.

(c) \Rightarrow (d): Define $Y_m := \sup_{n \geq m} X_n$. Then Y_m decreases P -a.s. to X . Since $\rho(X_n) \geq \rho(Y_n)$ by monotonicity, we get (d) from (c).

(d) \Rightarrow (e): We have to show that $\mathcal{C} := \{\rho \leq c\}$ is weak* closed for $c \in \mathbb{R}$. To this end, let $\mathcal{C}_r := \mathcal{C} \cap \{X \in L^\infty \mid \|X\|_\infty \leq r\}$ for $r > 0$. If (X_n) is a sequence in \mathcal{C}_r converging in L^1 to some random variable X , then there is a subsequence that converges P -a.s., and the Fatou property of ρ implies that $X \in \mathcal{C}_r$. Hence, \mathcal{C}_r is closed in L^1 and, due to convexity, also weakly closed in L^1 . Since the natural injection

$$(L^\infty, \sigma(L^\infty, L^1)) \longrightarrow (L^1, \sigma(L^1, L^\infty))$$

is continuous, \mathcal{C}_r is $\sigma(L^\infty, L^1)$ -closed in L^∞ . Thus, \mathcal{C} is weak* closed due to the Krein–Šmulian theorem (see, e.g., Theorem V.5.7 in [24]).

(e) \Rightarrow (f) is obvious, and (f) \Rightarrow (e) follows from cash invariance.

(e) \Rightarrow (b): Weak* lower semicontinuity allows us to repeat the proof of Theorem 1.10 for the weak* topology on L^∞ . The dual of L^∞ for this topology is given by L^1 , so that ρ^* is by definition concentrated on σ -additive measures absolutely continuous with respect to P . Only the compactness argument breaks down for $\mathcal{M}_1(P)$, so that we may no longer replace the supremum by a maximum in (18). All remaining arguments carry over with minor modifications. \square

Remark 1.14. The proof of Theorem 1.13 can be modified in a straightforward manner to cover representation theorems for convex risk measures on the Banach spaces $L^p(\Omega, \mathcal{F}, P)$ for $1 \leq p < \infty$. More precisely, let $q \in (1, \infty]$ be such that $\frac{1}{p} + \frac{1}{q} = 1$, and define

$$\mathcal{M}_1^q(P) := \left\{ Q \in \mathcal{M}_1(P) \mid \frac{dQ}{dP} \in L^q \right\}.$$

A convex risk measure ρ on L^p is of the form

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

if and only if it is lower semicontinuous on L^p , i.e., the Fatou property holds in the form

$$X_n \longrightarrow X \text{ in } L^p \implies \rho(X) \leq \liminf_{n \uparrow \infty} \rho(X_n). \quad (19)$$

In fact, one can show that (19) is automatically satisfied for *any* convex risk measure on L^p with $1 \leq p < \infty$; see Cheridito et al. [13, Proposition 3.8]. \diamond

Definition 1.15. A convex measure of risk ρ on L^∞ is called *sensitive* with respect to P if

$$\rho(-X) > \rho(0)$$

for all $X \in L_+^\infty$ such that $E[X] > 0$.

Sensitivity is also called *relevance*. In the context of Theorem 1.13, it is equivalent to the condition that $E_Q[X] > \alpha(Q)$ for some $Q \in \mathcal{M}_1(P)$ as soon as $X \geq 0$ is such that $E[X] > 0$.

Theorem 1.13 takes the following form for *coherent* measures of risk; the proof is the same as the one for Corollary 1.11.

Corollary 1.16. *A coherent measure of risk on L^∞ can be represented by a set $\mathcal{Q} \subset \mathcal{M}_1(P)$ if and only if the equivalent conditions of Theorem 1.13 are satisfied. In this case, the maximal representing subset of $\mathcal{M}_1(P)$ is given by*

$$\mathcal{Q}_{\max} := \{ Q \in \mathcal{M}_1(P) \mid \alpha_{\min}(Q) = 0 \}.$$

Moreover, ρ is sensitive if and only if $\mathcal{Q}_{\max} \sim P$ in the sense that for any $A \in \mathcal{F}$

$$P[A] = 0 \iff Q[A] = 0 \text{ for all } Q \in \mathcal{Q}_{\max}.$$

Continuity from above was one of the equivalent properties in Theorem 1.13. If one considers *continuity from below* in the form

$$X_n \nearrow X \implies \rho(X_n) \searrow \rho(X),$$

then it turns out that it is a stronger condition than continuity from above. One can show that continuity from below is equivalent to the following *Lebesgue property*:

$$\text{If } (X_n) \text{ are uniformly bounded and } X_n \rightarrow X \text{ } P\text{-a.s. then } \rho(X_n) \rightarrow \rho(X);$$

see [34, Remark 4.19] or [35, Remark 4.23]. The argument relies on the fact that continuity from below implies that the minimal penalty function is concentrated on $\mathcal{M}_1(P)$, as is proved in the next corollary for coherent risk measures.

Corollary 1.17. *For a coherent measure of risk ρ on L^∞ the following properties are equivalent:*

(a) ρ is continuous from below: $X_n \nearrow X \implies \rho(X_n) \searrow \rho(X)$.

(b) There exists a set $\mathcal{Q} \subset \mathcal{M}_1(P)$ representing ρ such that the supremum is attained:

$$\rho(X) = \max_{Q \in \mathcal{Q}} E_Q[-X] \quad \text{for all } X \in L^\infty.$$

(c) There exists a set $\mathcal{Q} \subset \mathcal{M}_1(P)$ representing ρ such that the set of densities

$$\mathcal{D} := \left\{ \frac{dQ}{dP} \mid Q \in \mathcal{Q} \right\}$$

is weakly compact in $L^1(\Omega, \mathcal{F}, P)$.

Proof. (c) \implies (a): This follows from Dini's lemma.

(a) \implies (b): Consider the representation $\rho(X) = \max_{Q \in \mathcal{Q}_{\max}} E_Q[-X]$, where \mathcal{Q}_{\max} is the maximal representing subset on $\mathcal{M}_{1,f}(P)$, and let $A_1 \subset A_2 \subset \dots$ be any decreasing sequence of events such that $\bigcap_n A_n = \emptyset$. Continuity from below implies that $Q[A_n] \rightarrow 0$ for each $Q \in \mathcal{Q}_{\max}$, so that $\mathcal{Q}_{\max} \subset \mathcal{M}_1(P)$.

(b) \implies (c): Without loss of generality, we can assume that \mathcal{D} is weakly closed in L^1 . For any $X \in L^\infty$, the continuous linear functional J_X on L^1 defined by $J_X(Z) := E[XZ]$ attains its infimum on \mathcal{D} . According to James' theorem (see, for instance, [28]), this implies weak compactness of \mathcal{D} . \square

A similar result as the preceding corollary also holds for convex risk measures on L^∞ ; see [34, Proposition 4.17] or [35, Proposition 4.21] and Jouini et al. [48]. We now give examples of coherent measures of risk that will be studied in more detail in Section 1.3.

Example 1.18. In our present context, where we require condition (9), the *worst-case risk measure* takes the form

$$\rho_{\max}(X) := -\text{ess inf } X = \inf \{ m \in \mathbb{R} \mid X + m \geq 0 \text{ } P\text{-a.s.} \}.$$

One can easily check that ρ_{\max} is coherent and satisfies the Fatou property. Moreover, the acceptance set of ρ_{\max} is equal to the positive cone L_+^∞ in L^∞ , and this implies $\alpha_{\min}(Q) = 0$ for any $Q \in \mathcal{M}_1(P)$. Thus,

$$\rho_{\max}(X) = \sup_{Q \in \mathcal{M}_1(P)} E_Q[-X].$$

Note however that the supremum on the right cannot be replaced by a maximum in case (Ω, \mathcal{F}, P) cannot be reduced to a finite model. Indeed, let $X \in L^\infty$ be such that X does not attain its essential infimum. Then there can be no $Q \in \mathcal{M}_1(P)$ such that $E_Q[X] = \text{ess inf } X = -\rho_{\max}(X)$. In this case, the preceding corollary shows that ρ_{\max} is not continuous from below. \diamond

Example 1.19. Let \mathcal{Q}_λ be the class of all $Q \in \mathcal{M}_1(P)$ whose density dQ/dP is bounded by $1/\lambda$ for some fixed parameter $\lambda \in (0, 1)$. The corresponding coherent risk measure

$$AV@R_\lambda(X) := \sup_{Q \in \mathcal{Q}_\lambda} E_Q[-X] \tag{20}$$

will be called the *Average Value at Risk* at level λ . This terminology will become clear in Section 1.3, which contains a detailed study of $AV@R_\lambda$. Note that the set of densities dQ/dP for $Q \in \mathcal{Q}_\lambda$ is weakly closed in L^1 . Moreover, it is weakly compact due to the Dunford–Pettis theorem. Thus, the supremum in (20) is actually attained, and Corollary 1.17 applies. An explicit construction of the maximizing measure will be given in the proof of Theorem 1.31. \diamond

Example 1.20. We take for \mathcal{Q} the class of all conditional distributions $P[\cdot \mid A]$ such that $A \in \mathcal{F}$ has $P[A] > \lambda$ for some fixed level $\lambda \in (0, 1)$. The coherent measure of risk induced by \mathcal{Q} ,

$$WCE_\lambda(X) := \sup \{ E[-X \mid A] \mid A \in \mathcal{F}, P[A] > \lambda \}, \tag{21}$$

is called the *worst conditional expectation* at level λ . We will show in Section 1.3 that it coincides with the Average Value at Risk of Example 1.19 if the underlying probability space is rich enough. \diamond

1.3 Value at Risk

As seen in Example 1.9, a common approach to the problem of measuring the risk of a financial position X consists in specifying a quantile of the distribution of X under the given probability measure P . In the sequel, we will first recall the notion of a quantile function. We will use the generic notation F_X for the distribution function of a random variable X . When the emphasis is on the law μ of X , we will also write F_μ .

Definition 1.21. A function $q_X : (0, 1) \rightarrow \mathbb{R}$ is called a *quantile function* for X if

$$F_X(q_X(s)-) \leq s \leq F_X(q_X(s)) \quad \text{for all } s \in (0, 1).$$

The left- and rightcontinuous inverse functions of F_X ,

$$q_X^-(s) = \sup\{x \in \mathbb{R} \mid F_X(x) < s\} \quad \text{and} \quad q_X^+(s) = \inf\{x \in \mathbb{R} \mid F_X(x) > s\},$$

are called the *lower* and *upper quantile functions*. The value $q_X(\lambda)$ of a quantile function at a given level $\lambda \in (0, 1)$ is called a λ -*quantile* of X .

The following lemma explains the reason for calling q_X^- and q_X^+ the upper and lower quantile functions.

Lemma 1.22. A function $q : (0, 1) \rightarrow \mathbb{R}$ is a quantile function X if and only if

$$q_X^-(s) \leq q(s) \leq q_X^+(s) \quad \text{for all } s \in (0, 1).$$

In particular, q_X^- and q_X^+ are quantile functions. Moreover, q_X^- is left-continuous, q_X^+ is right-continuous, and every quantile function q_X is increasing and satisfies $q_X(s-) = q_X^-(s)$ and $q_X(s+) = q_X^+(s)$ for all $s \in (0, 1)$. In particular, any two quantile functions coincide a.e. on $(0, 1)$.

Proof. We have $q_X^- \leq q_X^+$, and any quantile function q_X satisfies $q_X^- \leq q_X \leq q_X^+$, due to the definitions of q_X^- and q_X^+ . Hence, the first part of the assertion follows if we can show that $F_X(q_X^+(s)-) \leq s \leq F_X(q_X^-(s))$ for all s . But $x < q_X^+(s)$ implies $F_X(x) \leq s$ and $y > q_X^-(s)$ implies $F_X(y) \geq s$, which gives the result.

Next, the set $\{x \mid F_X(x) > s\}$ is the union of the sets $\{x \mid F_X(x) > s + \varepsilon\}$ for $\varepsilon < 0$, and so q_X^+ is right-continuous. An analogous argument shows the left-continuity of q_X^- . It is clear that both q_X^- and q_X^+ are increasing, so that the second part of the assertion follows. \square

Remark 1.23. The left- and right-continuous quantile functions can also be represented as

$$q_X^-(s) = \inf\{x \in \mathbb{R} \mid F_X(x) \geq s\} \quad \text{and} \quad q_X^+(s) = \sup\{x \in \mathbb{R} \mid F_X(x) \leq s\}.$$

To see this, note first that $q_X^-(s)$ is clearly dominated by the infimum. On the other hand, $y > q_X^-(s)$ implies $F_X(y) \geq s$, and we get $q_X^-(s) \geq \inf\{x \in \mathbb{R} \mid F_X(x) \geq s\}$. The proof for q_X^+ is analogous. \diamond

The following basic fact is well known. See, e.g., [35, Appendix A.3] for a proof.

Lemma 1.24. Let U be a random variable on a probability space (Ω, \mathcal{F}, P) with a uniform distribution on $(0, 1)$, i.e., $P[U \leq s] = s$ for all $s \in (0, 1)$. If q_X is a quantile function for the random variable X , then

$$\tilde{X}(\omega) := q_X(U(\omega))$$

has the same distribution as X . If, moreover, F_X is continuous, then $U := F_X(X)$ is uniformly distributed on $(0, 1)$, and $X = q_X(U)$ P -almost surely.

The second part of the preceding lemma implies that a probability space supports a random variable with uniform distribution on $(0, 1)$ if and only if it supports any non-constant random variable X with a continuous distribution. See Lemma 1.40 below for a more general result.

In this section, we will focus on the properties of $q_X^+(\lambda)$, viewed as a functional on a space of financial positions X . Many results presented here were first obtained by Artzner, Delbaen, Eber and Heath [3], Delbaen [20, 21], and Acerbi and Tasche [1]; see the notes in [35] for details.

Definition 1.25. Fix some level $\lambda \in (0, 1)$. For a financial position X , we define its *Value at Risk at level λ* as

$$V@R_\lambda(X) := -q_X^+(\lambda) = q_{-X}^-(1 - \lambda) = \inf\{m \mid P[X + m < 0] \leq \lambda\}. \quad (22)$$

In financial terms, $V@R_\lambda(X)$ is the smallest amount of capital which, if added to X and invested in the risk-free asset, keeps the probability of a negative outcome below the level λ . However, Value at Risk only controls the probability of a loss; it does not capture the size of such a loss if it occurs. Clearly, $V@R_\lambda$ is a monetary measure of risk on $\mathcal{X} = L^0$ and positively homogeneous. The following example shows that the acceptance set of $V@R_\lambda$ is typically not convex, and so $V@R_\lambda$ is *not* a convex measure of risk. In particular, $V@R_\lambda$ may penalize diversification instead of encouraging it, and this fact will also be illustrated by the example.

Example 1.26. 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, or P&L, 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

$$P\left[X_1 - \frac{w(\tilde{r} - r)}{1 + r} < 0\right] = P[1^{\text{st}} \text{ bond defaults}] = p \leq \lambda.$$

Hence,

$$V@R_\lambda(X_1) = -\frac{w(\tilde{r} - r)}{1 + r} < 0.$$

This means that the position X_1 is acceptable in the sense that it does not carry a positive Value at Risk, regardless of the possible loss of the entire investment w . In fact, from a $V@R$ point of view, an investment into the defaultable bond is actually “less risky” than a riskfree investment with P&L $R \equiv 0$ and $V@R_\lambda(R) = 0$.

Diversifying the portfolio by investing the amount $w/2$ into each of the two bonds leads to the position $Y := (X_1 + X_2)/2$. Let us assume that the two bonds default independently of each other, each of them with probability p . For realistic \tilde{r} , the probability that Y is negative is equal to the probability that at least one of the two bonds defaults: $P[Y < 0] = p(2 - p)$. If, for instance, $p = 0.009$ and $\lambda = 0.01$ then we have $p < \lambda < p(2 - p)$, hence

$$V@R_\lambda(Y) = \frac{w}{2} \cdot \left(1 - \frac{\tilde{r} - r}{1 + r}\right).$$

Typically, this value is close to one half of the invested capital w and thus constitutes a dramatic increase compared to the investment where the entire amount w is invested into one defaultable bond only. Thus, the acceptance set of $V@R_\lambda$ is not convex. This example also illustrates that $V@R$ 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. Thus, optimizing a portfolio with respect to $V@R_\lambda$ may lead to a concentration of the portfolio in one single asset with a sufficiently small default probability, but with an exposure to large losses. \diamond

In the remainder of this section, we will focus on monetary measures of risk which, in contrast to $V@R_\lambda$, are convex or even coherent on $\mathcal{X} := L^\infty$. In particular, we are looking for convex risk measures which come close to $V@R_\lambda$. A first guess might be that one should take the smallest convex measure of risk, continuous from above, which dominates $V@R_\lambda$. However, since $V@R_\lambda$ itself is not convex, the following proposition shows that such a smallest $V@R_\lambda$ -dominating convex measure of risk does not exist. A proof can be found in [20].

Proposition 1.27. For each $X \in L^\infty$ and each $\lambda \in (0, 1)$,

$$V@R_\lambda(X) = \min \{ \rho(X) \mid \rho \text{ is convex, continuous from above, and } \geq V@R_\lambda \}.$$

For the rest of this section, we concentrate on the following risk measure which is defined in terms of Value at Risk, but does satisfy the axioms of a coherent risk measure.

Definition 1.28. The *Average Value at Risk at level* $\lambda \in (0, 1]$ of a position $X \in L^\infty$ is given by

$$AV@R_\lambda(X) = \frac{1}{\lambda} \int_0^\lambda V@R_\gamma(X) d\gamma.$$

Sometimes, the Average Value at Risk is also called the ‘‘Conditional Value at Risk’’ or the ‘‘expected shortfall’’, and one writes $CV@R_\lambda(X)$ or $ES_\lambda(X)$. These terms are motivated by formulas (26) and (23) below, but they are potentially misleading: ‘‘Conditional Value at Risk’’ might also be used to denote the Value at Risk with respect to a conditional distribution, and ‘‘expected shortfall’’ might be understood as the expectation of the shortfall X^- . For these reasons, we prefer the term Average Value at Risk. Note that

$$AV@R_\lambda(X) = -\frac{1}{\lambda} \int_0^\lambda q_X(t) dt$$

by (22). In particular, the definition of $AV@R_\lambda(X)$ makes sense for any $X \in L^1(\Omega, \mathcal{F}, P)$ and we have, in view of Lemma 1.24,

$$AV@R_1(X) = -\int_0^1 q_X^+(t) dt = E[-X].$$

Remark 1.29. For $X \in L^\infty$, we have $\lim_{\lambda \downarrow 0} V@R_\lambda(X) = -\text{ess inf } X = \inf \{ m \mid P[X + m < 0] \leq 0 \}$. Hence, it makes sense to define

$$AV@R_0(X) := V@R_0(X) := -\text{ess inf } X,$$

which is the worst-case risk measure on L^∞ introduced in Example 1.18. Recall that it is continuous from above but in general not from below. \diamond

Lemma 1.30. For $\lambda \in (0, 1)$ and any λ -quantile q of X ,

$$AV@R_\lambda(X) = \frac{1}{\lambda} E[(q - X)^+] - q = \frac{1}{\lambda} \inf_{r \in \mathbb{R}} (E[(r - X)^+] - \lambda r). \quad (23)$$

Proof. Let q_X be a quantile function with $q_X(\lambda) = q$. By Lemma 1.24,

$$\frac{1}{\lambda} E[(q - X)^+] - q = \frac{1}{\lambda} \int_0^1 (q - q_X(t))^+ dt - q = -\frac{1}{\lambda} \int_0^\lambda q_X(t) dt = AV@R_\lambda(X).$$

This proves the first identity. For the second one, let f denote the convex function $f(r) := E[(r - X)^+] - \lambda r$. Note that the right-hand and left-hand derivatives of f are given by $f'_+(r) = F_X(r) - \lambda$ and $f'_-(r) = F_X(r-) - \lambda$. A point r is a minimizer of f iff $f'_+(r) \geq 0$ and $f'_-(r) \leq 0$, which is equivalent to r being a λ -quantile. This proves the second identity. \square

Theorem 1.31. For $\lambda \in (0, 1]$, $AV@R_\lambda$ is a coherent measure of risk which is continuous from below. It has the representation

$$AV@R_\lambda(X) = \max_{Q \in \mathcal{Q}_\lambda} E_Q[-X] \quad (24)$$

where \mathcal{Q}_λ is the set of all probability measures $Q \ll P$ whose density dQ/dP is P -a.s. bounded by $1/\lambda$. Moreover, \mathcal{Q}_λ is equal to the maximal set \mathcal{Q}_{\max} of Corollary 1.16.

The proof relies on the following version of the classical Neyman-Pearson lemma, which we recall here for the convenience of the reader. It is concerned with the infinite-dimensional optimization problem

$$\text{maximize } \tilde{E}[\psi] \text{ under the constraints } 0 \leq \psi \leq 1 \text{ and } E[\psi] = \lambda, \quad (25)$$

where $\lambda \in [0, 1]$ and \tilde{P} is probability measure such that $\tilde{P} \ll P$.

Proposition 1.32 (Neyman-Pearson lemma). *A solution to the problem (25) is given by*

$$\psi^0 = \mathbb{1}_{\{\varphi > q\}} + \kappa \mathbb{1}_{\{\varphi = q\}},$$

where q is a $(1 - \lambda)$ -quantile of $\varphi := d\tilde{P}/dP$ with respect to P and κ is defined as

$$\kappa := \begin{cases} 0 & \text{if } P[\varphi = q] = 0, \\ \frac{\lambda - P[\varphi > q]}{P[\varphi = q]} & \text{otherwise.} \end{cases}$$

Moreover, any other solution coincides with ψ^0 , P -a.s. on $\{\varphi \neq q\}$. In particular, ψ^0 is the P -a.s. unique $\sigma(\varphi)$ -measurable maximizer.

Proof. We first show that ψ^0 satisfies the constraints of the problem. Let F_φ denote the distribution function of φ under P . Then $P[\varphi > q] = 1 - F_\varphi(q) \leq \lambda$ and

$$P[\varphi = q] = F_\varphi(q) - F_\varphi(q-) \geq F_\varphi(q) - 1 + \lambda = \lambda - P[\varphi > q].$$

Hence $0 \leq \kappa \leq 1$ and in turn $0 \leq \psi^0 \leq 1$. The fact that $E[\psi^0] = \lambda$ is obvious.

Next, let ψ be any other measurable function satisfying the constraints. Then $(\psi^0 - \psi)(\varphi - q) \geq 0$. Hence

$$\tilde{E}[\psi^0] - \tilde{E}[\psi] = E[(\psi^0 - \psi)\varphi] \geq q(E[\psi^0] - E[\psi]) = 0.$$

Thus, ψ^0 solves the optimization problem.

Finally, suppose that ψ^* is another solution. Then $\tilde{E}[\psi^*] = \tilde{E}[\psi^0] = \lambda$ and also $E[\psi^*] = E[\psi^0]$, due to the already established fact that ψ^0 is a solution. Hence,

$$0 = \tilde{E}[\psi^0 - \psi^*] - q E[\psi^0 - \psi^*] = E[(\psi^0 - \psi^*)(\varphi - q)].$$

But we have seen above that $(\psi^0 - \psi^*)(\varphi - q) \geq 0$. Hence, $(\psi^0 - \psi^*)(\varphi - q) = 0$ P -a.s., i.e., $\psi^* = \psi^0$ P -a.s. on $\{\varphi \neq q\}$. \square

Proof of Theorem 1.31. Since $\mathcal{Q}_1 = \{P\}$, the assertion is obvious for $\lambda = 1$. For $0 < \lambda < 1$, We will show that the coherent risk measure $\rho_\lambda(X) := \sup_{Q \in \mathcal{Q}_\lambda} E_Q[-X]$ is such that the supremum in its definition is attained and that $\rho_\lambda(X) = AV@R_\lambda(X)$. Since both ρ_λ and $AV@R_\lambda$ are cash invariant and positively homogeneous, we may assume without loss of generality that $X < 0$ with $E[-X] = 1$. We define a measure $\tilde{P} \sim P$ by $d\tilde{P}/dP = -X$. Then

$$\rho_\lambda(X) = \frac{1}{\lambda} \sup \{ \tilde{E}[\psi] \mid 0 \leq \psi \leq 1, E[\psi] = \lambda \}.$$

Due to the Neyman-Pearson lemma, the supremum is attained by

$$\psi_0 = \mathbb{1}_{\{X < q\}} + \kappa \mathbb{1}_{\{X = q\}}$$

for a λ -quantile q of X and some $\kappa \in [0, 1]$ for which $E[\psi_0] = \lambda$. Note that a λ -quantile of X is the negative of a $(1 - \lambda)$ -quantile for $-X$. We get

$$\rho_\lambda(X) = \frac{1}{\lambda} \cdot \tilde{E}[\psi_0] = \frac{1}{\lambda} E[-X \psi_0].$$

Since $dQ_0 = \lambda^{-1} \psi_0 dP$ defines a probability measure in \mathcal{Q}_λ , we conclude that

$$\begin{aligned} \rho_\lambda(X) &= \max_{Q \in \mathcal{Q}_\lambda} E_Q[-X] = E_{Q_0}[-X] \\ &= \frac{1}{\lambda} (E[-X; X < q] - q(\lambda + P[X < q])) \\ &= \frac{1}{\lambda} E[(q - X)^+] - q \\ &= AV@R_\lambda(X), \end{aligned}$$

where we have used (23) in the last step. This proves (24).

It remains to prove that \mathcal{Q}_λ is the maximal set of Corollary 1.16. To this end, we show that

$$\sup_{X \in L^\infty} (E_Q[-X] - AV@R_\lambda(X)) = +\infty \quad \text{for } Q \notin \mathcal{Q}_\lambda.$$

We denote $\psi := dQ/dP$. There exist $\lambda' \in (0, \lambda)$ and $k > 1/\lambda'$ such that $P[\psi \wedge k \geq 1/\lambda'] > 0$. For $c > 0$ define $X^{(c)} \in L^\infty$ by

$$X^{(c)} := -c(\psi \wedge k) \mathbb{I}_{\{\psi \geq 1/\lambda'\}}.$$

Since

$$P[X^{(c)} < 0] = P\left[\psi \geq \frac{1}{\lambda'}\right] \leq \lambda' < \lambda,$$

we have $V@R_\lambda(X^{(c)}) = 0$, and (23) yields that

$$AV@R_\lambda(X^{(c)}) = \frac{1}{\lambda} E[-X^{(c)}] = \frac{c}{\lambda} E\left[\psi \wedge k; \psi \geq \frac{1}{\lambda'}\right].$$

On the other hand,

$$E_Q[-X^{(c)}] = c \cdot E\left[\psi \cdot (\psi \wedge k); \psi \geq \frac{1}{\lambda'}\right] \geq \frac{c}{\lambda'} E\left[\psi \wedge k; \psi \geq \frac{1}{\lambda'}\right].$$

Thus, the difference between $E_Q[-X^{(c)}]$ and $AV@R_\lambda(X^{(c)})$ becomes arbitrarily large as $c \uparrow \infty$. \square

Remark 1.33. The proof shows that for $\lambda \in (0, 1)$ the maximum in (24) is attained by the measure $Q_0 \in \mathcal{Q}_\lambda$, whose density is given by

$$\frac{dQ_0}{dP} = \frac{1}{\lambda} (\mathbb{I}_{\{X < q\}} + \kappa \mathbb{I}_{\{X = q\}}),$$

where q is a λ -quantile of X , and where κ is defined as

$$\kappa := \begin{cases} 0 & \text{if } P[X = q] = 0, \\ \frac{\lambda - P[X < q]}{P[X = q]} & \text{otherwise.} \end{cases} \quad \diamond$$

Corollary 1.34. For all $X \in L^\infty$,

$$\begin{aligned} AV@R_\lambda(X) &\geq WCE_\lambda(X) \\ &\geq E[-X \mid -X \geq V@R_\lambda(X)] \\ &\geq V@R_\lambda(X), \end{aligned} \quad (26)$$

where WCE_λ is the coherent risk measure defined in (21). Moreover, the first two inequalities are in fact identities if

$$P[X \leq q_X^+(\lambda)] = \lambda, \quad (27)$$

which is the case if X has a continuous distribution.

Proof. If $P[A] \geq \lambda$, then the density of $P[\cdot \mid A]$ with respect to P is bounded by $1/\lambda$. Therefore, Theorem 1.31 implies that $AV@R_\lambda$ dominates WCE_λ . Since

$$P[-X \geq V@R_\lambda(X) - \varepsilon] > \lambda,$$

we have

$$WCE_\lambda(X) \geq E[-X \mid -X \geq V@R_\lambda(X) - \varepsilon],$$

and the second inequality follows by taking the limit as $\varepsilon \downarrow 0$. Moreover, (23) shows that

$$AV@R_\lambda(X) = E[-X \mid -X \geq V@R_\lambda(X)]$$

as soon as (27) holds. \square

Remark 1.35. We will see in Corollary 1.47 that the two coherent risk measures $AV@R_\lambda$ and WCE_λ coincide if the underlying probability space is rich enough. If this is not the case, then the first inequality in (26) may be strict for some X ; see [1]. Moreover, the functional

$$E[-X \mid -X \geq V@R_\lambda(X)]$$

does not define a convex measure of risk. Hence, the second inequality in (26) cannot reduce to an identity in general; see [1]. \diamond

We have seen in Proposition 1.27 that there is no smallest convex risk measure dominating $V@R_\lambda$. But if we restrict our attention to the class of convex risk measures that dominate $V@R_\lambda$ and only depend on the distribution of a random variable, then the situation is different. In fact, we will see in Theorem 1.46 that $AV@R_\lambda$ is the smallest risk measure in this class, provided that the underlying probability space is rich enough. In this sense, Average Value at Risk can be regarded as the best conservative approximation to Value at Risk.

1.4 Law-invariant risk measures

Clearly, $V@R_\lambda$ and $AV@R_\lambda$ only involve the distribution of a position under the given probability measure P . In this section we study the class of all risk measures which share this property of law-invariance. Such risk measures were first discussed systematically by Kusuoka [53]. The extensions to the convex case were given by Dana [19], Föllmer and Schied [35], and Kunze [52].

Definition 1.36. A monetary measure of risk ρ on $\mathcal{X} = L^\infty(\Omega, \mathcal{F}, P)$ is called *law-invariant* if $\rho(X) = \rho(Y)$ whenever X and Y have the same distribution under P .

Throughout this section, we assume that the probability space (Ω, \mathcal{F}, P) is rich enough in the sense that it supports a random variable with a continuous distribution. This condition is satisfied if and only if (Ω, \mathcal{F}, P) is atomless. We can now formulate our first structure theorem for law-invariant convex risk measures.

Theorem 1.37. *Let ρ be a convex measure of risk and suppose that ρ is continuous from above. Then ρ is law-invariant if and only if its minimal penalty function $\alpha_{\min}(Q)$ depends only on the law of $\varphi_Q := \frac{dQ}{dP}$ under P when $Q \in \mathcal{M}_1(P)$. In this case, ρ has the representation*

$$\rho(X) = \sup_{Q \in \mathcal{M}_1(P)} \left(\int_0^1 q_{-X}(t) q_{\varphi_Q}(t) dt - \alpha_{\min}(Q) \right),$$

and the minimal penalty function satisfies

$$\alpha_{\min}(Q) = \sup_{X \in \mathcal{A}_\rho} \int_0^1 q_{-X}(t) q_{\varphi_Q}(t) dt = \sup_{X \in L^\infty} \left(\int_0^1 q_{-X}(t) q_{\varphi_Q}(t) dt - \rho(X) \right). \quad (28)$$

The condition of continuity from above can in fact be dropped: every law-invariant convex risk measure is automatically continuous from above, as was shown very recently by Jouini et al. [48]. We will see a particular case of this general fact in Theorem 1.56 below.

The proof of Theorem 1.37 uses the following general results on quantile functions. They will also be useful in the second part of these notes.

Lemma 1.38. *If $X = f(Y)$ for an increasing function f and q_Y is a quantile function for Y , then $f(q_Y(t))$ is a quantile function for X . In particular,*

$$q_X(t) = q_{f(Y)}(t) = f(q_Y(t)) \quad \text{for a.e. } t \in (0, 1),$$

for any quantile function q_X of X .

If f is decreasing, then $f(q_Y(1-t))$ is a quantile function for X . In particular,

$$q_X(t) = q_{f(Y)}(t) = f(q_Y(1-t)) \quad \text{for a.e. } t \in (0, 1).$$

Proof. If f is decreasing, then $q(t) := f(q_Y(1-t))$ satisfies

$$\begin{aligned} F_X(q(t)) &= P[f(Y) \leq f(q_Y(1-t))] \geq P[Y \geq q_Y(1-t)] \\ &= 1 - F_Y(q_Y(1-t)-) \geq t \geq P[Y > q_Y(1-t)] \geq F_X(q(t)-). \end{aligned}$$

Hence $q(t) = f(q_Y(1-t))$ is a quantile function. A similar argument applies to an increasing function f . \square

The following theorem is a version of the *Hardy–Littlewood inequalities*. They estimate the expectation $E[XY]$ in terms of quantile functions q_X and q_Y .

Theorem 1.39. *Let $X, Y \geq 0$ be two random variables on (Ω, \mathcal{F}, P) with quantile functions q_X and q_Y . Then,*

$$\int_0^1 q_X(1-s)q_Y(s) ds \leq E[XY] \leq \int_0^1 q_X(s)q_Y(s) ds.$$

Moreover, if $X = f(Y)$ and the lower (upper) bound is finite, then the lower (upper) bound is attained if and only if f can be chosen as a decreasing (increasing) function.

Proof. By Fubini's theorem,

$$E[XY] = E\left[\int_0^\infty \mathbb{1}_{\{X>x\}} dx \int_0^\infty \mathbb{1}_{\{Y>y\}} dy\right] = \int_0^\infty \int_0^\infty P[X > x, Y > y] dx dy.$$

Since

$$\begin{aligned} P[X > x, Y > y] &\geq (P[X > x] - P[Y \leq y])^+ \\ &= \int_0^1 \mathbb{1}_{\{F_Y(y) \leq s\}} \mathbb{1}_{\{s \leq 1 - F_X(x)\}} ds, \end{aligned}$$

and since

$$q_Z^+(s) = \sup\{x \geq 0 \mid F_Z(x) \leq s\} = \int_0^\infty \mathbb{1}_{\{F_Z(x) \leq s\}} dx \quad (29)$$

for any random variable $Z \geq 0$, another application of Fubini's theorem yields

$$E[XY] \geq \int_0^1 q_X^+(1-s)q_Y^+(s) ds = \int_0^1 q_X(1-s)q_Y(s) ds.$$

In the same way, the upper estimate follows from the inequality

$$P[X > x, Y > y] \leq P[X > x] \wedge P[Y > y] = \int_0^1 \mathbb{1}_{\{F_X(x) \leq s\}} \mathbb{1}_{\{F_Y(y) \leq s\}} ds.$$

For $X = f(Y)$,

$$E[XY] = E[f(Y)Y] = \int_0^1 f(q_Y(t))q_Y(t) dt, \quad (30)$$

due to Lemma 1.24, and so Lemma 1.38 implies that the upper and lower bounds are attained for increasing and decreasing functions, respectively.

Conversely, assume that $X = f(Y)$, and that the upper bound is attained and finite:

$$E[f(Y)Y] = \int_0^1 q_X(t)q_Y(t) dt < \infty. \quad (31)$$

Our aim is to show that P -a.s. $X = f(Y) = \tilde{f}(Y)$, where \tilde{f} is the increasing function on $[0, \infty)$ defined by $\tilde{f}(x) := q_X(F_Y(x))$ if x is a continuity point of F_Y , and by

$$\tilde{f}(x) := \frac{1}{F_Y(x) - F_Y(x-)} \int_{F_Y(x-)}^{F_Y(x)} q_X(t) dt$$

otherwise. Note that

$$\tilde{f}(q_Y) = E_\lambda[q_X | q_Y], \quad (32)$$

where $E_\lambda[\cdot | q_Y]$ denotes the conditional expectation with respect to q_Y under the Lebesgue measure λ on $(0, 1)$. Hence, (31) takes the form

$$\infty > \int_0^1 f(q_Y(t))q_Y(t) dt = \int_0^1 \tilde{f}(q_Y(t))q_Y(t) dt, \quad (33)$$

where we have used Lemma 1.24. Let ν denote the distribution of Y . By introducing the positive measures $d\mu = f d\nu$ and $d\tilde{\mu} = \tilde{f} d\nu$, (33) can be written as

$$\int_0^\infty \mu([y, \infty)) dy = \int x \mu(dx) = \int x \tilde{\mu}(dx) = \int_0^\infty \tilde{\mu}([y, \infty)) dy. \quad (34)$$

On the other hand, with g denoting the increasing function $\mathbb{1}_{[y, \infty)}$, the upper Hardy-Littlewood inequality, Lemma 1.38, and (32) yield

$$\mu([y, \infty)) = E[g(Y)f(Y)] \leq \int_0^1 q_{g(Y)}(t)q_X(t) dt = \int_0^1 g(q_Y(t))\tilde{f}(q_Y(t)) dt = \tilde{\mu}([y, \infty)).$$

In view of (34), we obtain $\mu = \tilde{\mu}$, hence $f = \tilde{f}$ ν -a.s. and $X = \tilde{f}(Y)$ P -almost surely. An analogous argument applies to the lower bound. \square

The following lemma generalizes the second part of Lemma 1.24.

Lemma 1.40. *If X is a random variable on an atomless probability space, then there exists a random variable with a uniform law on $(0, 1)$ such that $X = q_X(U)$ P -almost surely.*

Proof. We follow Ryll [59]. Without loss of generality, we may assume that $q_X = q_X^+$. Then $I_x := \{t \in (0, 1) | q_X(t) = x\}$ is a (possibly empty or degenerate) real interval with Lebesgue measure $\lambda(I_x) = P[X = x]$ for each $x \in \mathbb{R}$. Consider the set $D := \{x \in \mathbb{R} | P[X = x] > 0\}$, which is at most countable. For each $x \in D$, the probability space $(\Omega, \mathcal{F}, P[\cdot | X = x])$ is again atomless and hence supports a random variable U_x with a uniform law on I_x . That is, $P[U_x \in A | X = x] = \lambda(A \cap I_x) / \lambda(I_x)$ or, equivalently,

$$P[U_x \in A, X = x] = \lambda(A \cap I_x) \quad \text{for all measurable } A \subset (0, 1). \quad (35)$$

On $D^c = (0, 1) \setminus D$, q_X is one-to-one and hence admits a measurable inverse function F (which can actually be taken as F_X , but this fact will not be needed here). We let

$$U(\omega) := F(X(\omega))\mathbb{1}_{\{X(\omega) \notin D\}} + U_{X(\omega)}\mathbb{1}_{\{X(\omega) \in D\}},$$

which clearly is a measurable random variable. By definition we have $q_X(U(\omega)) = X(\omega)$ for all ω . It remains to show that U has a uniform law. To this end, take a measurable subset A of $(0, 1)$. Using (35) we get

$$\begin{aligned} P[U \in A] &= P[U \in A, X \notin D] + \sum_{x \in D} P[U \in A, X = x] \\ &= P[F(X) \in A, X \notin D] + \sum_{x \in D} \lambda(A \cap I_x). \end{aligned}$$

Now let I^c denote the complement of $\bigcup_{x \in D} I_x$ in $(0, 1)$. Then $\{X \notin D\} = \{X \in q_X(I^c)\}$ P -a.s. and hence

$$P[F(X) \in A, X \notin D] = P[X \in q_X(A \cap I^c)] = \lambda(q_X \in q_X(A \cap I^c)) = \lambda(A \cap I^c),$$

where we have used the fact that $\lambda \circ q_X^{-1} = P \circ X^{-1}$. This proves the result. \square

The preceding lemma and Theorem 1.39 imply the following result. Recall that we assume that (Ω, \mathcal{F}, P) is atomless.

Lemma 1.41. For $Y \in L^\infty$ and $X \in L^1$,

$$\int_0^1 q_X(t)q_Y(t) dt = \max_{\tilde{Y} \sim Y} E[X\tilde{Y}],$$

where $\tilde{Y} \sim Y$ indicates that \tilde{Y} is a random variable with the same law as Y . Moreover, the maximum is attained for $\tilde{Y} := q_Y(U)$, where U is as in Lemma 1.40, i.e., U has a uniform law on $(0, 1)$ and satisfies $X = q_X(U)$ P -a.s.

Proof. The upper Hardy-Littlewood inequality in Theorem 1.39 yields “ \geq ”. To prove the reverse inequality, let U be as in Lemma 1.40. According to Lemma 1.24, $\tilde{Y} = q_Y(U)$ then has the same law as Y . Hence,

$$E[X\tilde{Y}] = E[q_X(U)q_Y(U)] = \int_0^1 q_X(t)q_Y(t) dt.$$

□

Proof of Theorem 1.37. Suppose first that ρ is law-invariant. Then $X \in \mathcal{A}_\rho$ implies that $\tilde{X} \in \mathcal{A}_\rho$ for all $\tilde{X} \sim X$. Hence,

$$\alpha_{\min}(Q) = \sup_{X \in \mathcal{A}_\rho} E[-X\varphi_Q] = \sup_{X \in \mathcal{A}_\rho} \sup_{\tilde{X} \sim X} E[-\tilde{X}\varphi_Q] = \sup_{X \in \mathcal{A}_\rho} \int_0^1 q_{-X}(t)q_{\varphi_Q}(t) dt,$$

by Lemma 1.41. It follows that $\alpha_{\min}(Q)$ depends only on the law of φ_Q . In order to check the second identity in (28), note that $\tilde{X} := X + \rho(X)$ belongs to \mathcal{A}_ρ for any $X \in L^\infty$ and that $q_{-\tilde{X}} - \rho(X)$ is a quantile function for $-\tilde{X}$.

Conversely, let us assume that $\alpha_{\min}(Q)$ depends only on the law of φ_Q . Let us write $\tilde{Q} \sim Q$ to indicate that φ_Q and $\varphi_{\tilde{Q}}$ have the same law. Then Lemma 1.41 yields

$$\begin{aligned} \rho(X) &= \sup_{Q \in \mathcal{M}_1(P)} \left(E_Q[-X] - \alpha_{\min}(Q) \right) = \sup_{Q \in \mathcal{M}_1(P)} \sup_{\tilde{Q} \sim Q} \left(E[-X\varphi_{\tilde{Q}}] - \alpha_{\min}(Q) \right) \\ &= \sup_{Q \in \mathcal{M}_1(P)} \left(\int_0^1 q_{-X}(t)q_{\varphi_Q}(t) dt - \alpha_{\min}(Q) \right). \end{aligned}$$

□

Example 1.42. Let $u : \mathbb{R} \rightarrow \mathbb{R}$ be an increasing concave function, and suppose that a position $X \in L^\infty$ is acceptable if $E[u(X)] \geq c$, where c is a given constant in the interior of $u(\mathbb{R})$. We have seen in Example 1.8 that the corresponding acceptance set induces a convex risk measure ρ . Clearly, ρ is law-invariant, and it is not difficult to show that ρ is continuous from below and, hence, from above; see [34, Proposition 4.59] or [35, Proposition 4.104]. Moreover, the corresponding minimal penalty function can be computed as

$$\alpha_{\min}(Q) = \inf_{\lambda > 0} \frac{1}{\lambda} \left(\int_0^1 \ell^*(\lambda \cdot q_{\varphi_Q}(t)) dt - c \right),$$

where

$$\ell^*(y) = \sup_{x \in \mathbb{R}} (xy + u(-x)) = \sup_{x \in \mathbb{R}} (xy - \ell(x))$$

is the Fenchel-Legendre transform of the convex increasing loss function $\ell(x) := -u(-x)$; see [33, Theorem 10], [34, Theorem 4.61] or [35, Theorem 4.106]. \diamond

For a probability density $\varphi = dQ/dP$, the functional

$$\rho_\varphi(X) := \sup_{\tilde{\varphi} \sim \varphi} E[-X\tilde{\varphi}] = \int_0^1 q_{-X}(t)q_\varphi(t) dt$$

appearing in Theorem 1.37 is sometimes called the *maximal correlation risk measure*. The following theorem shows that ρ_φ can be represented as a mixture of the risk measures $AV@R_\lambda$ and hence is itself a coherent measure of risk. Recall that we assume that (Ω, \mathcal{F}, P) is atomless.

Theorem 1.43. *Let $\varphi = dQ/dP$ for some $Q \in \mathcal{M}_1(P)$. Then there exists a probability measure μ on $(0, 1]$ such that*

$$\int_0^1 q_{-X}(t)q_\varphi(t) dt = \int_{(0,1]} AV@R_\lambda(X) \mu(d\lambda).$$

In particular, a convex measure of risk ρ is law-invariant and continuous from above if and only if

$$\rho(X) = \sup_{\mu \in \mathcal{M}_1((0,1])} \left(\int_{(0,1]} AV@R_\lambda(X) \mu(d\lambda) - \beta_{\min}(\mu) \right), \quad (36)$$

where

$$\beta_{\min}(\mu) = \sup_{X \in \mathcal{A}_\rho} \int_{(0,1]} AV@R_\lambda(X) \mu(d\lambda).$$

Proof. Since $q_{-X}(t) = V@R_{1-t}(X)$ and $q_\varphi(t) = q_\varphi^+(t)$ for a.e. $t \in (0, 1)$,

$$\int_0^1 q_{-X}(t)q_\varphi(t) dt = \int_0^1 V@R_t(X)q_\varphi^+(1-t) dt.$$

Since q_φ^+ is increasing and right-continuous, we can write $q_\varphi^+(t) = \nu((1-t, 1])$ for some positive locally finite measure ν on $(0, 1]$. Moreover, the measure μ given by $\mu(dt) = t\nu(dt)$ is a probability measure on $(0, 1]$:

$$\int_{(0,1]} t\nu(dt) = \int_0^1 \nu((s, 1]) ds = \int_0^1 q_\varphi^+(s) ds = E[\varphi] = 1.$$

Thus,

$$\begin{aligned} \int_0^1 q_{-X}(t)q_\varphi(t) dt &= \int_0^1 V@R_t(X) \int_{(t,1]} \frac{1}{s} \mu(ds) dt \\ &= \int_{(0,1]} \frac{1}{s} \int_0^s V@R_t(X) dt \mu(ds) \\ &= \int_{(0,1]} AV@R_s(X) \mu(ds). \end{aligned} \quad (37)$$

□

The second assertion in Theorem 1.43 takes the following form for coherent measures of risk.

Corollary 1.44. *A coherent risk measure ρ is continuous from above and law-invariant if and only if*

$$\rho(X) = \sup_{\mu \in \mathcal{M}} \int_{(0,1]} AV@R_\lambda(X) \mu(d\lambda)$$

for some set $\mathcal{M} \subset \mathcal{M}_1((0, 1])$.

The preceding result is due to Kusuoka [53]. We point out once more that the condition of continuity from above can actually be dropped according to a recent result by Jouini et al. [48].

Law-invariant convex risk measures enjoy the following Jensen-type inequality, which is due to H. Föllmer and taken from [61]. Here we give a proof based on Lemma 1.30 and Theorem 1.43.

Corollary 1.45. *Assume that ρ is a convex risk measure which is continuous from above and law-invariant. Then, for $X \in L^\infty$ and any σ -algebra $\mathcal{G} \subset \mathcal{F}$,*

$$\rho(E[X | \mathcal{G}]) \leq \rho(X),$$

and in particular

$$\rho(E[X]) = \rho(0) - E[X] \leq \rho(X).$$

Proof. By Jensen's inequality for conditional expectations,

$$E[(r - E[X | \mathcal{G}])^+] \leq E[(r - X)^+]$$

for any $r \in \mathbb{R}$. Hence, Lemma 1.30 implies that the first inequality holds for $\rho := AV@R_\lambda$. But this is enough, due to Theorem 1.43. The second inequality follows from the first by taking $\mathcal{G} = \{\emptyset, \Omega\}$. \square

In contrast to Proposition 1.27, the following theorem shows that $AV@R_\lambda$ is the best conservative approximation to $V@R_\lambda$ in the class of all *law-invariant* convex measures of risk which are continuous from above, given our standing assumption that (Ω, \mathcal{F}, P) is atomless. This result is due to Delbaen [20]

Theorem 1.46. *$AV@R_\lambda$ is the smallest law-invariant convex measure dominating $V@R_\lambda$.*

Proof. That $AV@R_\lambda$ dominates $V@R_\lambda$ was already stated in (26). Suppose now that ρ is another law-invariant convex risk measure which dominates $V@R_\lambda$ and which is continuous from above. We must show that, for a given $X \in L^\infty$,

$$\rho(X) \geq AV@R_\lambda(X). \quad (38)$$

Take $\varepsilon > 0$, and let $A := \{-X \geq V@R_\lambda(X) - \varepsilon\}$ and

$$Y := E[X | X \mathbb{1}_{A^c}] = X \cdot \mathbb{1}_{A^c} + E[X | A] \cdot \mathbb{1}_A.$$

Since $Y > q_X^+(\lambda) + \varepsilon \geq E[X | A]$ on A^c , we get $P[Y < E[X | A]] = 0$. On the other hand, $P[Y \leq E[X | A]] \geq P[A] > \lambda$, and this implies that $V@R_\lambda(Y) = E[-X | A]$. Since ρ dominates $V@R_\lambda$, we have $\rho(Y) \geq E[-X | A]$. Thus,

$$\rho(X) \geq \rho(Y) = E[-X | -X \geq V@R_\lambda(X) - \varepsilon],$$

by Corollary 1.45. Taking $\varepsilon \downarrow 0$ yields

$$\rho(X) \geq E[-X | -X \geq V@R_\lambda(X)].$$

If the distribution of X is continuous, Corollary 1.34 states that the conditional expectation on the right equals $AV@R_\lambda(X)$, and we obtain (38). If the distribution of X is not continuous, we denote by D the set of all points x such that $P[X = x] > 0$ and take any bounded random variable $Z \geq 0$ with a continuous distribution. Such a random variable exists due to our assumption that (Ω, \mathcal{F}, P) is atomless. Note that $X_n := X + \frac{1}{n} Z \mathbb{1}_{\{X \in D\}}$ has a continuous distribution. Indeed, for any y ,

$$P[X_n = y] = P[Y = y, Y \notin D] + \sum_{x \in D} P[Y = x, Z = n(y - x)] = 0.$$

Moreover, X_n decreases to X . The inequality (38) holds for each X_n and extends to X by continuity from above. \square

Corollary 1.47. *$AV@R_\lambda$ and WCE_λ coincide under our standing assumption that the probability space is atomless.*

Proof. We know from Corollary 1.34 that $WCE_\lambda(X) = AV@R_\lambda(X)$ if X has a continuous distribution. Repeating the approximation argument at the end of the preceding proof yields $WCE_\lambda(X) = AV@R_\lambda(X)$ for each $X \in L^\infty$. \square

Since $AV@R_\lambda$ is coherent, continuous from below, and law-invariant, any mixture

$$\rho_\mu(X) := \int AV@R_\lambda \mu(d\lambda) \quad (39)$$

for some probability measure μ on $(0, 1]$ has the same properties. According to Remark 1.29, we may set $AV@R_0(X) = -\text{ess inf } X$ so that we can extend the definition (39) to probability measures μ on the

closed interval $[0, 1]$. However, ρ_μ will only be continuous from above and not from below if $\mu(\{0\}) > 0$, because $AV@R_0$ is not continuous from below. Our next goal is to derive a representation of the risk measure ρ_μ in terms of the Choquet integral with respect to the set function $c_\psi(A) := \psi(P[A])$, where ψ is the nonlinear function constructed in the following lemma.

Lemma 1.48. *By defining $\psi(0) = 0$ and*

$$\psi(x) = \mu(\{0\}) + \int_0^x \int_{(t,1]} s^{-1} \mu(ds) dt, \quad 0 < x \leq 1, \quad (40)$$

we get a one-to-one correspondence between probability measures μ on $[0, 1]$ and increasing concave functions $\psi : [0, 1] \rightarrow [0, 1]$ with $\psi(0) = 0$ and $\psi(1) = 1$.

Proof. Suppose first that μ is given and ψ is defined by (40). Then ψ is concave and increasing on $(0, 1]$. Moreover,

$$\psi(1) = \mu(\{0\}) + \int_{(0,1]} \frac{1}{s} \int_0^1 \mathbb{I}_{\{t < s \leq 1\}} dt \mu(ds) = \mu(\{0\}) + \mu((0, 1]) = 1.$$

Conversely, if ψ is given, then its right-hand derivative $\psi'_+(t)$ is a decreasing right-continuous function on $(0, 1)$ and can be written as $\psi'_+(t) = \nu((t, 1])$ for some locally finite positive measure ν on $(0, 1]$. We first define μ on $(0, 1]$ by $\mu(dt) = t\nu(dt)$. Then, by Fubini's theorem,

$$\mu((0, 1]) = \int_0^1 \int_{(0,1]} \mathbb{I}_{\{t < s\}} \nu(ds) dt = 1 - \psi(0+) \leq 1.$$

Hence, setting $\mu(\{0\}) := \psi(0+)$ defines a probability measure μ on $[0, 1]$, for which (40) holds. \square

Theorem 1.49. *For a probability measure μ on $[0, 1]$, let ψ be the concave function defined in Lemma 1.48. Then, for $X \in L^\infty$,*

$$\begin{aligned} \rho_\mu(-X) &= \psi(0+)AV@R_0(-X) + \int_0^1 q_X(t)\psi'(1-t) dt \\ &= \int_{-\infty}^0 (\psi(P[X > x]) - 1) dx + \int_0^\infty \psi(P[X > x]) dx. \end{aligned}$$

Proof. Using the fact that $V@R_\lambda(-X) = q_X^-(1-\lambda)$, we get as in (37) that

$$\int_{(0,1]} AV@R_\lambda(-X) \mu(d\lambda) = \int_0^1 q_X(t)\psi'(1-t) dt.$$

Hence, we obtain the first identity. For the second one, we will first assume $X \geq 0$. Then, using (29) and Fubini's theorem, we obtain

$$\begin{aligned} \int_0^1 q_X(t)\psi'(1-t) dt &= \int_0^\infty \int_0^1 \mathbb{I}_{\{F_X(x) \leq 1-t\}} \psi'(t) dt dx \\ &= \int_0^\infty \psi(1 - F_X(x)) dx - \psi(0+) \text{ess sup } X, \end{aligned}$$

since $\int_0^y \psi'(t) dt = (\psi(y) - \psi(0+))\mathbb{I}_{\{y > 0\}}$. This proves the second identity for $X \geq 0$, since $\psi(0+) = \mu(\{0\})$ and $\text{ess sup } X = AV@R_0(-X)$. If $X \in L^\infty$ is arbitrary, we consider $X+C$, where $C := -\text{ess inf } X$. The cash invariance of ρ_μ yields

$$\begin{aligned} C + \rho_\mu(-X) &= \int_0^\infty \psi(P[X > x - C]) dx \\ &= C + \int_{-\infty}^0 (\psi(P[X > x]) - 1) dx + \int_0^\infty \psi(P[X > x]) dx. \end{aligned}$$

\square

Example 1.50. Clearly, the risk measure $AV@R_\lambda$ is itself of the form ρ_μ where $\mu = \delta_\lambda$. For $\lambda > 0$, the corresponding concave distortion function is given by

$$\psi(t) = \left(\frac{t}{\lambda}\right) \wedge 1 = \frac{1}{\lambda} (t \wedge \lambda).$$

Thus, we obtain yet another representation of $AV@R_\lambda$:

$$AV@R_\lambda(-X) = \frac{1}{\lambda} \int_0^\infty P[X > x] \wedge \lambda dx \quad \text{for } X \in L_+^\infty.$$

◇

As another consequence of Theorem 1.49, we obtain an explicit description of the maximal representing set $\mathcal{Q}_\mu \subset \mathcal{M}_1(P)$ for the coherent risk measure ρ_μ , which was first obtained by Carlier and Dana [11] in the case of a sufficiently regular distortion function ψ .

Theorem 1.51. *Let μ be a probability measure on $[0, 1]$, and let ψ be the corresponding concave function defined in Lemma 1.48. Then ρ_μ can be represented as*

$$\rho_\mu(X) = \sup_{Q \in \mathcal{Q}_\mu} E_Q[-X],$$

where the set \mathcal{Q}_μ is given by

$$\begin{aligned} \mathcal{Q}_\mu &:= \left\{ Q \in \mathcal{M}_1(P) \mid Q[A] \leq \psi(P[A]) \text{ for all } A \in \mathcal{F} \right\} \\ &= \left\{ Q \in \mathcal{M}_1(P) \mid \varphi := \frac{dQ}{dP} \text{ satisfies } \int_t^1 q_\varphi(s) ds \leq \psi(1-t) \text{ for } t \in (0, 1) \right\}. \end{aligned}$$

Moreover, \mathcal{Q}_μ is the maximal subset of $\mathcal{M}_1(P)$ that represents ρ_μ .

Proof. The risk measure ρ_μ is coherent and continuous from above. By Corollary 1.16, it can be represented by taking the supremum of expectations over the set $\mathcal{Q}_{\max} = \{Q \in \mathcal{M}_1(P) \mid \alpha_{\min}(Q) = 0\}$. It from (13) that $Q \in \mathcal{Q}_{\max}$ if and only if $E_Q[-X] \leq \rho_\mu(X)$ for all $X \in L^\infty$. By the second identity in Theorem 1.49, this condition is equivalent to $Q[A] \leq \psi(P[A])$ for all $A \in \mathcal{F}$.

In order to get the second representation of \mathcal{Q}_μ , we use (28) and the first identity in Theorem 1.49 to see that a measure $Q \in \mathcal{M}_1(P)$ with density $\varphi = dQ/dP$ belongs to \mathcal{Q}_{\max} if and only if

$$\int_0^1 q_X(s) q_\varphi(s) ds \leq \rho_\mu(-X) = \psi(0+) AV@R_0(-X) + \int_0^1 q_X(s) \psi'(1-s) ds \quad (41)$$

for all $X \in L^\infty$. For constant random variables $X \equiv t$, we have $q_X = \mathbb{I}_{[t,1]}$ a.e., and so we obtain

$$\int_t^1 q_\varphi(s) ds \leq \psi(0+) + \int_t^1 \psi'(1-s) ds = \psi(1-t)$$

for all $t \in (0, 1)$. Hence $\mathcal{Q}_{\max} \subset \mathcal{Q}_\mu$. For the proof of the converse inclusion, we show that the density φ of a fixed measure $Q \in \mathcal{Q}_\mu$ satisfies (41) for any given $X \in L^\infty$. To this end, let ν be the positive finite measure on $[0, 1]$ such that $q_X^\pm(s) = \nu([0, s])$. Using Fubini's theorem and the definition of \mathcal{Q}_μ , we get

$$\begin{aligned} \int_0^1 q_X(s) q_\varphi(s) ds &= \int_{[0,1]} \int_t^1 q_\varphi(s) ds \nu(dt) \\ &\leq \int_{[0,1]} \psi(1-t) \nu(dt) \\ &= \psi(0+) \nu([0, 1]) + \int_0^1 \psi'(1-s) \int_{[0,s]} \nu(dt) ds, \end{aligned}$$

which coincides with the right-hand side of (41). □

1.5 Comonotonic law-invariant risk measures

In many situations, the risk $\rho(X + Y)$ of a combined position will be strictly lower than the sum of the individual risks $\rho(X)$ and $\rho(Y)$ because one position serves as a hedge against adverse changes in the other position. If, on the other hand, there is no way for X to work as a hedge for Y then we may want the risk simply to add up. In order to make this idea precise, we introduce the notion of comonotonicity. Our main goal in this section is to show that a law-invariant convex risk measure ρ is comonotonic if and only if it is of the form

$$\rho_\mu(X) := \int AV@R_\lambda(X) \mu(d\lambda)$$

for some probability measure μ on $[0, 1]$. In other words, comonotonicity characterizes those law-invariant convex risk measures which quantify the risk of a position as the expected loss with respect to a concave distortion of the underlying probability measure P . Law-invariant comonotonic coherent risk measures were first characterized by Kusuoka [53]. The reader can find further results on comonotonic risk measures in [20, 21] and [35].

Definition 1.52. Two measurable functions X and Y on (Ω, \mathcal{F}) are called *comonotone* if

$$(X(\omega) - X(\omega'))(Y(\omega) - Y(\omega')) \geq 0 \quad \text{for all } (\omega, \omega') \in \Omega \times \Omega. \quad (42)$$

A monetary measure of risk ρ is called *comonotonic* if

$$\rho(X + Y) = \rho(X) + \rho(Y)$$

whenever X and Y are comonotone.

Lemma 1.53. *If ρ is a monetary measure of risk defined on the space of bounded measurable functions and if ρ is comonotonic, then ρ is positively homogeneous.*

Proof. Note that (X, X) is a comonotone pair. Hence $\rho(2X) = 2\rho(X)$. An iteration of this argument yields $\rho(rX) = r\rho(X)$ for all rational numbers $r \geq 0$. Positive homogeneity now follows from the Lipschitz continuity of ρ ; see Lemma 1.2. \square

The following lemma is taken from Denneberg [23].

Lemma 1.54. *Two measurable functions X and Y on (Ω, \mathcal{F}) are comonotone if and only if there exists a third measurable function Z on (Ω, \mathcal{F}) and increasing functions f and g on \mathbb{R} such that $X = f(Z)$ and $Y = g(Z)$.*

Proof. Clearly, $X := f(Z)$ and $Y := g(Z)$ are comonotone for given Z , f , and g . Conversely, suppose that X and Y are comonotone and define Z by $Z := X + Y$. We show that $z := Z(\omega)$ has a unique decomposition as $z = x + y$, where $(x, y) = (X(\omega'), Y(\omega'))$ for some $\omega' \in \Omega$. Having established this, we can put $f(z) := x$ and $g(z) := y$. The existence of the decomposition as $z = x + y$ follows by taking $x := X(\omega)$ and $y := Y(\omega)$, so it remains to show that these are the only possible values x and y . To this end, let us suppose that $X(\omega) + Y(\omega) = z = X(\omega') + Y(\omega')$ for some $\omega' \in \Omega$. Then

$$X(\omega) - X(\omega') = -(Y(\omega) - Y(\omega')),$$

and comonotonicity implies that this expression vanishes. Hence $x = X(\omega')$ and $y = Y(\omega')$.

Next, we check that both f and g are increasing functions on $Z(\Omega)$. So let us suppose that

$$X(\omega_1) + Y(\omega_1) = z_1 \leq z_2 = X(\omega_2) + Y(\omega_2).$$

This implies

$$X(\omega_1) - X(\omega_2) \leq -(Y(\omega_1) - Y(\omega_2)).$$

Comonotonicity thus yields that $X(\omega_1) - X(\omega_2) \leq 0$ and $Y(\omega_1) - Y(\omega_2) \leq 0$, whence $f(z_1) \leq f(z_2)$ and $g(z_1) \leq g(z_2)$. Thus, f and g are increasing on $Z(\Omega)$, and it is straightforward to extend them to increasing functions defined on \mathbb{R} . \square

The following lemma implies in particular that $V@R_\lambda$ and $AV@R_\lambda$ are comonotonic.

Lemma 1.55. *If X and Y is a pair of comonotone random variables on (Ω, \mathcal{F}, P) , then $q_X + q_Y$ is a quantile function for $X + Y$. In particular,*

$$q_{X+Y}(t) = q_X(t) + q_Y(t) \quad \text{for a.e. } t.$$

Proof. By Lemma 1.54, $X = f(Z)$ and $Y = g(Z)$ for some random variable Z and increasing functions f and g . Applying Lemma 1.38 to the increasing function $h := f + g$ shows that $h(q_Z) = f(q_Z) + g(q_Z)$ is a quantile function for $X + Y$. Another application of the same lemma yields that $q_X + q_Y$ is a quantile function for $X + Y$. \square

The following theorem shows on the one hand that the risk measures $AV@R_\lambda$ may be viewed as the extreme points in the convex class of all comonotonic law-invariant convex risk measures on L^∞ that are continuous from above. This part of the result was first proved by Kusuoka [53]. The theorem also shows that every comonotonic law-invariant convex risk measure is automatically continuous from above, and this fact was first observed by Kunze [52].

Theorem 1.56. *On an atomless probability space, the class of risk measures*

$$\rho_\mu(X) := \int AV@R_\lambda(X) \mu(d\lambda), \quad \mu \in \mathcal{M}_1([0, 1]),$$

is precisely the class of all law-invariant convex risk measures on L^∞ that are comonotonic. In particular, any convex measure of risk that is law-invariant and comonotonic is also coherent and continuous from above.

Proof. Comonotonic additivity of ρ_μ follows from Lemma 1.55 and the first representation in Theorem 1.49.

Now assume that ρ is a law-invariant convex measure of risk that is also comonotonic. Then ρ is a coherent risk measure by Lemma 1.53 and hence subadditive. Consider the set function $A \mapsto \rho(-\mathbb{I}_A)$. Since ρ is law-invariant and coherent, there exists an increasing function ψ on $[0, 1]$ such that $\psi(0) = 0$, $\psi(1) = 1$, and $\rho(-\mathbb{I}_A) = \psi(P[A]) =: c_\psi(A)$. Note that $\mathbb{I}_{A \cup B}$ and $\mathbb{I}_{A \cap B}$ is a pair of comonotone functions for all $A, B \in \mathcal{F}$. Hence, comonotonicity and subadditivity of ρ imply

$$\begin{aligned} c_\psi(A \cap B) + c_\psi(A \cup B) &= \rho(-\mathbb{I}_{A \cap B}) + \rho(-\mathbb{I}_{A \cup B}) = \rho(-\mathbb{I}_{A \cap B} - \mathbb{I}_{A \cup B}) \\ &= \rho(-\mathbb{I}_A - \mathbb{I}_B) \\ &\leq c_\psi(A) + c_\psi(B). \end{aligned} \tag{43}$$

To verify the concavity of ψ , we shall show that $\psi(y) \geq (\psi(x) + \psi(z))/2$ whenever $0 \leq x \leq z \leq 1$ and $y = (x + z)/2$. To this end, we will construct two sets $A, B \subset \mathcal{F}$ such that $P[A] = P[B] = y$, $P[A \cap B] = x$, and $P[A \cup B] = z$. We then get $\psi(x) + \psi(z) \leq 2\psi(y)$ from (43) and in turn the concavity of ψ . In order to construct the two sets A and B , take a random variable U with a uniform distribution on $[0, 1]$, which exists due to our assumption that (Ω, \mathcal{F}, P) is atomless. Then

$$A := \{0 \leq U \leq y\} \quad \text{and} \quad B := \{z - y \leq U \leq z\}$$

are as desired.

Theorem 1.49 shows that the Choquet integral with respect to c_ψ can be identified with a risk measure ρ_μ , where μ is obtained from ψ via Lemma 1.48. Let us now show that ρ and ρ_μ coincide on simple random variables of the form

$$X = \sum_{i=1}^n a_i \mathbb{I}_{A_i}, \quad a_i \in \mathbb{R}, A_i \in \mathcal{F}.$$

Since these random variables are dense in L^∞ , Lemma 1.2 will then imply that $\rho = \rho_\mu$. In order to show that $\rho_\mu(X) = \rho(X)$ for X as above, we may assume without loss of generality that $a_1 \geq a_2 \geq \dots \geq a_n$ and that the sets A_i are disjoint. Thus, we can write $X = \sum_{i=1}^n b_i \mathbb{1}_{B_i}$, where $b_i := a_i - a_{i+1}$, $a_{n+1} := 0$, and $B_i := \bigcup_{k=1}^i A_k$. Since $\sum_{i=1}^{k-1} b_i \mathbb{1}_{B_i}$ and $b_k \mathbb{1}_{B_k}$ are comonotone and $\rho(-\mathbb{1}_A) = c_\psi(A) = \rho_\mu(-\mathbb{1}_A)$,

$$\rho(-X) = \sum_{i=1}^n b_i \rho(-\mathbb{1}_{B_i}) = \sum_{i=1}^n b_i \rho_\mu(-\mathbb{1}_{B_i}) = \rho_\mu(-X).$$

□

Remark 1.57. Let $\psi : [0, 1] \rightarrow [0, 1]$ be an increasing function with $\psi(0) = 0$ and $\psi(1) = 1$. The preceding proof shows that the concavity of ψ is equivalent to the fact that the set function $c_\psi(A) := \psi(P[A])$ is *submodular* or *2-alternating* in the sense of Choquet:

$$c_\psi(A \cap B) + c_\psi(A \cup B) \leq c_\psi(A) + c_\psi(B) \quad \text{for } A, B \in \mathcal{F}.$$

This property of submodularity will play an important role in Sections 2.3 and 2.4. ◇

1.6 Risk measures arising from robust preferences

In this section, we will see how risk measures arise in a natural way from numerical representations of the preferences of an investor. As a motivation, let us first consider the following simple thought experiment. Suppose the investor is offered a bet β_1 that pays off +1000 € or -1000 €, both with known probability $p = 0.5$. The alternative would be to reject the offered bet, and this could be regarded as accepting the “bet” β_2 with the certain payoff 0 €, which is also identical to the expected payoff of the risky bet β_1 . An investor who is *risk-averse* will thus prefer β_2 over the risky bet β_1 . Now consider a third bet β_3 that also yields either +1000 € or -1000 €, but this time we assume that no information on the success probability is provided. That is, the investor is facing *model uncertainty* or *ambiguity*, which is sometimes also called *Knightian uncertainty*. Although the possible payoffs of β_3 and β_1 are identical, it is reasonable to assign some value to the information on the success probability given for β_1 . Hence, β_1 should be preferred over β_3 . That is, the underlying decision rule should exhibit a feature one might call *uncertainty aversion*. In this section, our aim is to outline a corresponding theory of choice that was developed by Schmeidler [65] and Gilboa and Schmeidler [40]. In particular, we wish to highlight its connections to coherent risk measures.

The general aim of a theory of choice is to give an axiomatic foundation and corresponding mathematical representation theory for a normative decision rule by means of which one can reach decisions when presented with several alternatives. Our starting point is the classical theory of expected utility as developed by John von Neumann and Oscar Morgenstern; see, e.g., Kreps [51] or [34, 35, Chapter 2] for introductions. It deals with monetary bets whose outcome probabilities are known. Such a bet can be regarded as a Borel probability measure μ on \mathbb{R} . More precisely, we will consider here the space

$$\mathcal{M}_b := \mathcal{M}_b(\mathbb{R}) = \{ \mu \in \mathcal{M}_1(\mathbb{R}) \mid \mu([-c, c]) = 1 \text{ for some } c \geq 0 \}$$

of boundedly supported Borel probability measures. The decision rule is usually taken as a *preference relation* or *preference order* \succ on \mathcal{M}_b , i.e., \succ is a binary relation on \mathcal{M}_b that is *asymmetric*

$$\mu \succ \nu \quad \Rightarrow \quad \nu \not\succeq \mu$$

and *negative transitive*

$$\mu \succ \nu \text{ and } \lambda \in \mathcal{M}_b \quad \Rightarrow \quad \mu \succ \lambda \text{ or } \lambda \succ \nu;$$

see, e.g., [51] or [34, 35, Chapter 2] for details. The corresponding *weak preference order* $\mu \succeq \nu$ is defined via $\mu \succeq \nu \Leftrightarrow \nu \not\succeq \mu$. If both $\mu \succeq \nu$ and $\nu \succeq \mu$ hold, we will write $\mu \sim \nu$. Dealing with a preference order is greatly facilitated if one has a *numerical representation*, namely a function $R : \mathcal{M}_b \rightarrow \mathbb{R}$ such that

$$\mu \succ \nu \quad \Leftrightarrow \quad R(\mu) > R(\nu).$$

John von Neumann and Oscar Morgenstern formulated a set of axioms that are necessary and sufficient for the existence of a numerical representation R of *von Neumann-Morgenstern form*, that is,

$$R(\mu) = \int U(x) \mu(dx) \quad (44)$$

for a function $U : \mathbb{R} \rightarrow \mathbb{R}$. The two main axioms are:

- the *Archimedean axiom*: for any triple $\mu \succ \lambda \succ \nu$ there are $\alpha, \beta \in (0, 1)$ such that $\alpha\mu + (1 - \alpha)\nu \succ \lambda \succ \beta\mu + (1 - \beta)\nu$;
- the *independence axiom*: for all $\mu, \nu \in \mathcal{M}$, the relation $\mu \succ \nu$ implies $\alpha\mu + (1 - \alpha)\lambda \succ \alpha\nu + (1 - \alpha)\lambda$ for all $\lambda \in \mathcal{M}$ and all $\alpha \in (0, 1]$.

These two axioms are equivalent to the existence of an affine numerical representation R . To obtain an integral representation (44) for this affine functional on \mathcal{M}_b one needs some additional regularity condition such as topological assumptions on the level sets of \succ ; see [51] and [34, 35, Chapter 2].

The preference order is called *monotone* if

$$y > x \quad \iff \quad \delta_y \succ \delta_x.$$

Clearly, monotonicity holds if and only if the function U in (44) is strictly increasing. One says that the preference order exhibits *risk aversion* if the certain amount $m(\mu) := \int x \mu(dx)$ is preferred over the risky lottery μ , i.e.,

$$\delta_{m(\mu)} \succ \mu \quad \text{for any } \mu \in \mathcal{M}_b \text{ such that } \mu \neq \delta_{m(\mu)}.$$

Risk aversion is equivalent to the strict concavity of U , and if U is both increasing and strictly concave, it is called a *utility function*.

Now we wish to extend this setting to the case in which the probabilities of outcomes may be subject to uncertainty. This is achieved by randomizing lotteries via an exterior probability space (Ω, \mathcal{F}) . More precisely, we will consider a space $\tilde{\mathcal{X}}$ defined as the set of all stochastic kernels $\tilde{X}(\omega, dy)$ from (Ω, \mathcal{F}) to \mathbb{R} for which there exists a constant $c \geq 0$ such that

$$\tilde{X}(\omega, [-c, c]) = 1 \quad \text{for all } \omega \in \Omega.$$

In mathematical economics, the elements of $\tilde{\mathcal{X}}$ are sometimes called *horse race lotteries*. The space of standard lotteries \mathcal{M}_b has a natural embedding into $\tilde{\mathcal{X}}$ by identifying $\mu \in \mathcal{M}_b$ with the constant map $\tilde{X}(\omega) = \mu$ for all ω .

Now consider a given preference order \succ on $\tilde{\mathcal{X}}$. We will assume that \succ is compatible with the embedding $\mathcal{M}_b \subset \tilde{\mathcal{X}}$ in the sense that

$$\tilde{Y} \succeq \tilde{X} \quad \text{if } \tilde{Y}(\omega) \succeq \tilde{X}(\omega) \text{ for all } \omega \in \Omega.$$

We will furthermore assume the following extension of the Archimedean axiom of classical von Neumann-Morgenstern theory.

- *Archimedean axiom*: if $\tilde{X}, \tilde{Y}, \tilde{Z} \in \tilde{\mathcal{X}}$ are such that $\tilde{Z} \succ \tilde{Y} \succ \tilde{X}$, then there are $\alpha, \beta \in (0, 1)$ with

$$\alpha\tilde{Z} + (1 - \alpha)\tilde{X} \succ \tilde{Y} \succ \beta\tilde{Z} + (1 - \beta)\tilde{X}.$$

We also need an extended version of the independence axiom:

- *certainty independence*: for $\tilde{X}, \tilde{Y} \in \tilde{\mathcal{X}}$, $\tilde{Z} \equiv \mu \in \mathcal{M}_b$, and $\alpha \in (0, 1]$ we have

$$\tilde{X} \succ \tilde{Y} \quad \iff \quad \alpha\tilde{X} + (1 - \alpha)\tilde{Z} \succ \alpha\tilde{Y} + (1 - \alpha)\tilde{Z}.$$

These two immediately imply that the restriction of \succ to \mathcal{M}_b satisfies both the classical Archimedean axiom and the independence axiom. Hence, it admits an affine numerical representation $R : \mathcal{M}_b \rightarrow \mathbb{R}$, and to simplify things, we will assume henceforth that R is of von Neumann-Morgenstern form (44) for some function $U : \mathbb{R} \rightarrow \mathbb{R}$.

The main new axiom is:

- *uncertainty aversion*: if $\tilde{X}, \tilde{Y} \in \tilde{\mathcal{X}}$ are such that $\tilde{X} \simeq \tilde{Y}$, then

$$\alpha \tilde{X} + (1 - \alpha) \tilde{Y} \succeq \tilde{X} \quad \text{for all } \alpha \in [0, 1].$$

In order to motivate the term ‘‘uncertainty aversion’’, consider the following simple example. For $\Omega := \{0, 1\}$ define

$$\tilde{Z}_i(\omega) := \delta_{1000} \cdot \mathbb{1}_{\{i\}}(\omega) + \delta_0 \cdot \mathbb{1}_{\{1-i\}}(\omega), \quad i = 0, 1.$$

Suppose that an agent is indifferent between the choices \tilde{Z}_0 and \tilde{Z}_1 , which both involve the same kind of uncertainty. In the case of uncertainty aversion, the convex combination $\tilde{Y} := \alpha \tilde{Z}_0 + (1 - \alpha) \tilde{Z}_1$ is weakly preferred to both \tilde{Z}_0 and \tilde{Z}_1 . It takes the form

$$\tilde{Y}(\omega) = \begin{cases} \alpha \delta_{1000} + (1 - \alpha) \delta_0 & \text{for } \omega = 1, \\ \alpha \delta_0 + (1 - \alpha) \delta_{1000} & \text{for } \omega = 0. \end{cases}$$

This convex combination now allows for upper and lower probability bounds in terms of α , and this means that model uncertainty is reduced in favor of risk. For $\alpha = 1/2$, the resulting lottery $\tilde{Y}(\omega) \equiv \frac{1}{2}(\delta_{1000} + \delta_0)$ is independent of the scenario ω , i.e., model uncertainty is completely eliminated.

Theorem 1.58. *Under the above conditions, there exists a unique extension of R to a numerical representation $\tilde{R} : \mathcal{X} \rightarrow \mathbb{R}$, and \tilde{R} is of the form*

$$\tilde{R}(\tilde{X}) = \min_{Q \in \mathcal{Q}} E_Q[R(\tilde{X})] = \min_{Q \in \mathcal{Q}} E_Q \left[\int U(x) \tilde{X}(\cdot, dx) \right]$$

for a convex set $\mathcal{Q} \subset \mathcal{M}_{1,f}(\Omega, \mathcal{F})$.

A proof of this theorem can be found in [40] or in [34, 35, Section 2.5].

Remark 1.59. Let us comment on the axiom of certainty independence. It extends the independence axiom for preferences on \mathcal{M}_b to our present setting, but only under the restriction that one of the two contingent lotteries \tilde{X} and \tilde{Y} is *certain*, i.e., does not depend on the scenario $\omega \in \Omega$. Without this restriction, the extended independence axiom would lead to a so-called *Savage representation*

$$\tilde{R}(\tilde{X}) = E_Q[R(\tilde{X})]$$

in terms of a *subjective measure* Q . But there are good reasons for *not* requiring full independence for all $\tilde{Z} \in \tilde{\mathcal{X}}$. As an example, take $\Omega = \{0, 1\}$ and define $\tilde{X}(\omega) = \delta_\omega$, $\tilde{Y}(\omega) = \delta_{1-\omega}$, and $\tilde{Z} = \tilde{X}$. An agent may prefer \tilde{X} over \tilde{Y} , thus expressing the implicit view that scenario 1 is somewhat more likely than scenario 0. At the same time, the agent may like the idea of *hedging* against the occurrence of scenario 0, and this could mean that the certain lottery

$$\frac{1}{2}(\tilde{Y} + \tilde{Z}) \equiv \frac{1}{2}(\delta_0 + \delta_1)$$

is preferred over the contingent lottery

$$\frac{1}{2}(\tilde{X} + \tilde{Z}) \equiv \tilde{X},$$

thus violating the independence assumption in its unrestricted form. In general, the role of \tilde{Z} as a hedge against scenarios unfavorable for \tilde{Y} requires that \tilde{Y} and \tilde{Z} are *not* comonotone, where comonotonicity means:

$$\tilde{Y}(\omega) \succeq \tilde{Y}(\eta) \iff \tilde{Z}(\omega) \succeq \tilde{Z}(\eta). \quad (45)$$

Thus, the wish to hedge would still be compatible with the following stronger version of certainty independence, called

- *comonotonic independence*: For $\tilde{X}, \tilde{Y}, \tilde{Z} \in \tilde{\mathcal{X}}$ and $\alpha \in (0, 1]$

$$\tilde{X} \succ \tilde{Y} \iff \alpha \tilde{X} + (1 - \alpha) \tilde{Z} \succ \alpha \tilde{Y} + (1 - \alpha) \tilde{Z}$$

whenever \tilde{Y} and \tilde{Z} are comonotone.

This stronger requirement holds iff the set \mathcal{Q} in the preceding theorem is such that $\gamma(A) := \sup_{Q \in \mathcal{Q}} Q[A]$ is *submodular*:

$$\gamma(A \cup B) + \gamma(A \cap B) \leq \gamma(A) + \gamma(B);$$

see Schmeidler [65], p. 582, or [35, Section 4.7]. Compare also with Remark 1.57. \diamond

The space \mathcal{X} of all bounded measurable function on (Ω, \mathcal{F}) can be embedded into $\tilde{\mathcal{X}}$ by virtue of the mapping

$$\mathcal{X} \ni X \longmapsto \delta_X \in \tilde{\mathcal{X}}. \quad (46)$$

In this way, \mathcal{X} can be identified with the set of all uncertain payoffs. The preceding theorem implies that the restriction of \succ to \mathcal{X} admits the numerical representation

$$R_{\mathcal{X}}(X) = \inf_{Q \in \mathcal{Q}} E_Q[U(X)], \quad (47)$$

and it is this representation in which we are really interested. Note, however, that it is necessary to it is necessary to formulate the axiom of uncertainty aversion on the larger space of uncertain lotteries. But even without its axiomatic foundation, such a representation of preferences in the face of model uncertainty by a subjective utility assessment $R_{\mathcal{X}}(X)$ is highly plausible as it stands. It may be viewed as a *robust* approach to the problem of model uncertainty: The agent has in mind a whole collection of possible probabilistic views of the given set of scenarios and takes a *worst case approach* in evaluating the payoff of a given financial position.

Let us now emphasize the downside rather than the upside by switching from a utility functional $R_{\mathcal{X}}$ to the associated *loss functional* $L := -R_{\mathcal{X}}$, and let us assume that U is a utility function. Then our representation takes the form

$$L(X) = \sup_{Q \in \mathcal{Q}} E_Q[\ell(-X)]$$

where ℓ denotes the convex increasing *loss function* on \mathbb{R} defined by $\ell(x) = -U(-x)$. Now suppose that an agent finds a position X *acceptable* if $L(X)$ does not exceed a given bound x_0 . How do we determine the amount of capital which is needed to turn a given position $X \in \mathcal{X}$ into an acceptable position by adding this amount? In order to answer this question, consider the convex acceptance set

$$\mathcal{A}_L := \left\{ X \in \mathcal{X} \mid L(X) = \sup_{Q \in \mathcal{Q}} E_Q[\ell(-X)] \leq x_0 \right\},$$

where x_0 is an interior point in the range of ℓ . Recalling Proposition 1.5, we see that \mathcal{A} induces a convex measure of risk ρ_L . Applying the results of the preceding section, we can conclude that ρ_L admits a representation of the form

$$\rho_L(X) = \sup_{Q \in \mathcal{M}_1} \left(E_Q[-X] - \alpha_L(Q) \right).$$

Thus, the problem is reduced to the computation of a suitable penalty function α_L . To this end, let us introduce the Fenchel-Legendre transform ℓ^* of the loss function ℓ defined by

$$\ell^*(z) := \sup_{x \in \mathbb{R}} (zx - \ell(x)).$$

By combining the formula stated in Example 1.42 with Proposition 1.12, we get the following result.

Theorem 1.60. *Suppose \mathcal{Q} is a set of equivalent probability measures. Then the convex risk measure corresponding to the acceptance set \mathcal{A} can be represented in terms of the penalty function*

$$\alpha_L(P) = \inf_{\lambda > 0} \frac{1}{\lambda} \left(x_0 + \inf_{Q \in \mathcal{Q}} E_Q \left[\ell^* \left(\lambda \frac{dP}{dQ} \right) \right] \right).$$

Thus, $\alpha_L(P) < \infty$ only if $P \ll Q$ for at least some $Q \in \mathcal{Q}$.

Example 1.61. For the exponential loss function $\ell(x) = e^x$ and $x_0 = 1$, the penalty function in Theorem 1.60 takes the form

$$\alpha_L(P) = \inf_{Q \in \mathcal{Q}} H(P|Q);$$

where

$$H(P|Q) = \begin{cases} E_Q \left[\frac{dP}{dQ} \log \frac{dP}{dQ} \right] & \text{if } P \ll Q, \\ +\infty & \text{otherwise,} \end{cases}$$

denotes the *relative entropy* of P with respect to Q ; see [35].

2 Risk measures and robust optimization in financial markets

In this chapter, we will extend our setting by assuming that financial positions arise in a multivariate market model. In such a model, it may be possible to eliminate risk by using appropriate hedging strategies, and this idea should be considered when measuring the risk of positions. In Section 2.1, we show how (super-)hedgeability can be used as a criterion for acceptability of positions, and we identify the corresponding risk measures. We also discuss how hedging can be used in order to relax a given acceptability criterion. In Section 2.2, we address the problem of risk-minimal hedging when the risk criterion is defined in terms of a risk measure. In Section 2.3, we consider an optimal investment problem under uncertainty aversion. Thus, we are back in the setting of Section 1.6, and it is more natural to consider utility rather than risk, i.e., we are looking for investment strategies that maximize a robust utility functional. In particular, we give a general criterion under which the robust problem can be reduced to a standard one. In Section 2.4, we will then discuss a number of examples.

2.1 Measures of risk in a financial market with convex constraints

In this section, we will assume that positions are—at least to some extent—contingent on a financial market, and we will show how one can combine a given risk measure with the idea of risk reduction by hedging. Moreover, we will see how various arbitrage valuation principles correspond to certain convex risk measures. Thus, the concept of a risk measures can unify and combine many common static and dynamic approaches to risk.

The most popular market models in mathematical finance are based on time-continuous price processes. However, insofar arbitrage theory is concerned, these models exhibit certain pathologies that stem from the idealization of trading in continuous time. For instance, it is well known that even the standard Black-Scholes model admits arbitrage opportunities unless one excludes certain trading strategies such as doubling strategies embedded into a finite time interval. Therefore, we will restrict ourselves to discrete-time market models, for which the arbitrage theory is much simpler and easier to handle. These market models are typically *incomplete*, i.e., they involve intrinsic risks which cannot be hedged away completely. Hence, the need for combining a static risk measure with dynamic hedging arises in a natural manner. This section is based on joint work of Hans Föllmer and Schied [33, 34, 35].

We consider a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t=0, \dots, T}, P)$ and a market where one bond and d risky assets are traded. The price of the bond will be assumed to be normalized to 1, and the (correspondingly discounted) price process of the risky assets is denoted by $S_t = (S_t^1, \dots, S_t^d)$. We will assume that

$$S_t^i \geq 0 \quad \text{for } i = 1, \dots, d \text{ and } t = 0, \dots, T. \quad (48)$$

Any d -dimensional predictable process ξ gives rise to a self-financing trading strategy; ξ_t^i is the number of shares held of the i^{th} asset during the trading period $t - 1 \rightsquigarrow t$, and

$$V_t = V_0 + \sum_{k=1}^t \xi_k \cdot (S_k - S_{k-1})$$

is the associated value process for an initial investment V_0 . Here, $\xi \cdot S = \sum_{i=1}^d \xi^i S^i$ is the Euclidean scalar product. Recall that an arbitrage opportunity is a self-financing trading strategy such that $V_T \geq V_0$ and $P[V_T > V_0] > 0$. The existence of arbitrage opportunities can be regarded as a market inefficiency, and one usually insists on arbitrage-free market models. Due to the fundamental theorem of asset pricing, the market model does not admit arbitrage opportunities if and only if there exists a measure $P^* \sim P$ under which S is a martingale; see, e.g., [34, Theorem 5.17] or [35, Theorem 5.17]. The set of all these equivalent martingale measures will be denoted by \mathcal{P} .

Now consider a financial position $X \in L^\infty(P)$. X can be interpreted as “riskless” if $X \geq 0$ or, more generally, if the “risky part” of X can be hedged at no additional cost. The latter means that we can find a suitable hedging portfolio ξ such that

$$X + \sum_{t=1}^T \xi_t \cdot (S_t - S_{t-1}) \geq 0 \quad P\text{-a.s.} \quad (49)$$

Due to the fact that X is bounded, (49) is only possible if ξ is *admissible* in the sense that there is a constant $c = c(\xi)$ such that the associated gains process satisfies

$$G_T(\xi) := \sum_{t=1}^T \xi_t \cdot (S_t - S_{t-1}) \geq -c \quad P\text{-a.s.} \quad (50)$$

Thus, we define the following set of acceptable positions in L^∞ :

$$\mathcal{A}_0 := \{ X \in L^\infty \mid \exists \xi \text{ with } X + G_T(\xi) \geq 0 \text{ } P\text{-a.s.} \}.$$

Theorem 2.1. *Suppose that $\inf\{m \in \mathbb{R} \mid m \in \mathcal{A}_0\} > -\infty$. Then $\rho_0 := \rho_{\mathcal{A}_0}$ is a coherent measure of risk. Moreover, ρ_0 is sensitive in the sense of Definition 1.15 if and only if the market model is arbitrage-free, i.e., $G_T(\xi) \geq 0$ P -a.s. implies $G_T(\xi) = 0$ P -almost surely. In this case, ρ_0 is continuous from above and can be represented in terms of the set \mathcal{P} of equivalent martingale measures for the price process S :*

$$\rho_0(X) = \sup_{P^* \in \mathcal{P}} E^*[-X]. \quad (51)$$

Proof. The fact that ρ_0 is a coherent measure of risk follows from Proposition 1.6. If the model is arbitrage-free, then the superhedging duality theorem (see, e.g., [35, Corollary 7.9]) yields the representation (51), and it follows that ρ_0 is sensitive and continuous from above.

Conversely, suppose that ρ_0 is sensitive, but the market model admits an arbitrage opportunity ξ . Then there is some $k > 0$ such that $0 \leq G_T(\xi)$ P -a.s. and $P[G_T(\xi) \wedge k > 0] > 0$. It follows that $X := -G_T(\xi) \wedge k \in \mathcal{A}_0$. However, the sensitivity of ρ_0 implies that

$$\rho_0(X) > \rho_0(0) = 0.$$

Thus, we arrive at a contradiction. \square

There are several reasons why it may make sense to allow in (49) only strategies ξ that belong to a proper subset \mathcal{S} of the class of all strategies. For instance, if the resources available to an investor are limited, only those strategies should be considered for which the initial investment in risky assets is below a certain amount. Such a restriction corresponds to an upper bound on V_0 . There may be other constraints. For instance, *short sales constraints* are lower bounds on the number of shares in the portfolio. In view of *market illiquidity*, the investor may also wish to avoid holding too many shares of one single asset, since the market capacity may not suffice to resell the shares. Such constraints will be taken into account by assuming throughout the remainder of this section that \mathcal{S} has the following properties:

- (a) $0 \in \mathcal{S}$.
- (b) \mathcal{S} is *predictably convex*: If $\xi, \eta \in \mathcal{S}$ and h is a predictable process with $0 \leq h \leq 1$, then the process

$$h_t \xi_t + (1 - h_t) \eta_t, \quad t = 1, \dots, T,$$

belongs to \mathcal{S} .

(c) For each $t \in \{1, \dots, T\}$, the set

$$\mathcal{S}_t := \{ \xi_t \mid \xi \in \mathcal{S} \}$$

is closed in $L^0(\Omega, \mathcal{F}_{t-1}, P; \mathbb{R}^d)$.

(d) Each $\xi \in \mathcal{S}$ is admissible.

Moreover, we will assume that the price increments satisfy the following *non-redundance condition*: For all $t \in \{1, \dots, T\}$ and $\xi_t \in \mathcal{S}_t$,

$$\xi_t \cdot (S_t - S_{t-1}) = 0 \text{ P-a.s.} \implies \xi_t = 0 \text{ P-a.s.} \quad (52)$$

In a first step, we define the non-empty set

$$\mathcal{A}^{\mathcal{S}} := \left\{ X \in L^\infty(P) \mid \text{there exists } \xi \in \mathcal{S} \text{ with } X + G_T(\xi) \geq 0 \text{ P-a.s.} \right\}, \quad (53)$$

of acceptable positions which can be hedged with strategies in \mathcal{S} at no cost. By Proposition 1.6, $\mathcal{A}^{\mathcal{S}}$ induces the convex measure of risk

$$\rho^{\mathcal{S}}(X) := \rho_{\mathcal{A}^{\mathcal{S}}}(X) = \inf \left\{ m \in \mathbb{R} \mid m + X \in \mathcal{A}^{\mathcal{S}} \right\} \quad (54)$$

provided that

$$\rho^{\mathcal{S}}(0) > -\infty. \quad (55)$$

Note that (55) holds, in particular, if \mathcal{S} does not contain arbitrage opportunities. We will assume (55) throughout this section.

The following problems arise:

- Compute the minimal penalty function of $\rho^{\mathcal{S}}$.
- Give criteria under which $\rho^{\mathcal{S}}$ is continuous from above.

Let us consider the first problem.

Proposition 2.2. *The minimal penalty function $\alpha_{\min}^{\mathcal{S}}(\cdot)$ of $\rho^{\mathcal{S}}$ is given by*

$$\alpha_{\min}^{\mathcal{S}}(Q) = E_Q[A_T^Q] \quad \text{for } Q \in \mathcal{M}_1(P). \quad (56)$$

Here A^Q is the predictable increasing process defined by

$$A_0^Q := 0, \quad A_{t+1}^Q - A_t^Q := \operatorname{ess\,sup}_{\xi \in \mathcal{S}} \left[\xi_{t+1} \cdot \left(E_Q[S_{t+1} \mid \mathcal{F}_t] - S_t \right) \right]. \quad (57)$$

In particular, $\rho^{\mathcal{S}}$ has the representation

$$\rho^{\mathcal{S}}(X) = \sup_{Q \in \mathcal{M}_1(P)} (E_Q[-X] - E_Q[A_T^Q]) \quad (58)$$

if $\rho^{\mathcal{S}}$ is continuous from above.

Proof. First note that $E_Q[\xi_t \cdot (S_t - S_{t-1}) \mid \mathcal{F}_{t-1}]$ is well-defined and satisfies

$$E_Q[\xi_t \cdot (S_t - S_{t-1}) \mid \mathcal{F}_{t-1}] = \xi_t \cdot \left(E_Q[S_t \mid \mathcal{F}_{t-1}] - S_{t-1} \right) \quad (59)$$

for every $\xi \in \mathcal{S}$. To see this, observe first that, by predictable convexity, also $\xi^{(t)} \in \mathcal{S}$, where

$$\xi_s^{(t)} := \begin{cases} \xi_t & \text{if } s = t, \\ 0 & \text{otherwise.} \end{cases}$$

By assumption every element in \mathcal{S} is admissible in the sense of (50), and thus there is some constant c with $\xi_t \cdot (S_t - S_{t-1}) = G_T(\xi^{(t)}) \geq -c$ P-a.s. Using our assumption (48) that prices are non-negative, we

see that the conditional expectation is well defined, and we obtain (59) by adding the \mathcal{F}_{t-1} -measurable term $\xi_t \cdot S_{t-1}$ to both sides of the equation.

Next, if $X \in \mathcal{A}^S$ there exists $\xi \in \mathcal{S}$ with $-X \leq G(\xi)$ P -a.s. By using (59), we obtain that for $Q \ll P$

$$E_Q[-X] \leq E_Q[G(\xi)] \leq E_Q[A_T^Q]. \quad (60)$$

Hence, we conclude that

$$\alpha_{\min}^S(Q) = \sup_{X \in \mathcal{A}^S} E_Q[-X] \leq E_Q[A_T^Q].$$

Now we turn to the proof of the converse inequality. To this end, we show first that

$$\Psi := \left\{ \sum_{t=1}^T E_Q[\xi_t \cdot (S_t - S_{t-1}) | \mathcal{F}_{t-1}] \mid \xi \in \mathcal{S} \right\}$$

is directed upwards in the sense that for $\psi_1, \psi_2 \in \Psi$ there is $\psi_3 \in \Psi$ with $\psi_3 \geq \psi_1 \vee \psi_2$. For $\xi, \tilde{\xi} \in \mathcal{S}$ let

$$B_t := \left\{ E_Q[\xi_t \cdot (S_t - S_{t-1}) | \mathcal{F}_{t-1}] > E_Q[\tilde{\xi}_t \cdot (S_t - S_{t-1}) | \mathcal{F}_{t-1}] \right\},$$

and define $\xi'_t \in \mathcal{S}$ by

$$\xi'_t := \xi_t \mathbb{1}_{B_t} + \tilde{\xi}_t \mathbb{1}_{B_t^c}.$$

Then clearly

$$\begin{aligned} E_Q[\xi'_t \cdot (S_t - S_{t-1}) | \mathcal{F}_{t-1}] \\ = E_Q[\xi_t \cdot (S_t - S_{t-1}) | \mathcal{F}_{t-1}] \vee E_Q[\tilde{\xi}_t \cdot (S_t - S_{t-1}) | \mathcal{F}_{t-1}], \end{aligned}$$

and therefore Ψ is directed upwards. It follows that $A_T^Q = \text{ess sup } \Psi$ is the limit of an increasing sequence in Ψ ; see, e.g., [35, Appendix A.5]. Hence, by monotone convergence,

$$\begin{aligned} E_Q[A_T^Q] &= \sup_{\xi \in \mathcal{S}} E_Q \left[\sum_{t=1}^T E_Q[\xi_t \cdot (S_t - S_{t-1}) | \mathcal{F}_{t-1}] \right] \\ &= \sup_{\xi \in \mathcal{S}} \sup_{k \in \mathbb{N}} E_Q[G_T(\xi) \wedge k]. \end{aligned}$$

Admissibility yields that $-(G_T(\xi) \wedge k) \in \mathcal{A}^S \subseteq \mathcal{A}_{\rho^S}$, and thus

$$E_Q[A_T^Q] = \sup_{\xi \in \mathcal{S}} \sup_{k \in \mathbb{N}} E_Q[G_T(\xi) \wedge k] \leq \sup_{X \in \mathcal{A}_{\rho^S}} E_Q[X] = \alpha_{\min}^S(Q).$$

This concludes the proof. \square

Let us now turn to the second problem above. That is, we are looking for criteria that guarantee that ρ is continuous from above. It will turn out that the absence of arbitrage opportunities in \mathcal{S} is a sufficient condition under a certain regularity requirement. It can in turn be characterized by the following class of local supermartingale measures.

Definition 2.3. By \mathcal{P}_S we denote the class of all probability measures $\tilde{P} \sim P$ such that

$$S_t \in \mathcal{L}^1(\tilde{P}) \quad \text{for all } t, \quad (61)$$

and such that the gains process of any trading strategy in \mathcal{S} is a local \tilde{P} -supermartingale.

The following result follows from Theorem 9.29 of [35].

Theorem 2.4. *If \mathcal{P}_S is nonempty, then ρ^S is continuous from above and admits the representations*

$$\rho^S(X) = \sup_{Q \in \mathcal{M}_1(P)} (E_Q[-X] - E_Q[A_T^Q]) = \sup_{Q \in \mathcal{Q}_S} (E_Q[-X] - E_Q[A_T^Q]), \quad (62)$$

where \mathcal{Q}_S denotes the set of all $Q \sim P$ such that $E_Q[A_T^Q] < \infty$ and such that $E_Q[|S_{t+1} - S_t| | \mathcal{F}_t] < \infty$ P -a.s. for all t .

We now give a sufficient condition on \mathcal{S} to guarantee that $\mathcal{P}_{\mathcal{S}}$ is nonempty. To this end, denote the cone generated by \mathcal{S} by $\mathcal{R}_t^0 := \{\lambda \xi_t \mid \xi_t \in \mathcal{S}_t \text{ and } \lambda \geq 0\}$, and let \mathcal{R}_t be its L^0 -closure.

Theorem 2.5. *In addition to the above assumptions, suppose that $\mathcal{R}_t \cap (L^\infty)^d \subset \mathcal{R}_t^0$. Then the following conditions are equivalent:*

- (a) $\rho^{\mathcal{S}}$ is sensitive.
- (b) \mathcal{S} contains no arbitrage opportunities, i.e., for $\xi \in \mathcal{S}$, $G_T(\xi) \geq 0$ P -a.s. implies $G_T(\xi) = 0$ P -a.s.
- (c) $\mathcal{P}_{\mathcal{S}} \neq \emptyset$.

Proof. (a) \Rightarrow (b): Suppose that $\rho^{\mathcal{S}}$ is sensitive, but \mathcal{S} contains an arbitrage opportunity ξ . Then there exists $k > 0$ such that $P[G_T(\xi) \wedge k > 0] > 0$. Then $X := -G_T(\xi) \wedge k$ is bounded and satisfies $X + G(\xi) \geq 0$, i.e., $X \in \mathcal{A}$. It follows that $\rho^{\mathcal{S}}(X) \leq 0$. But this contradicts the sensitivity of $\rho^{\mathcal{S}}$ and the facts that $0 \geq X$ and $P[X < 0] > 0$.

(b) \Rightarrow (c): By standard arguments, the existence of arbitrage opportunities in \mathcal{S} is equivalent to the existence of some t and $\xi_t \in \mathcal{S}_t \cap (L^\infty)^d$ such that $\xi_t \cdot (S_t - S_{t-1}) \geq 0$ P -a.s. and $P[\xi_t \cdot (S_t - S_{t-1}) > 0] > 0$; see Lemma 9.11 in [35]. The condition $\mathcal{R}_t \cap (L^\infty)^d \subset \mathcal{R}_t^0$ hence guarantees that we may replace \mathcal{S}_t by \mathcal{R}_t . Now we can apply Theorem 9.9 of [35]; the condition $\mathcal{R}_t \cap (L^\infty)^d \subset \mathcal{R}_t^0$ is missing in the statements of Theorem 9.9 and Lemma 9.13 of [35], as was kindly pointed out to us by Konstantinos Kardaras and Sven Lickfeld.

(c) \Rightarrow (a): Due to (c) and (62), we have $\rho^{\mathcal{S}}(X) \geq \sup_{\tilde{P} \in \mathcal{P}_{\mathcal{S}}} \tilde{E}[-X]$, and this implies the sensitivity of $\rho^{\mathcal{S}}$. \square

Example 2.6. Let C_t be a closed convex subset of \mathbb{R}^d and suppose that $0 \in C_t$. Take \mathcal{S} as the class of all predictable processes ξ such that $\xi_t \in C_t$ P -a.s. for all t . Then \mathcal{S} satisfies our conditions (a) through (d). If, in addition, the cones generated by the convex sets C_t are closed in \mathbb{R}^d , then the condition $\mathcal{R}_t \cap (L^\infty)^d \subset \mathcal{R}_t^0$ is also satisfied. This case includes short sales constraints and constraints on the size of a long position. These types of constraints are modeled by taking $C_t = [a_t^1, b_t^1] \times \cdots \times [a_t^d, b_t^d]$ for certain numbers a_t^k, b_t^k such that $-\infty \leq a_t^k \leq 0 \leq b_t^k \leq \infty$.

Example 2.7. Let \mathcal{S} denote the set of all predictable ξ such that $a \leq \xi_t \cdot S_{t-1} \leq b$ for certain numbers a, b such that $-\infty \leq a < 0 < b \leq \infty$. This class \mathcal{S} corresponds to constraints on the capital invested into risky assets. It satisfies conditions (a) through (d). We claim that \mathcal{S} does not contain arbitrage opportunities if and only if the unconstrained market is arbitrage-free, so that we have $\mathcal{P}_{\mathcal{S}} = \mathcal{P}$. To prove this, note that the existence of an arbitrage opportunity in the unconstrained market is equivalent to the existence of some t and some \mathcal{F}_{t-1} -measurable ξ_t such that $\xi_t \cdot (S_t - S_{t-1}) \geq 0$ P -a.s. and $P[\xi_t \cdot (S_t - S_{t-1}) > 0] > 0$ (see Proposition 5.11 in [35]). Next, there exists a constant $c > 0$ such that these properties are shared by $\tilde{\xi}_t := \xi_t \mathbb{I}_{\{|\xi_t \cdot S_{t-1}| \leq c\}}$ and in turn by $\varepsilon \tilde{\xi}_t$, where $\varepsilon > 0$. But $\varepsilon \tilde{\xi}_t \in \mathcal{S}_t$ if ε is small enough.

Example 2.8. Suppose that \mathcal{S} consists of all bounded predictable processes with non-negative components, and that $\mathcal{P}_{\mathcal{S}} \neq \emptyset$. Then

$$\rho^{\mathcal{S}}(X) = \sup_{\tilde{P} \in \mathcal{P}_{\mathcal{S}}} \tilde{E}[X].$$

To see this, note first that A_T^Q can only take the values 0 and $+\infty$, due to the fact that \mathcal{S} is a cone. Hence, $E_Q[A_T^Q] = 0$ for all $Q \in \mathcal{Q}_{\mathcal{S}}$. Moreover, $G := G(\xi)$ is a local Q -supermartingale for $Q \in \mathcal{Q}_{\mathcal{S}}$ and $\xi \in \mathcal{S}$. To prove this, denote by $\tau_n(\omega)$ the first time t at which

$$|\xi_{t+1}(\omega)| > n \quad \text{or} \quad E_Q[|S_{t+1} - S_t| \mid \mathcal{F}_t](\omega) > n.$$

If such a t does not exist, let $\tau_n(\omega) := T$. Then τ_n is a stopping time. Since

$$|G_{(t+1) \wedge \tau_n} - G_{t \wedge \tau_n}| \leq \mathbb{I}_{\{\tau_n \geq t+1\}} |\xi_{t+1}| \cdot |S_{t+1} - S_t|,$$

$G_{\tau_n \wedge t}$ belongs to $\mathcal{L}^1(Q)$, and

$$\begin{aligned} E_Q[G_{(t+1) \wedge \tau_n} - G_{t \wedge \tau_n} \mid \mathcal{F}_t] &= \mathbb{1}_{\{\tau_n \geq t+1\}} \xi_{t+1} \cdot (E_Q[S_{t+1} \mid \mathcal{F}_t] - S_t) \\ &\leq A_{(t+1) \wedge \tau_n}^Q - A_{t \wedge \tau_n}^Q = 0. \end{aligned}$$

This proves that G_{τ_n} is a Q -supermartingale, i.e., $Q \in \mathcal{P}_S$.

Remark 2.9. In a continuous-time financial market model where the price process S follows a special semimartingale under P , one can similarly define a predictably convex set \mathcal{S} of admissible integrands and a corresponding convex measure of risk ρ . If one assumes in addition that the set $\{\int \xi dS \mid \xi \in \mathcal{S}\}$ is closed in the semimartingale or Émery topology, the optional decomposition theorem of [31] will imply a representation (62) of ρ . The penalty function $\alpha(Q)$ can be described as $\alpha(Q) = E_Q[A_T^Q]$ provided that Q satisfies the following three conditions are fulfilled: Q is equivalent to P , every process $\int \xi dS$ with $\xi \in \mathcal{S}$ is a special semimartingale under Q , and Q admits the upper variation process A^Q for the set $\{\int \xi dS \mid \xi \in \mathcal{S}\}$. One can take $\alpha(Q) = \infty$ for measures Q which do not satisfy one of these conditions.

Let us now relax the condition of acceptability in (53). We no longer insist that the final outcome of an acceptable position, suitably hedged, should always be non-negative. Instead, we only require that the hedged position is acceptable in terms of a given convex risk measure $\rho_{\mathcal{A}}$ with acceptance set \mathcal{A} . Since we cannot be sure that the hedged position is bounded from above, we define

$$\bar{\mathcal{A}} := \{X \in L^\infty \mid \exists \xi \in \mathcal{S}, A \in \mathcal{A} \text{ with } X + G_T(\xi) \geq A \text{ } P\text{-a.s.}\}. \quad (63)$$

If $\rho_{\mathcal{A}}$ is normalized, then $\mathcal{A} \subset \bar{\mathcal{A}}$ and hence

$$\rho_{\mathcal{A}} \geq \bar{\rho} := \rho_{\bar{\mathcal{A}}}.$$

From now on, we assume that

$$\bar{\rho} > -\infty, \quad (64)$$

which implies our assumption (55) for ρ^S . Note also that we have $\bar{\mathcal{A}} = \mathcal{A}^S$ if $\rho_{\mathcal{A}}$ is the worst-case measure.

Proposition 2.10. *The minimal penalty function $\bar{\alpha}_{\min}$ for $\bar{\rho}$ is given by*

$$\bar{\alpha}_{\min}(Q) = \alpha_{\min}^S(Q) + \alpha_{\min}(Q),$$

where α_{\min}^S is the minimal penalty function for ρ^S and α_{\min} is the minimal penalty function for $\rho_{\mathcal{A}}$.

Proof. We claim that

$$\{X \mid \bar{\rho}(X) < 0\} \subseteq \{X^S + A \mid X^S \in \mathcal{A}^S, A \in \mathcal{A}\} \subseteq \bar{\mathcal{A}}. \quad (65)$$

If $\bar{\rho}(X) < 0$, then there exists $A \in \mathcal{A}$ and $\xi \in \mathcal{S}$ such that $X + G_T(\xi) \geq A$. Therefore $X^S := X - A \in \mathcal{A}^S$. Conversely, if $X^S \in \mathcal{A}^S$ then $X^S + G_T(\xi) \geq 0$ for some $\xi \in \mathcal{S}$. Hence, for any $A \in \mathcal{A}$, we get $X^S + A + G_T(\xi) \geq A \in \mathcal{A}$, i.e., $X := X^S + A \in \bar{\mathcal{A}}$.

In view of (65), we have

$$\begin{aligned} \bar{\alpha}_{\min}(Q) &= \sup_{X \in \bar{\mathcal{A}}} E_Q[-X] = \sup_{X: \bar{\rho}(X) < 0} E_Q[-X] \\ &= \sup_{X^S \in \mathcal{A}^S} \sup_{A \in \mathcal{A}} E_Q[-X^S - A] \\ &= \alpha_{\min}^S(Q) + \alpha_{\min}(Q). \quad \square \end{aligned}$$

Barrieu and El Karoui [7] (see also the references therein) study the risk measure $\bar{\rho}$ in more general situation and relate it to the inf-convolution of the risk measures $\rho_{\mathcal{A}}$ and ρ^S . By this operation they can characterize optimal risk transfers in financial markets. The assumption (64) is not as harmless as it might seem. Consider, for instance, the situation of Example 2.8, suppose that the price process S is

bounded, and take $AV@R_\lambda$ as $\rho_{\mathcal{A}}$. The preceding proposition implies that the minimal penalty function of $\bar{\rho}$ is given by

$$\bar{\alpha}_{\min}(Q) = \begin{cases} 0 & \text{if } Q \in \mathcal{P}_S \text{ and } dQ/dP \leq 1/\lambda, \\ +\infty & \text{otherwise.} \end{cases}$$

There may, however, be no measures in \mathcal{P}_S whose density is bounded by $1/\lambda$, in which case (64) is not satisfied. An example in continuous time would be the standard Black & Scholes model. For this reason, the alternative approach to combining hedging with subjective risk measurement as presented in the next section may be more appropriate. It will also have the additional advantage that it yields optimal hedging strategies.

It should also be mentioned that there are some recent approaches to *dynamic* risk measures in financial markets; see Artzner et al. [4, 5, 6], Delbaen [22], Föllmer and Schied [34, 35], Frittelli and Scandolo [39], Riedel [57], Weber [67], to mention only a few.

2.2 Efficient hedging with $AV@R_\lambda$ and other convex risk measures

Let us consider the discrete-time market model of the preceding section. We assume that the model is arbitrage-free, which is equivalent to the assumption that the set \mathcal{P} of equivalent martingale measure for S is nonempty. Let $H \geq 0$ be the discounted payoff of a European claim, and consider an investor who is short in H , i.e., at time T , the investor must deliver the random amount $H(\omega)$. In the situation of Theorem 2.1, the risk of the short position $-H$ is given by

$$\pi_{\text{sup}}(H) := \rho_0(-H) := \sup_{P^* \in \mathcal{P}} E^*[H],$$

where we assume that the right-hand side is finite. In fact, the theory of superhedging tells us $\pi_{\text{sup}}(H)$ is equal to the cost of superreplicating H , i.e., there exists a trading strategy ξ such that

$$V_T = \pi_{\text{sup}}(H) + G_T(\xi) \geq H \quad P\text{-a.s.},$$

see, e.g., [35, Corollary 7.9]. By using such a superhedging strategy, the seller of H can cover almost any possible obligation which may arise from the sale of H and thus eliminate completely the corresponding risk. In many cases, however, the cost $\pi_{\text{sup}}(H)$ will be much too expensive from a practical point of view. And even if H is attainable, a complete elimination of risk by using a replicating strategy for H would consume the entire proceeds from the sale of H , and any opportunity of making a profit would be lost along with the risk.

Let us therefore suppose that the investor is unwilling to put up the initial amount of capital required by a superhedge and is ready to accept some risk. What is the optimal partial hedge which can be achieved with a given smaller amount of capital? In order to make this question precise, we need a risk measure ρ expressing the seller's attitude towards risk, and suppose the investor is only willing to put up a smaller amount

$$v \in (0, \pi_{\text{sup}}(H)).$$

This means that the investor is ready to take some risk: Any "partial" hedging strategy whose value process V satisfies the capital constraint $V_0 \leq v$ will generate a non-trivial *shortfall*

$$(H - V_T)^+.$$

Our aim is to minimize the *shortfall risk*

$$\rho(-(H - V_T)^+)$$

among all admissible strategies satisfying the capital constraint $V_0 \leq v$ (here we assume that ρ is defined on a suitable function space, so that the shortfall risk is well defined). Alternatively, we could minimize the cost under a given bound on the shortfall risk. In other words, the problem consists in constructing strategies which are *efficient* with respect to the trade-off between cost and shortfall risk.

It turns out that the construction of the optimal hedging strategy is carried out in two steps. The first one is to solve the “static” problem of minimizing

$$\rho(-(H - X)^+)$$

among all \mathcal{F}_T -measurable random variables $X \geq 0$ which satisfy the constraints

$$\sup_{P^* \in \mathcal{P}} E^*[X] \leq v.$$

If X^* solves this problem, then so does $H \wedge X^*$. Hence, we may assume that $0 \leq X^* \leq H$. Thus, we can reformulate the problem as

$$\text{minimize } \rho(X - H) \text{ subject to } 0 \leq X \leq H \text{ and } \sup_{P^* \in \mathcal{P}} E^*[X] \leq v. \quad (66)$$

The next step is to fit the terminal value V_T of an admissible strategy to the optimal profile X^* . It turns out that this step can be carried out without any further assumptions on our risk measure ρ . Thus, we assume at this point that the optimal X^* of step one is granted, and we construct the corresponding optimal strategy.

Proposition 2.11. *A superhedging strategy for a solution X^* of (66) with initial investment $\pi_{\text{sup}}(X^*)$ has minimal shortfall risk among all admissible strategies whose value process satisfies the capital constraint $V_0 \leq v$.*

Proof. Let V be the value process of any admissible strategy such that $V_0 \leq v$. Due to Doob’s systems theorem (e.g., [35, Theorem 5.15]), V is a martingale under any $P^* \in \mathcal{P}$, and so $\sup_{P^* \in \mathcal{P}} E^*[V_T] = V_0 \leq v$. Thus, $X := H \wedge V_T$ satisfies the constraints in (66), and we get

$$\rho(-(H - V_T)^+) = \rho(X - H) \geq \rho(X^* - H).$$

Next let V^* be a superhedging strategy for X^* with initial investment $\pi_{\text{sup}}(X^*) = \sup_{P^* \in \mathcal{P}} E^*[X^*]$. Then we have $V_0^* = \sup_{P^* \in \mathcal{P}} E^*[X^*] \leq v$ and $V_T^* \geq 0$. Moreover, $V_T^* \geq X^*$ P -a.s., and thus

$$\rho(X^* - H) = \rho(-(H - X^*)^+) \geq \rho(-(H - V_T^*)^+).$$

This concludes the proof. \square

Let us now return to the static problem defined by (66).

Proposition 2.12. *If ρ is lower semicontinuous with respect to a.s. convergence of random variables in the class $\{-X \mid 0 \leq X \leq H\}$, then there exists a solution of the static optimization problem (66). In particular, there exists a solution if H is bounded and ρ is continuous from above.*

The proof needs the the following variant of Komlos’ principle of subsequences [49].

Lemma 2.13. *Let X_n be a sequence in L^0 such that $\sup_n |X_n| < \infty$ P -almost surely. Then there exists a sequence of convex combinations*

$$Y_n \in \text{conv}\{X_n, X_{n+1}, \dots\}$$

which converges P -almost surely to some $Y \in L^0$.

Proof. We can assume without loss of generality that $\sup_n |X_n| \leq 1$ P -a.s.; otherwise we consider the sequence $\tilde{X}_n := X_n / \sup_n |X_n|$. Then (X_n) is a bounded sequence in the Hilbert space L^2 . Since the closed unit ball in L^2 is weakly compact, the sequence (X_n) has an accumulation point $Y \in L^2$. For each n , the accumulation point Y belongs to the L^2 -closure \mathcal{C}_n of $\text{conv}\{X_n, X_{n+1}, \dots\}$, due to the fact that a closed convex set in L^2 is also weakly closed. Thus, we can find $Y_n \in \text{conv}\{X_n, X_{n+1}, \dots\}$ such that $E[|Y_n - Y|^2] \leq \frac{1}{n^2}$. This sequence (Y_n) converges P -a.s. to Y . \square

Proof Proposition 2.12. Take X_n with $0 \leq X_n \leq H$ and $\sup_{P^* \in \mathcal{P}} E^*[X_n] \leq v$ such that $\rho(X_n - H)$ converges to the infimum A of the shortfall risk. We can use Lemma 2.13 to select convex combinations $Y_n \in \text{conv}\{X_n, X_{n+1}, \dots\}$ which converge P -a.s. to some Y . Then $0 \leq Y \leq H$ and Fatou's lemma yields that

$$E^*[Y] \leq \liminf_{n \uparrow \infty} E^*[Y_n] \leq v$$

for all $P^* \in \mathcal{P}$. The lower semicontinuity of ρ implies that

$$\rho(Y - H) \leq \liminf_{n \uparrow \infty} \rho(Y_n - H).$$

Moreover, the right-hand side is equal to A , due to the convexity of ρ . Hence, Y is the desired minimizer. \square

Combining Proposition 2.12 and Proposition 2.11 yields the existence of an optimal hedging strategy under risk aversion in a general arbitrage-free market model. So far, all arguments were practically the same as in the paper [32] by Föllmer and Leukert.

Beyond the general existence statement of Proposition 2.12, it is sometimes possible to obtain an explicit formula for the optimal solution of the static problem if the market model is complete. Recall that model completeness is equivalent to the uniqueness of the equivalent martingale measure, i.e., to the condition $\mathcal{P} = \{P^*\}$. Thus, the static optimization problem simplifies to

$$\text{minimize } \rho(X - H) \text{ subject to } 0 \leq X \leq H \text{ and } E^*[X] \leq v.$$

By substituting Y for $H - X$, this is equivalent to the problem

$$\text{minimize } \rho(-Y) \text{ subject to } 0 \leq Y \leq H \text{ and } E^*[Y] \geq \tilde{v}, \quad (67)$$

where $\tilde{v} := E^*[H] - v$. We will now solve this problem in the case $\rho = AV@R_\lambda$ and thus recover a recent result by Sekine [66]. Note, however, that our proof is significantly shorter. It relies on the general idea that a minimax problem can be transformed into a standard minimization problem by using a duality result for the expression involving the maximum. In the case of $AV@R_\lambda$, we can use Lemma 1.30, and this idea was first applied by Eichhorn and Römisch [25]. See also Quenez [56] and Hernández-Hernández and Schied [42] for applications of this general idea to optimal investment problems as discussed in Sections 2.3 and 2.4 below.

Theorem 2.14. *Suppose that the price density $\varphi := dP^*/dP$ has a continuous distribution under P . Then the problem (67) has a unique solution Y^* for $\rho = AV@R_\lambda$. Moreover, there exists a critical capital v^* such that*

$$Y^* = H \mathbb{1}_{\{\varphi \geq c\}} \quad \text{if } \tilde{v} \leq v^*,$$

and

$$Y^* = H \mathbb{1}_{\{\varphi \geq c\}} + H \wedge r^* \mathbb{1}_{\{\varphi < c\}} \quad \text{if } \tilde{v} > v^*, \quad (68)$$

for certain constants $c, r^* > 0$.

Proof. Lemma 1.30 gives

$$AV@R_\lambda(-Y) = \frac{1}{\lambda} \min_{r \in \mathbb{R}} (E[(Y - r)^+] + \lambda r) = \frac{1}{\lambda} \min_{r \geq 0} (E[(Y - r)^+] + \lambda r)$$

for $Y \geq 0$. Hence, Y^* must solve

$$\text{minimize } E[(Y - r^*)^+] \text{ subject to } 0 \leq Y \leq H \text{ and } E^*[Y] \geq \tilde{v}, \quad (69)$$

if $r^* \geq 0$ is such that

$$AV@R_\lambda(-Y^*) = \frac{1}{\lambda} E[(Y^* - r^*)^+] + r^* \quad (70)$$

By Lemma 1.30, r^* is a λ -quantile for Y^* .

Let us now solve (69). To this end, we consider first the case in which $r^* = 0$. By writing $Y = H(1-\psi)$, we see that the solution is provided by the Neyman-Pearson lemma in the form of Proposition 1.32. Indeed, ψ^* corresponding to Y^* must be a solution of the problem

$$\text{maximize } E_Q[\psi] \text{ subject to } 0 \leq \psi \leq 1 \text{ and } E_{Q^*}[\psi] \leq 1 - \tilde{v}/E^*[H],$$

where

$$\frac{dQ}{dP} = \frac{H}{E[H]} \quad \text{and} \quad \frac{dQ^*}{dP^*} = \frac{H}{E^*[H]}.$$

Thus, ψ^* is of the form

$$\psi^* = \mathbb{I}_{\{\frac{dQ}{dQ^*} > \tilde{c}\}} = \mathbb{I}_{\{\varphi < c\}} \quad Q\text{-a.s.}$$

for certain constants \tilde{c} and c , and in turn $Y^* = H \mathbb{I}_{\{\varphi \geq c\}}$.

Now we consider the case $r^* > 0$. Note first that we must have $Y^* \geq H \wedge r^*$. Indeed, let us assume $P[Y^* < H \wedge r^*] > 0$. Then we could obtain a strictly lower risk $AV@R_\lambda(-Y^*)$ either by decreasing the level r^* in case $P[Y^* \leq H \wedge r^*] = 1$ or, in case $P[Y^* > H \wedge r^*] > 0$, by shifting mass of Y^* from $\{Y^* > H \wedge r^*\}$ to the set $\{Y^* < H \wedge r^*\}$. Thus, we have to minimize $E[(\hat{Y} + H \wedge r^* - r^*)^+]$ subject to $0 \leq \hat{Y} \leq H - H \wedge r^*$ and $E^*[\hat{Y}] \geq \hat{v} := \tilde{v} - E^*[H \wedge r^*]$. Any \hat{Y} satisfying these constraints must be concentrated on $\{H > r^*\}$, so that the problem is equivalent to

$$\text{minimize } E[\hat{Y}] \text{ subject to } 0 \leq \hat{Y} \leq H - H \wedge r^* \text{ and } E^*[\hat{Y}] \geq \hat{v}. \quad (71)$$

But this problem is equivalent to the one for $r^* = 0$ if we replace H by $H - H \wedge r^*$. Hence, it is solved by $\hat{Y}^* = (H - H \wedge r^*) \mathbb{I}_{\{\varphi \geq c\}}$ for some constant c . It follows that

$$Y^* = \hat{Y}^* + H \wedge r^* = H \mathbb{I}_{\{\varphi \geq c\}} + H \wedge r^* \mathbb{I}_{\{\varphi < c\}}.$$

The preceding arguments also yield that for any $r^* = r \leq \tilde{v}$ there exists a unique solution $Y_{r,\tilde{v}}$ to the minimization problem (69). Clearly, we have $Y_{r,\tilde{w}} \geq Y_{r,\tilde{v}}$ if $\tilde{w} \geq \tilde{v}$. Moreover, we can find some $r(\tilde{v})$ such that $E[(Y_{r,\tilde{v}} - r)^+] + \lambda r$ is minimal. We leave it as an exercise for the reader to prove that $r(\tilde{w}) \geq r(\tilde{v})$ if $\tilde{w} \geq \tilde{v}$. This fact then yields the uniqueness of solutions as well as the existence of the critical value v^* . \square

The following example is taken from [61].

Example 2.15. In case $H \equiv 1$, the solution can be determined explicitly if φ has a continuous and strictly increasing distribution function. In order to compute the solution, let us fix $\tilde{v} \in (0, 1)$ and let

$$Y_r = r + (1-r) \mathbb{I}_{\{\varphi \geq c\}},$$

where $r = r(c)$ is such that $E^*[Y_r] = \tilde{v}$, i.e.,

$$r(c) = \frac{\tilde{v} - E[\varphi; \varphi \geq c]}{E[\varphi; \varphi < c]}.$$

This makes sense as long as $c \geq c_0$, where c_0 is defined via $\tilde{v} = E[\varphi; \varphi \geq c_0]$. Due to the Neyman-Pearson lemma, $Y_{r(c)}$ then is the optimal solution of the problem (69) for $r^* := r(c)$. The preceding theorem states that the solution of our original problem is within the class $\{Y_{r(c)} \mid c \geq c_0\}$. Thus, we have to minimize

$$\lambda AV@R_\lambda(-Y_{r(c)}) = \int_{1-\lambda}^1 q_{Y_{r(c)}}(s) ds = \lambda r + (1-r) \int_{1-\lambda}^1 \mathbb{I}_{\{q_\varphi(s) \geq c\}} ds$$

over $c \geq c_0$. Here we have used Lemma 1.38 in the second identity. This minimization problem can be simplified further by using the reparameterization $c = q_\varphi(t)$, where q_φ is the (unique) quantile function for φ under P . Indeed, by letting

$$\varrho(t) := r(q_\varphi(t)) = 1 - \frac{1 - \tilde{v}}{\Phi(t)}$$

for $\Phi(t) := \int_0^t q_\varphi(s) ds$, we simply have to minimize the function

$$\begin{aligned} R(t) &:= \lambda AV @ R_\lambda(-Y_{\varrho(t)}) = \lambda \varrho(t) + (1 - \varrho(t)) \int_{1-\lambda}^1 \mathbb{I}_{[t,1]}(s) ds \\ &= \lambda \varrho(t) + (1 - \varrho(t))(\lambda - (t - 1 + \lambda)^+) \\ &= \lambda - (1 - v) \frac{(t - 1 + \lambda)^+}{\Phi(t)} \end{aligned}$$

over $t \geq t_0 := F_\varphi(c_0)$. For $t \leq 1 - \lambda$, we get $R(t) = 1$, which cannot be optimal. Moreover, it is easy to see that the function

$$t \mapsto \frac{t - 1 + \lambda}{\Phi(t)}$$

has a unique maximizer $t_\lambda \in (1 - \lambda, 1]$, which will define the solution as soon as $t_\lambda \geq t_0$ and as long as $t = t_0$ does not give a better result. If $t_\lambda \leq t_0$, then R has no minimizer on $(t_0, 1]$, and it follows that $t^* = t_0$. Thus, let us finally compare $R(t_\lambda)$ against $R(t_0)$ in case $t_\lambda > t_0$. Since $t_\lambda > 1 - \lambda$, we have

$$R(t_\lambda) = \lambda - (1 - v) \frac{(t_\lambda - 1 + \lambda)^+}{\Phi(t_\lambda)} = \lambda - \Phi(t_0) \frac{(t_\lambda - 1 + \lambda)^+}{\Phi(t_\lambda)}$$

and

$$R(t_0) = \lambda - (t_0 + \lambda - 1)^+ = \lambda - \Phi(t_0) \frac{(t_0 - 1 + \lambda)^+}{\Phi(t_0)}.$$

Since t_λ is the unique maximizer of the function $t \mapsto (t - 1 + \lambda)^+ / \Phi(t)$, we thus see that $R(t_\lambda)$ is strictly better than $R(t_0)$. Hence the solution is defined by

$$t^* := t_0 \vee t_\lambda.$$

Note also that, for $\|\varphi\|_\infty > \lambda^{-1}$, t_λ is the unique solution of the equation

$$q_\varphi(t_\lambda)(t_\lambda - 1 + \lambda) = \Phi(t_\lambda).$$

Clearly, t_λ is independent of \tilde{v} , while t_0 decreases from 1 to 0 as \tilde{v} increases from 0 to 1. Thus, by taking \tilde{v}^* as the capital level for which $t_\lambda = t_0$, we see that the optimal solution has the form $Y^* = \mathbb{I}_{\{\varphi \geq q_\varphi(t_0)\}}$ for $\tilde{v} \leq \tilde{v}^*$ and $Y^* = r^* + (1 - r^*) \mathbb{I}_{\{\varphi \geq q_\varphi(t_\lambda)\}}$ for $\tilde{v} > \tilde{v}^*$, where $r^* = 1 - \frac{1 - \tilde{v}^*}{\Phi(t_\lambda)} > 0$. \diamond

2.3 Optimal investment under Knightian uncertainty

In this section, we return to the setting of Section 1.6, where an investor assesses payoffs at some time T by a robust utility functional of the form

$$X \mapsto \inf_{Q \in \mathcal{Q}} E_Q[U(X)].$$

Here, \mathcal{Q} is a set of probability measures and U is a strictly concave utility function. The problem we are considering is to find trading strategies that are optimal with respect to the investor's preferences. This problem was first discussed systematically by Quenez [56] and Schied [62]. Our presentation draws from this second paper.

In this section, we consider a standard complete market model in continuous time. That is, we consider a market model consisting of one bond and d risky assets, whose price processes are denoted by $S = (S_t^i)_{0 \leq t \leq T, i=1, \dots, d}$. We may assume without loss of generality that the price of the bond is constant. The process S is assumed to be a semimartingale on the filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{0 \leq t \leq T}, P)$, and we emphasize that this includes the case of a discrete-time market model, in which prices are adjusted only at times $t = 0, 1, \dots, T$: just set $S_t := S_{[t]}$ and $\mathcal{F}_t := \mathcal{F}_{[t]}$ for arbitrary $t \in [0, T]$. We assume that \mathcal{F}_0 is P -trivial and that the market is complete in the sense that there exists a unique probability measure P^* that is equivalent to P and under which S is a d -dimensional local martingale. In a discrete-time

setting, market completeness implies that Ω can be chosen as a finite set, and this will simplify certain assumptions on our set \mathcal{Q} .

A self-financing trading strategy can be regarded as a pair (x, ξ) , where $x \in \mathbb{R}$ is the initial investment and $\xi = (\xi_t^i)_{0 \leq t \leq T, i=1, \dots, d}$ is a predictable and S -integrable process. The value process X associated with (x, ξ) is given by $X_0 = x$ and

$$X_t = X_0 + \int_0^t \xi_r dS_r, \quad 0 \leq t \leq T.$$

For $x \in \mathbb{R}$ given, we denote by $\mathcal{X}(x)$ the set of all such processes X with $X_0 \leq x$ which are admissible in the sense that $X_t \geq 0$ for $0 \leq t \leq T$ and whose terminal wealth X_T has a well-defined robust utility

$$\inf_{Q \in \mathcal{Q}} E_Q[U(X_T)].$$

Now we can state our main problem:

$$\text{maximize } \inf_{Q \in \mathcal{Q}} E_Q[U(X_T)] \text{ among all } X \in \mathcal{X}(x). \quad (72)$$

Definition 2.16. Let \mathcal{Q} be a set of probability measures absolutely continuous with respect to P^* . $Q_0 \in \mathcal{Q}$ is called a *least favorable measure* with respect to P^* if the density $\pi = dP^*/dQ_0$ (taken in the sense of the Lebesgue decomposition) satisfies

$$Q_0[\pi \leq t] = \inf_{Q \in \mathcal{Q}} Q[\pi \leq t] \quad \text{for all } t > 0.$$

In the sequel, we will assume that \mathcal{Q} is a convex set. Moreover, we will assume throughout this note that \mathcal{Q} is equivalent to P^* in the following sense:

$$P^*[A] = 0 \quad \iff \quad Q[A] = 0 \text{ for all } Q \in \mathcal{Q}. \quad (73)$$

Clearly, our problem (72) would not be well-posed without the implication “ \implies ”. The converse implication is economically natural, since a position with a positive price should lead to a non-vanishing utility. Note that (73) is strictly weaker than the condition that *every* measure in \mathcal{Q} is equivalent to P^* , which is often assumed in papers on model uncertainty; for a discussion see Cont [17].

Now we can state our first main result. It reduces the robust utility maximization problem to a standard utility maximization problem plus the computation of a least favorable measure, which is *independent* of the utility function.

Theorem 2.17. *Suppose that \mathcal{Q} admits a least favorable measure $Q_0 \sim P^*$. Then the robust utility maximization problem (72) is equivalent to the standard utility maximization problem with respect to Q_0 , i.e., to (72) with \mathcal{Q} replaced by $\mathcal{Q}_0 := \{Q_0\}$. More precisely, $X_T^* \in \mathcal{X}(x)$ solves the robust problem (72) if and only if it solves the standard problem for Q_0 , and the corresponding value functions are equal, whether there exists a solution or not:*

$$\sup_{X \in \mathcal{X}(x)} \inf_{Q \in \mathcal{Q}} E_Q[U(X_T)] = \sup_{X \in \mathcal{X}(x)} E_{Q_0}[U(X_T)], \quad \text{for all } x.$$

This result has the following economic consequence. Let \succ denote the preference order induced by our robust utility functional, i.e.,

$$X \succ Y \quad \iff \quad \inf_{Q \in \mathcal{Q}} E_Q[U(X)] > \inf_{Q \in \mathcal{Q}} E_Q[U(Y)].$$

Then, although \succ does not satisfy the axioms of (subjective) expected utility theory, optimal investment decisions with respect to \succ are still made in accordance with the Savage/Anscombe-Aumann version of expected utility, provided that we take Q_0 as the subjective probability measure. The surprising part is that this subjective measure neither depends on the initial investment $x = X_0$ nor on the choice of the utility function U . If \mathcal{Q} does not admit a least favorable measure, then it is still possible that the

robust problem is equivalent to a standard utility maximization problem with a subjective measure Q , which then, however, will depend on x and U . This can be shown by suitably adjusting the arguments in Proposition 2.20 below; see Gundel [41] for details.

Let us now show that the condition $Q_0 \sim P^*$ is always satisfied if \mathcal{Q} is convex and closed in total variation. Recall that \mathcal{Q} is closed in total variation if and only if $\{dQ/dP^* \mid Q \in \mathcal{Q}\}$ is closed in $L^1(P^*)$.

Lemma 2.18. *Suppose that \mathcal{Q} is convex and closed in total variation. Then every least favorable measure Q_0 is equivalent to P^* .*

Proof. Due to our assumptions and the Halmos-Savage theorem, \mathcal{Q} contains a measure $Q_1 \sim P^*$. We get

$$1 = Q_0[\pi < \infty] = \lim_{t \uparrow \infty} Q_0[\pi \leq t] = \lim_{t \uparrow \infty} \inf_{Q \in \mathcal{Q}} Q[\pi \leq t] \leq Q_1[\pi < \infty].$$

Hence, also $P^*[\pi < \infty] = 1$ and in turn $P^* \ll Q_0$. \square

We also have the following converse to Theorem 2.17:

Theorem 2.19. *Suppose $Q_0 \in \mathcal{Q}$ is such that for all utility functions and all $x > 0$ the robust utility maximization problem (72) is equivalent to the standard utility maximization problem with respect to Q_0 . Then Q_0 is a least favorable measure in the sense of Definition 2.16.*

The proof will show that in the preceding Theorem the class of all utility functions can be replaced by the smaller class of all bounded and continuously differentiable utility functions.

Let us next state the following elementary characterization of least favorable measures, which is a variant of Theorem 6.1 in [46].

Proposition 2.20. *For $Q_0 \in \mathcal{Q}$ with $Q_0 \sim P^*$ and $\pi := dP^*/dQ_0$, the following conditions are equivalent.*

(a) Q_0 is a least favorable measure for P^* .

(b) For all decreasing functions $f : (0, \infty) \rightarrow \mathbb{R}$ such that $\inf_{Q \in \mathcal{Q}} E_Q[f(\pi) \wedge 0] > -\infty$,

$$\inf_{Q \in \mathcal{Q}} E_Q[f(\pi)] = E_{Q_0}[f(\pi)].$$

(c) For all increasing functions $g : (0, \infty) \rightarrow \mathbb{R}$ such that $\sup_{Q \in \mathcal{Q}} E_Q[g(\pi) \vee 0] < \infty$,

$$\sup_{Q \in \mathcal{Q}} E_Q[g(\pi)] = E_{Q_0}[g(\pi)].$$

(d) Q_0 minimizes

$$I_\Phi(P^*|Q) := \int \Phi\left(\frac{dQ}{dP^*}\right) dP^*$$

among all $Q \in \mathcal{Q}$, for all continuous convex functions $\Phi : [0, \infty) \rightarrow \mathbb{R}$ such that $I_\Phi(P^*|Q)$ is finite for some $Q \in \mathcal{Q}$.

Proof. (a) \Leftrightarrow (b): According to the definition, Q_0 is a least favorable measure if and only if $Q_0 \circ \pi^{-1}$ stochastically dominates $Q \circ \pi^{-1}$ for all $Q \in \mathcal{Q}$. Hence, if f is bounded, then the equivalence of (a) and (b) is just the standard characterization of stochastic dominance (see, e.g., [35, Theorem 2.71]). If f is unbounded but satisfies $\inf_{Q \in \mathcal{Q}} E_Q[f(\pi) \wedge 0] > -\infty$, then assertion (b) holds for $f_N := (-N) \vee f \wedge 0$. Thus, for all $Q \in \mathcal{Q}$ and $N \in \mathbb{N}$,

$$E_Q[f_N(\pi)] \geq E_{Q_0}[f_N(\pi)] \geq E_{Q_0}[f(\pi) \wedge 0] > -\infty.$$

By sending N to infinity, it follows that $E_Q[f(\pi) \wedge 0] \geq E_{Q_0}[f(\pi) \wedge 0]$ for every $Q \in \mathcal{Q}$. After using a similar argument on $0 \vee f(\pi)$, we get

$$E_Q[f(\pi)] = E_Q[f(\pi) \vee 0] + E_Q[f(\pi) \wedge 0] \geq E_{Q_0}[f(\pi)] \quad \text{for all } Q \in \mathcal{Q}.$$

(b) \Leftrightarrow (c) follows by changing signs.

(b) \Rightarrow (d): Clearly, $I_\Phi(P^*|Q)$ is well-defined and larger than $\Phi(1)$ for each $Q \ll P$. Now take a $Q_1 \in \mathcal{Q}$ with $I_\Phi(P^*|Q_1) < \infty$, and denote by $\Phi'_+(x)$ the right-hand derivative of Φ at $x \geq 0$. Suppose first that Φ'_+ is bounded. Since $\Phi(y) - \Phi(x) \geq \Phi'_+(x)(y - x)$, we have

$$I_\Phi(P^*|Q_1) - I_\Phi(P^*|Q_0) \geq \int \Phi'_+(\pi^{-1}) \left(\frac{dQ_1}{dP^*} - \frac{dQ_0}{dP^*} \right) dP^* = \int f(\pi) dQ_1 - \int f(\pi) dQ_0,$$

where $f(x) := \Phi'_+(1/x)$ is a bounded decreasing function. Therefore $\int f(\pi) dQ_1 \geq \int f(\pi) dQ_0$, and Q_0 minimizes $I_\Phi(P^*|\cdot)$ on \mathcal{Q} . If Φ'_+ is unbounded, one can either use a straightforward approximation argument or apply [35, Corollary 2.62].

(d) \Rightarrow (b): It is enough to prove (b) for continuous bounded decreasing functions f . For such a function f let $\Phi(x) := \int_1^x f(1/t) dt$. Then Φ is convex. For $Q_1 \in \mathcal{Q}$ we let $Q_t := tQ_1 + (1-t)Q_0$ and $h(t) := I_\Phi(P^*|Q_t)$. The right-hand derivative of h satisfies $0 \leq h'_+(0) = \int f(\pi) dQ_1 - \int f(\pi) dQ_0$, and the proof is complete. \square

An important method for solving optimal investment problems is the duality approach via convex analysis, which is sometimes also called the ‘martingale method’. Below we will state such a result which is valid in our present setting. It follows immediately by combining Theorem 2.17 with Proposition 2.20 and Theorem 2.0 of Kramkov and Schachermayer [50], which is the corresponding result for standard utility maximization problems. We have to assume that U is continuously differentiable and satisfies the Inada conditions

$$U'(0) := \lim_{x \downarrow 0} U'(x) = +\infty \quad \text{and} \quad U'(\infty) := \lim_{x \uparrow \infty} U'(x) = 0.$$

We denote by

$$u(x) = \sup_{X \in \mathcal{X}(x)} \inf_{Q \in \mathcal{Q}} E_Q[U(X_T)], \quad x > 0,$$

the value function of the problem (72). Let

$$V(y) = \sup_{x > 0} [U(x) - xy], \quad y > 0,$$

denote the convex conjugate of U and define the function

$$I := -V' = (U')^{-1}.$$

We also define the convex function

$$v(y) = \inf_{Q \in \mathcal{Q}} E_Q[V(y \cdot \pi)], \quad y > 0.$$

Theorem 2.21. *Suppose that \mathcal{Q} admits a least favorable measure $Q_0 \sim P^*$ and that $u(x)$ is finite for some $x > 0$. Then:*

(a) $u(x)$ is finite for all $x > 0$, and $v(y) < \infty$ for $y > 0$ sufficiently large. The function v is continuously differentiable in the interior (y_0, ∞) of its effective domain. The function u is continuously differentiable on $(0, \infty)$ and strictly concave on $(0, x_0)$, where $x_0 := -\lim_{y \downarrow y_0} v'(y)$. For $x, y > 0$,

$$v(y) = \sup_{x > 0} [u(x) - xy] \quad \text{and} \quad u(x) = \inf_{y > 0} [v(y) + xy].$$

Moreover, $u'(0) := \lim_{x \downarrow 0} u'(x) = \infty$ and $v'(\infty) = \lim_{y \uparrow \infty} v'(y) = 0$.

(b) For $x < x_0$ there exists a unique solution $X^*(x) \in \mathcal{X}(x)$ of (72), and its terminal wealth is of the form

$$X_T^*(x) = I(y \cdot \pi), \quad \text{for } y = u'(x).$$

(c) For $0 < x < x_0$ and $y < y_0$,

$$u' = x^{-1} \sup_{Q \in \mathcal{Q}} E_Q[X_T^*(x) U'(X_T^*(x))] \quad \text{and} \quad v'(y) = E^*[V'(y \cdot \pi)].$$

A duality theorem for robust utility maximization in *incomplete* markets was first obtained by Quenez [56] under the additional assumption that *every* measure $Q \in \mathcal{Q}$ is equivalent to a given reference measure P . It was extended to the general case by Schied and Wu [64]; see also Föllmer and Gundel [30] for a different technique and Schied [63] for a further extension. One of the important features of these kind of duality results is that they reduce the original maximin problem to a dual problem, which then is of reduced complexity as it just consists in minimizing over a certain set of controls. We have seen this principle already in the proof of Theorem 2.14. In robust utility maximization, it has been employed by Quenez [56] together stochastic control techniques involving backward stochastic differential equations. Hernández-Hernández and Schied [42] use duality combined with partial differential equations to characterize the value function of a robust utility problem as the unique bounded classical solution of a nonlinear Hamilton-Jacobi-Bellman equation.

Let us now turn to the proofs of Theorems 2.17 and 2.19. Let X^* be a solution of the standard utility maximization problem for the least favorable measure Q_0 . Then it is well known that $X_T^* = I(y\pi)$ for some constant $y > 0$. Thus, one easily checks via Proposition 2.20 that X^* is also a solution of the robust utility maximization problem. However, in order to show the full equivalence of the two problems, we must also take care of the situation in which the standard problem has no solution. Our key result is the following proposition.

Proposition 2.22. *Let $Q_0 \sim P^*$ be a least favorable measure and $\pi = dP^*/dQ_0$.*

(a) *For any $X \in \mathcal{X}(x)$ there exists $\tilde{X} \in \mathcal{X}(x)$ such that*

$$\inf_{Q \in \mathcal{Q}} E_Q[U(\tilde{X}_T)] \geq \inf_{Q \in \mathcal{Q}} E_Q[U(X_T)]$$

and such that $\tilde{X}_T = f(\pi)$ for some deterministic decreasing function $f : (0, \infty) \rightarrow [0, \infty)$.

(b) *The terminal wealth of any solution X^* of (72) is of the form $X_T^* = f^*(\pi)$ for a deterministic decreasing function $f^* : (0, \infty) \rightarrow [0, \infty)$.*

Proof of Proposition 2.22:. (a) By market completeness, it suffices to construct a decreasing function $f \geq 0$ such that $E^*[f(\pi)] \leq x$ and

$$\inf_{Q \in \mathcal{Q}} E_Q[U(f(\pi))] \geq \inf_{Q \in \mathcal{Q}} E_Q[U(X_T)]. \quad (74)$$

To this end, we denote by $F_Y(x) := Q_0[Y \leq x]$ the distribution function and by $q_Y(t)$ a quantile function of a random variable Y with respect to the probability measure Q_0 .

Let us define a function f by

$$f(t) := \begin{cases} q_{X_T}(1 - F_\pi(t)) & \text{if } F_\pi \text{ is continuous at } t, \\ \frac{1}{F_\pi(t) - F_\pi(t-)} \int_{F_\pi(t-)}^{F_\pi(t)} q_{X_T}(1 - s) ds & \text{otherwise.} \end{cases} \quad (75)$$

Then f is decreasing and satisfies $f(q_\pi) = E_\lambda[h | q_\pi]$, where λ is the Lebesgue measure and $h(t) := q_{X_T}(1 - t)$. Hence, Jensen's inequality for conditional expectations and Lemma 1.38 show that

$$\begin{aligned} \inf_{Q \in \mathcal{Q}} E_Q[U(X_T)] &\leq E_{Q_0}[U(X_T)] = \int_0^1 U(h(t)) dt \\ &\leq \int_0^1 U(E_\lambda[h | q_\pi](t)) dt = \int_0^1 U(q_{f(\pi)}(1 - t)) dt \\ &= E_{Q_0}[U(f(\pi))] = \inf_{Q \in \mathcal{Q}} E_Q[U(f(\pi))], \end{aligned} \quad (76)$$

where we have used Proposition 2.20 in the last step. Thus, f satisfies (74).

It remains to show that $f(\pi)$ satisfies the capital constraint. To this end, we first use the lower Hardy-Littlewood inequality:

$$x \geq E^*[X_T] = E_{Q_0}[\pi X_T] \geq \int_0^1 q_\pi(t) q_{X_T}(1-t) dt. \quad (77)$$

Here we may replace $q_{X_T}(1-t) = h(t)$ by $E_\lambda[h|q_\pi](t) = f(q_\pi(t))$. We then get

$$\int_0^1 q_\pi(t) q_{X_T}(1-t) dt = \int_0^1 q_\pi(t) f(q_\pi(t)) dt = E_{Q_0}[\pi f(\pi)] = E^*[f(\pi)]. \quad (78)$$

Thus, f is as desired.

(b) Now suppose X^* solves (72). If X_T^* is not Q_0 -a.s. $\sigma(\pi)$ -measurable, then $Y := E_{Q_0}[X_T^* | \pi]$ must satisfy

$$E_{Q_0}[U(Y)] > E_{Q_0}[U(X_T^*)], \quad (79)$$

due to the strict concavity of U . If we define \tilde{f} as in (75) with Y replacing X_T , then the proof of part (a) yields that

$$E^*[\tilde{f}(\pi)] = E_{Q_0}[\pi \tilde{f}(\pi)] \leq E_{Q_0}[\pi Y] = E_{Q_0}[\pi X_T^*] \leq x,$$

and by (76) and (79),

$$\inf_{Q \in \mathcal{Q}} E_Q[U(\tilde{f}(\pi))] \geq E_{Q_0}[U(Y)] > E_{Q_0}[U(X_T^*)] \geq \inf_{Q \in \mathcal{Q}} E_Q[U(X_T^*)],$$

in contradiction to the optimality of X^* . Thus, X_T^* is necessarily $\sigma(\pi)$ -measurable and can hence be written as a (not yet necessarily decreasing) function of π .

If we define f^* as in (75) with X_T^* replacing X_T , then $f^*(\pi)$ is the terminal wealth of yet another solution in $\mathcal{X}(x)$. Clearly, we must have $E^*[X_T^*] = x = E^*[f^*(\pi)]$. Thus, (77) and (78) yield that $E_{Q_0}[\pi X_T^*] = \int_0^1 q_\pi(t) q_{X_T^*}(1-t) dt$. But then the ‘‘only if’’ part of the lower Hardy-Littlewood inequality together with the $\sigma(\pi)$ -measurability of X_T^* imply that X_T^* is a decreasing function of π . \square

Proof of Theorem 2.17. Proposition 2.22 implies that in solving the robust utility maximization problem (72) we may restrict ourselves to strategies whose terminal wealth is a decreasing function of π . By Propositions 2.20, the robust utility of a such a terminal wealth is the same as the expected utility with respect to Q_0 . On the other hand, taking $\mathcal{Q}_0 := \{Q_0\}$ in Proposition 2.22 implies that the standard utility maximization problem for Q_0 also requires only strategies whose terminal wealth is a decreasing function of π . Therefore, the two problems are equivalent, and Theorem 2.17 is proved. \square

Proof of Theorem 2.19. Let (U_n) be a sequence of nonnegative and continuously differentiable utility functions that increase uniformly to the concave increasing function $U(x) := x \wedge 1$. Uniform convergence of U_n implies convergence of the corresponding value functions:

$$u_0^n(x) := \sup_{X \in \mathcal{X}(x)} E_{Q_0}[U_n(X_T)] \nearrow \sup_{X \in \mathcal{X}(x)} E_{Q_0}[U(X_T)] =: u_0(x). \quad (80)$$

If we assume that $U_1'(x)$ decreases fast enough to 0 as $x \uparrow \infty$, then $E^*[I_1^+(c\pi)] < \infty$ for all $c > 0$, where $\pi := dP^*/dQ_0$ and I_1^+ is the inverse of U_1' on $(0, U_1'(0))$ and $I_1^+(x) = 0$ for $x \geq U_1'(0)$. Market completeness and [35, Theorem 3.39] guarantee that, for every $0 < x \leq 1$ and each $n \in \mathbb{N}$, there exists a solution $X^n \in \mathcal{X}(x)$ for the standard utility maximization problem with utility function U_n under Q_0 . Note that the preceding two statements also remain true for $P^* \ll Q_0$ and that $X_T^n = 0$ on $\{\pi = \infty\}$.

By a Komlos-type argument (see Lemma 3.3 in [50]), there exists a sequence $Y_n \in \text{conv}\{X_T^n, X_T^{n+1}, \dots\}$ which converges P^* -a.s. to some random variable $X_T^* \geq 0$, which satisfies $E^*[X_T^*] \leq x$, due to Fatou's

lemma. Hence, X_T^* corresponds to a value process $X^* \in \mathcal{X}(x)$. Let us write Y_n as the convex combination $Y_n = \sum_{k \geq n} \alpha_{k,n} X_T^k$, where only finitely many $\alpha_{k,n}$ are nonzero. Then,

$$\begin{aligned} u_0(x) &\geq E_{Q_0}[U(X_T^*)] = \lim_{n \uparrow \infty} E_{Q_0}[U(Y_n)] \geq \limsup_{n \uparrow \infty} \sum_{k \geq n} \alpha_{k,n} E_{Q_0}[U_k(X_T^k)] \\ &= \limsup_{n \uparrow \infty} \sum_{k \geq n} \alpha_{k,n} u_0^k(x) = u_0(x), \end{aligned}$$

due to (80). Hence, X^* is optimal for the utility maximization problem with U and Q_0 . Since U is constant on $[1, \infty)$, we must have $0 \leq X_T^* \leq 1$ P^* -almost surely. Thus, X_T^* is a solution to the problem of maximizing $E_{Q_0}[U(X)] = E_{Q_0}[X]$ under the constraints $0 \leq X \leq 1$ and $E^*[X] \leq x$. Hence, the generalized Neyman-Pearson lemma in the form of Proposition 1.32 implies that $X_T^* = \mathbb{I}_{\{\pi < q\}} + \kappa \mathbb{I}_{\{\pi = q\}}$, where q can be any x -quantile for the law of π under P^* , and κ is a $[0, 1]$ -valued random variable. In particular,

$$X_T^* = \mathbb{I}_{\{\pi \leq q\}} \quad P^*\text{-a.s. for } x \text{ with } P^*[\pi = q] = 0. \quad (81)$$

Note also that the x -quantile q is unique if $P^*[\pi = q] > 0$.

Next, if $Q \in \mathcal{Q}$ is given, then

$$\begin{aligned} E_Q[U(X_T^*)] &= \lim_{n \uparrow \infty} E_Q[U(Y_n)] \geq \limsup_{n \uparrow \infty} \sum_{k \geq n} \alpha_{k,n} E_Q[U_k(X_T^k)] \\ &\geq \limsup_{n \uparrow \infty} \sum_{k \geq n} \alpha_{k,n} E_{Q_0}[U_k(X_T^k)] = \limsup_{n \uparrow \infty} \sum_{k \geq n} \alpha_{k,n} u_0^k(x) \\ &= u_0(x) = E_{Q_0}[U(X_T^*)], \end{aligned} \quad (82)$$

where we have used the fact that $E_Q[U_k(X_T^k)] \geq E_{Q_0}[U_k(X_T^k)]$ for all k . This inequality follows from the hypothesis of the theorem: X_T^k solves both the standard and the robust utility maximization problems, and the corresponding value functions are equal, i.e.,

$$\inf_{Q \in \mathcal{Q}} E_Q[U_k(X_T^k)] = \sup_{X \in \mathcal{X}(x)} \inf_{Q \in \mathcal{Q}} E_Q[U_k(X_T)] = \sup_{X \in \mathcal{X}(x)} E_{Q_0}[U_k(X_T)] = E_{Q_0}[U_k(X_T^k)].$$

Finally, combining (82) with (81) yields $Q[\pi \leq q] = E_Q[U(X_T^*)] \geq E_{Q_0}[U(X_T^*)] = Q_0[\pi \leq q]$ for all but countably many and, in turn, all $q \in [0, 1]$. \square

2.4 Examples of least favorable measures

In this section, we will discuss two classes of examples in which least favorable measures can be determined. The first is a Black-Scholes market with uncertain drift. The second is provided by the classical Huber-Strassen theory, where \mathcal{Q} is the σ -core of a submodular capacity.

Utility maximization with uncertain drift

Consider a Black-Scholes market model with a riskless bond, B_t , of which we assume $B_t \equiv 1$ and with d risky assets $S_t = (S_t^1, \dots, S_t^d)$ that satisfy an SDE of the form

$$dS_t^i = S_t^i \sum_{j=1}^d \sigma_t^{ij} dW_t^j + \alpha_t^i S_t^i dt \quad (83)$$

with a d -dimensional Brownian motion W and a volatility matrix σ_t that has full rank. Now suppose the investor is uncertain about the “true” future drift $\alpha_t = (\alpha_t^1, \dots, \alpha_t^d)$ in the market: any drift α is possible that is adapted to the filtration generated by W and satisfies $\alpha_t \in C_t$, where C_t is a nonrandom bounded closed convex subset of \mathbb{R}^d . Let us denote by \mathcal{A} the set of all such processes α . This uncertainty in the choice of the drift can be expressed by the set

$$\mathcal{Q} := \left\{ Q \mid S \text{ has drift } \alpha^Q \in \mathcal{A} \text{ under } Q \right\}.$$

Under P^* the drift α in (83) vanishes. It turns out that the optimal investment problem with *uncertain drift* can be solved by transforming it into a problem for *uncertain volatility* as studied by El Karoui et al. [27]. To this end, we denote by α_t^0 the element in C_t that minimizes the norm $|\sigma_t^{-1}x|$ among all $x \in C_t$

Proposition 2.23. *Suppose that σ_t is deterministic and that both α_t^0 and σ_t are continuous in t . Then \mathcal{Q} admits a least favorable measure Q_0 with respect to P^* , which is characterized by having the drift α^0 .*

Proof. We will use arguments from [27] to check condition (d) of Proposition 2.20. The density process of $Q \in \mathcal{Q}$ with respect to P^* has the form

$$Z_t^Q := \frac{dQ}{dP^*} \Big|_{\mathcal{F}_t} = \exp \left(\int_0^t \lambda_s dW_s^* - \frac{1}{2} \int_0^t |\lambda_s|^2 ds \right),$$

where $\lambda_s = \sigma_s^{-1} \alpha_s^Q$ and W^* is a d -dimensional P^* -Brownian motion. Similarly, the density process $Z := Z^{Q_0}$ will involve the deterministic integrand $\gamma_s := \sigma_s^{-1} \alpha_s^0$. Let Φ be a convex function on \mathbb{R}_+ . We may assume without loss of generality that Φ has at most polynomial growth. Then $v(t, x) := E^*[\Phi(xZ_t)]$ is a solution of the Black-Scholes equation $v_t = \frac{1}{2} |\gamma_t|^2 x^2 v_{xx}$. This fact and Itô's formula show that

$$dv(T-t, Z_t^Q) = v_x(T-t, Z_t^Q) dZ_t^Q + \frac{1}{2} (Z_t^Q)^2 v_{xx}(T-t, Z_t^Q) (|\lambda_t|^2 - |\gamma_t|^2) dt.$$

One easily checks that the first term on the right is a martingale increment. Moreover, v is convex and $|\lambda_t|^2 \geq |\gamma_t|^2$ by definition of α^0 . Hence, $v(T-t, Z_t^Q)$ is a submartingale and

$$E^*[\Phi(Z_T^Q)] = E^*[v(0, Z_T^Q)] \geq v(T, Z_0^Q) = E^*[\Phi(Z_T)].$$

□

An obvious question is whether the strong condition that the volatility σ_t and the drift α^0 are deterministic can be relaxed. One case of interest is, for instance, a local volatility model in which the equation (83) is replaced by the one-dimensional SDE

$$dS_t = \sigma(t, S_t) S_t dW_t + \alpha_t S_t dt. \quad (84)$$

In this case, however, it will typically no longer be optimal to take the drift that is closest to the riskneutral case $\alpha \equiv 0$. The reason is that the utility of an investment can be reduced both by a small drift and by a large volatility, and these two requirements may be competing with each other. This effect may also destroy the existence of a least favorable measure; see Hernández-Hernández and Schied [42] for the discussion of a related tradeoff effect. See also Schied [62] for examples in the setting of Proposition 2.23, in which a path-dependent volatility σ_t or drift α_t^0 prevent the existence of a least favorable measure.

Examples within the Huber-Strassen theory

In the preceding section, the way of determining the set \mathcal{Q} was to specify a “confidence set” around an estimate of a certain market parameter and to take for \mathcal{Q} the class of all measures that are consistent with this confidence set. In practice, however, one would rather try to assign a high weight to the original estimate, while a measure concentrated on the outmost edge of the confidence set should receive a lower weight. This idea illustrates that the set \mathcal{Q} may arise in a more complicated manner from the investor's preference relation than in the ad hoc approach of the preceding section.

The complexity of determining the set \mathcal{Q} is reduced if one imposes additional assumptions on the underlying preference order. For instance, we have already discussed the assumption of *comonotonic independence*, which is reasonable insofar comonotonic positions cannot act as mutual hedges; see Remark 1.59. Mathematically, comonotonic independence is essentially equivalent to the fact that the nonadditive set function

$$\gamma(A) := \sup_{Q \in \mathcal{Q}} Q[A], \quad A \in \mathcal{F}_T,$$

is submodular in the sense of Choquet:

$$\gamma(A \cup B) + \gamma(A \cap B) \leq \gamma(A) + \gamma(B) \quad \text{for } A, B \in \mathcal{F}_T;$$

see Remark 1.59.

Assumption 2.24. *Consider the following set of conditions.*

- (a) γ is submodular.
- (b) \mathcal{Q} is maximal in the sense that it contains every measure Q with $Q[A] \leq \gamma(A)$ for all $A \in \mathcal{F}_T$.
- (c) There exists a Polish topology on Ω such that \mathcal{F}_T is the corresponding Borel field and \mathcal{Q} is compact.

Let us also comment on conditions (b) and (c) in Assumption 2.24. Condition (c) guarantees that γ is a capacity in the sense of Choquet [16]. Condition (b) implies that \mathcal{Q} is convex and closed in total variation. Hence, Lemma 2.18 yields that any least favorable measure must be equivalent to P^* . Moreover, under assumption (a), the set $\mathcal{Q} = \{Q \mid Q \leq \gamma\}$ is equal to

$$\left\{ Q \mid E_Q[X] \leq \int_0^\infty \gamma(X > t) dt \text{ for all } X \in L_+^\infty \right\}$$

see, e.g., [35, Theorem 4.88].

Consider the submodular set function

$$\nu_t(A) := t\gamma(A) - P^*[A], \quad A \in \mathcal{F}_T. \quad (85)$$

It is shown in Lemmas 3.1 and 3.2 of [46] that under Assumption 2.24 there exists a decreasing family $(A_t)_{t>0} \subset \mathcal{F}_T$ such that A_t minimizes ν_t and such that $A_t = \bigcup_{s>t} A_s$.

Definition 2.25 (Huber and Strassen). The function

$$\frac{dP^*}{d\gamma}(\omega) = \inf\{t \mid \omega \notin A_t\}, \quad \omega \in \Omega,$$

is called the *Radon-Nikodym derivative* of P^* with respect to γ .

The terminology ‘‘Radon-Nikodym derivative’’ comes from the fact that $dP^*/d\gamma$ coincides with the usual Radon-Nikodym derivative dP^*/dQ in case where $\mathcal{Q} = \{Q\}$; see [46]. We will need the following simple lemma:

Lemma 2.26. *Condition (73) implies that $P[0 < \frac{dP^*}{d\gamma} < \infty] = 1$.*

Proof. Let ν_t be as in (85). Clearly, $\frac{dP^*}{d\gamma}(\omega) = \infty$ if and only if $\omega \in A_\infty := \bigcap_{0 < t < \infty} A_t$. Since $\nu_t(A_t) \leq \nu_t(\emptyset) = 0$, we have $\gamma(A_t) \leq 1/t$. It follows that $\gamma(A_\infty) = 0$, which by (73) implies that $P[A_\infty] = 0$.

Letting $A_0 := \bigcup_{0 < t < \infty} A_t$, we see that $\frac{dP^*}{d\gamma}(\omega) = 0$ if and only if $\omega \in A_0^c$. From $\nu_t(A_t) \leq \nu_t(\Omega) = t - 1$, we find that $P^*[A_t^c] \leq t(1 - \gamma(A_t))$. As $t \downarrow 0$ we thus get $P^*[A_0^c] = 0$. \square

Let us now state the Huber-Strassen theorem from [46] in a form in which it will be needed here.

Theorem 2.27 (Huber-Strassen). *Under Assumption 2.24, \mathcal{Q} admits a least favorable measure Q_0 with respect to any probability measure R on (Ω, \mathcal{F}_T) . Moreover, if $R = P^*$ and \mathcal{Q} satisfies (73), then Q_0 is equivalent to P^* and given by*

$$dQ_0 = \left(\frac{dP^*}{d\gamma} \right)^{-1} dP^*.$$

Together with Theorem 2.17, we get a complete solution of the robust utility maximization problem within the large class of utility functionals that arise from sets \mathcal{Q} as in Assumption 2.24. Before discussing particular examples, let us state the following converse of the Huber-Strassen theorem in order to clarify the role of condition (a) in Assumption 2.24.

Theorem 2.28. *Suppose Ω is a Polish space with Borel field \mathcal{F}_T and \mathcal{Q} is a compact set of probability measures. If every probability measure on (Ω, \mathcal{F}_T) admits a least favorable measure $Q_0 \in \mathcal{Q}$, then $\gamma(A) = \sup_{Q \in \mathcal{Q}} Q[A]$ is submodular.*

For finite probability spaces, Theorem 2.28 is due to Huber and Strassen [46]. In the form stated above, it was proved by Lembcke [54]. An alternative formulation was given by Bednarski [10].

Let us now turn to the discussion of examples. The following example class was first studied by Bednarski [9] under slightly different conditions than here. These examples also play a role in the theory of law-invariant risk measures; see Kusuoka [53] and Sections 4.4 through 4.7 in [35].

Example 2.29. The following class of submodular set functions arises in Yaari's [68] "dual theory of choice under risk". Let $\psi : [0, 1] \rightarrow [0, 1]$ be an increasing concave function with $\psi(0) = 0$ and $\psi(1) = 1$. In particular, ψ is continuous on $(0, 1]$. We define γ by

$$\gamma(A) := \psi(P[A]), \quad A \in \mathcal{F}.$$

Then γ is submodular, and the set \mathcal{Q} of all probability measures Q on (Ω, \mathcal{F}_T) with $Q[A] \leq \gamma(A)$ can be described in terms of ψ ; see Theorem 1.51. If (Ω, \mathcal{F}_T) is a standard Borel space, then there exists a compact metric topology on Ω whose Borel field is \mathcal{F}_T . For such a topology, \mathcal{Q} is weakly compact, and so Assumption 2.24 is satisfied and \mathcal{Q} admits a least favorable measure Q_0 . It can be explicitly determined in the case in which $\psi(t) = (t\lambda^{-1}) \wedge 1$ for some $\lambda \in (0, 1)$. Indeed, it follows from Example 2.15 that the Radon-Nikodym derivative of P^* with respect to γ is given by

$$\pi = \frac{dP^*}{d\gamma} = c \cdot (\varphi \vee q_\varphi(t_\lambda)),$$

where c is the normalizing constant and t_λ is the unique maximizer of the function $t \mapsto (t-1+\lambda)^+ / \Phi(t)$.
 \diamond

Example 2.30 (Weak information). Let Y be a measurable function on (Ω, \mathcal{F}_T) , and denote by μ its law under P^* . For $\nu \sim \mu$ given, let

$$\mathcal{Q} := \left\{ Q \ll P^* \mid Q \circ Y^{-1} = \nu \right\}.$$

The robust utility maximization problem for this set \mathcal{Q} was studied by Baudoin [8], who coined the terminology "weak information". The interpretation behind the set \mathcal{Q} is that an investor has full knowledge about the pricing measure P^* but is uncertain about the true distribution P of market prices and only knows that a certain functional Y of the stock price has distribution ν .

Define Q_0 by

$$dQ_0 = \frac{d\nu}{d\mu}(Y) dP^*.$$

Then $Q_0 \in \mathcal{Q}$ and the law of $\pi := dQ_0/dP^* = d\mu/d\nu(Y)$ is the same for all $Q \in \mathcal{Q}$. Hence, Q_0 satisfies the definition of a least favorable measure. The same procedure can be applied to *any* measure $R \sim P^*$. In fact, \mathcal{Q} fits into the framework of the Huber-Strassen theory, as is shown in the following proposition. Least favorable measures for this setting of "weak information" and its generalizations were first analyzed in robust statistics; see Huber [45], section 10.3, and Plachky and Rüschemdorf [55].
 \diamond

Proposition 2.31. *Suppose (Ω, \mathcal{F}_T) is a standard Borel space. Then the set \mathcal{Q} defined in Example 2.30 satisfies Assumption 2.24. In particular, $\gamma(A) := \sup_{Q \in \mathcal{Q}} Q[A]$ is submodular.*

Proof. If Q is a probability measure with $Q[\cdot] \leq \gamma(\cdot)$, then

$$Q[Y \leq t] \leq \gamma(Y \leq t) = \nu((-\infty, t]).$$

Using the same argument on $\{Y > t\}$ shows that Y has law ν under Q . Hence, \mathcal{Q} is maximal in the sense of part (b) of Assumption 2.24.

To prove that part (a) holds we will use Theorem 2.28. To this end, we may choose a compact metric topology on Ω such that Y is continuous and \mathcal{F}_T is the Borel σ -algebra. Write $P^* = \mu K^* := \int \mu(dy) K^*(y, \cdot)$, where $K^*(y, \cdot) = P^*[\cdot | Y = y]$ is a regular conditional expectation given Y . If $R \ll P^*$, then $\eta := R \circ Y^{-1} \ll \nu$ and R can be written as ηK_R , where K_R is a stochastic kernel such that $K_R(y, \cdot) \ll K^*(y, \cdot)$ for η -a.e. y . Let $\nu = \nu_a + \nu_s$ be the Lebesgue decomposition of ν with respect to η into the absolutely continuous part $\nu_a \ll \eta$ and into the singular part ν_s . If we let $Q_0 := \nu_a K_R + \nu_s K^*$, then $Q_0 \in \mathcal{Q}$ and

$$\pi = \frac{dR}{dQ_0} = \frac{d\eta}{d\nu}(Y).$$

Again, the distribution of π is the same for all $Q \in \mathcal{Q}$, and it follows that Q_0 is a least favorable measure. If $R \not\ll P^*$, then it is clear that any measure Q_0 will be least favorable for R if it is least favorable for the absolutely continuous part of R . \square

In the 1970's and 1980's, explicit formulas for Radon-Nikodym derivatives with respect to capacities were found in a number of examples such as sets \mathcal{Q} defined in terms of ε -contamination or via probability metrics like total variation or Prohorov distance; we refer to Chapter 10 in the book [45] by Huber and the references therein. But, unless Ω is finite, these examples fail to satisfy either implication in (73). Nevertheless, they are still interesting for discrete-time market models.

References

- [1] Acerbi, C., Tasche, D., On the coherence of expected shortfall. *Journal of Banking and Finance* 26 (7), 1487-1503, 2002.
- [2] Anscombe, F. J., Aumann, R. J., A definition of subjective probability. *Ann. Math. Statistics* 34 (1963), 199-205.
- [3] Artzner, P., Delbaen, F., Eber, J.-M., Heath, D., Coherent measures of risk. *Math. Finance* 9 (1999), 203-228.
- [4] Artzner, P., Delbaen, F., Eber, J.-M., Heath, D., Ku, H. Coherent multiperiod risk adjusted values and Bellman's principle. To appear in *Annals of Operations Research*.
- [5] Artzner, P., Delbaen, F., Eber, J.-M., Heath, D., Ku, H. Coherent multiperiod risk measurement. Preprint, ETH Zürich (2003).
- [6] Artzner, P., Delbaen, F., Eber, J.-M., Heath, D., Ku, H., Multiperiod risk and coherent multiperiod risk measurement. Preprint, ETH Zürich (2003).
- [7] Barrieu, P., El Karoui, N. Inf-convolution of risk measures and optimal risk transfer. *Finance Stoch.* 9, no. 2, 269-298 (2005).
- [8] Baudoin, F. *Conditioned stochastic differential equations: theory, examples and application to finance*. *Stochastic Process. Appl.* 100, 109-145 (2002).
- [9] T. Bednarski, On solutions of minimax test problems for special capacities, *Z. Wahrsch. Verw. Gebiete* 58 (1981) 397-405.
- [10] T. Bednarski, Binary experiments, minimax tests and 2-alternating capacities, *Ann. Statist.* 10 (1982) 226-232.
- [11] Carlier, G., Dana, R. A. Core of convex distortions of a probability. *J. Econom. Theory* 113 (2003), no. 2, 199-222.
- [12] Carlier, G., Dana, R. A. Insurance contracts with deductibles and upper limits. Preprint, Ceremade, Université Paris Dauphine (2002).
- [13] Cheridito, P., Delbaen, F., Kupper, M. Coherent and convex monetary risk measures for bounded càdlàg processes. *Stoch. Proc. Appl.* 112 (2004), 1-22.
- [14] Cheridito, P., Delbaen, F., Kupper, M. Coherent and convex monetary risk measures for unbounded càdlàg processes. *Finance Stoch.* 9 (2005), 1713-1732.
- [15] Cheridito, P., Delbaen, F., Kupper, M. Dynamic monetary risk measures for bounded discrete-time processes. *Electronic J. Probab.* 11 (2006), 57-106.

- [16] Choquet, G. Theory of capacities. *Ann. Inst. Fourier* 5, 131-295 (1953/54)
- [17] R. Cont, Model uncertainty and its impact on the pricing of derivative instruments. *Math. Finance* 16, 519-542, (2006).
- [18] Cvitanic, J., Karatzas, I., Generalized Neyman-Pearson lemma via convex duality. *Bernoulli* 7 (2001), 79–97.
- [19] Dana, R.-A., A representation result for concave Schur concave functions. *Math. Finance* 15 (2005), 613-634.
- [20] Delbaen, F., Coherent measures of risk on general probability spaces. In: *Advances in Finance and Stochastics. Essays in Honour of Dieter Sondermann*, Springer-Verlag, 2002, 1–37.
- [21] Delbaen, F., Coherent risk measures. *Cattedra Galileiana. Scuola Normale Superiore, Classe di Scienze*, Pisa, 2000.
- [22] Delbaen, F., The structure of m -stable sets and in particular of the set of riskneutral measures. To appear in: *Seminaire Probab.* 39 (2006).
- [23] Denneberg, D., Non-additive measure and integral. *Theory Decis. Lib. Ser. B Math. Statist. Meth.* 27, Kluwer Academic Publishers, Dordrecht, 1994.
- [24] Dunford, N., Schwartz, J., *Linear Operators. Part I: General Theory*. Interscience Publishers, New York, 1958.
- [25] A. Eichhorn and W. Römisch. Polyhedral risk measures in stochastic programming. *SIAM J. Optim.* 16, no. 1, 69–95 (2005).
- [26] Ekeland, I., Temam, R., *Convex analysis and variational problems*. *Stud. Math. Appl.* 1, North-Holland Publishing Co., Amsterdam–Oxford; American Elsevier Publishing Co., Inc., New York, 1976.
- [27] N. El Karoui, M. Jeanblanc-Picqué, and S. Shreve, Robustness of the Black and Scholes formula, *Math. Finance* 8 (1998) 93–126.
- [28] Floret, K., *Weakly compact sets*. *Lecture Notes in Math.* 801, Springer-Verlag, Berlin, 1980.
- [29] Föllmer, H., Probabilistic Aspects of Financial Risk. Plenary Lecture at the Third European Congress of Mathematics. *Proceedings of the European Congress of Mathematics, Barcelona 2000*, Birkhäuser, Basel, 2001.
- [30] Föllmer, H., Gundel, A. *Robust projections in the class of martingale measures*. To appear in *Illinois J. Math.*
- [31] Föllmer, H., Kramkov, D., Optional decompositions under constraints. *Probab. Theory Related Fields* 109 (1997), 1–25.
- [32] Föllmer, H., Leukert, P., Efficient hedging: cost versus shortfall risk. *Finance Stoch.* 4 (2000), 117–146.
- [33] Föllmer, H., Schied, A., Convex measures of risk and trading constraints. *Finance Stoch.* 6, 429–447. (2002).
- [34] H. Föllmer and A. Schied, *Stochastic Finance: An Introduction in Discrete Time*, Walter de Gruyter & Co., Berlin, de Gruyter Studies in Mathematics 27, 2002.
- [35] H. Föllmer and A. Schied, *Stochastic Finance: An Introduction in Discrete Time*, 2nd revised and extended edition. Walter de Gruyter & Co., Berlin, de Gruyter Studies in Mathematics 27, 2004.
- [36] H. Föllmer and A. Schied, Robust preferences and convex measures of risk. *Advances in finance and stochastics*, 39–56, Springer, Berlin, 2002.
- [37] Frittelli, M., Rosazza Gianin, E., Putting order in risk measures. *J. Banking & Finance* 26 (2002), 1473-1486.
- [38] Frittelli, M., Rosazza Gianin, E., Dynamic convex risk measures. In: *New Risk Measures in Investment and Regulation*, G. Szegö ed., John Wiley & Sons, New York, 2003.
- [39] Frittelli, M., Scandolo, G., Risk measures and capital requirements for processes. To appear in *Math. Finance*.
- [40] Gilboa, I., Schmeidler, D., Maxmin expected utility with non-unique prior. *J. Math. Econom.* 18 (1989), 141–153.
- [41] A. Gundel, Robust utility maximization in complete and incomplete market models, *Finance Stoch.* 9, no. 2, 151–176 (2005).

- [42] Hernández-Hernández, D., Schied, A., Robust utility maximization in a stochastic factor model. *Statistics & Decisions* 24, No. 3, 109-125 (2006).
- [43] Heath, D., Back to the future. Plenary lecture, First World Congress of the Bachelier Finance Society, Paris (2000).
- [44] Heath, D., Ku, H., Pareto equilibria with coherent measures of risk. *Math. Finance* 14, no. 2 (2004), 163-172.
- [45] Huber, P., Robust statistics. Wiley Ser. Probab. Math. Statist., Wiley, New York, 1981.
- [46] Huber, P., Strassen, V., Minimax tests and the Neyman-Pearson lemma for capacities. *Ann. Statist.* 1 (1973), 251-263.
- [47] Jaschke, S., Küchler, U. Coherent Risk Measures and Good-Deal Bounds. *Finance Stoch.* 5, no. 2 (2001), 181-200
- [48] E. Jouini, W. Schachermayer, N. Touzi: Law invariant risk measures have the Fatou property. Preprint, Université Paris Dauphine (2005).
- [49] Komlos, J., A generalization of a problem of Steinhaus. *Acta Math. Acad. Sci. Hung.* 18 (1967), 217-229.
- [50] D. Kramkov and W. Schachermayer, The asymptotic elasticity of utility functions and optimal investment in incomplete markets, *Ann. Appl. Probab.* 9 (1999) no. 3, 904-950.
- [51] Kreps, D., Notes on the theory of choice. Westview Press, Boulder, CO, 1988.
- [52] Kunze, M. Verteilungsinvariante konvexe Risikomaße. Diplomarbeit, Humboldt-Universität zu Berlin, 2003.
- [53] Kusuoka, S., On law invariant coherent risk measures. *Adv. Math. Econ.* 3 (2001), 83-95.
- [54] J. Lembcke, The necessity of strongly subadditive capacities for Neyman-Pearson minimax tests, *Monatsh. Math.* 105 (1988) 113-126.
- [55] Plachky, D., Rüschemdorf, L. Conservation of the UMP—resp. maximin-property of statistical tests under extensions of probability measures. Goodness-of-fit (Debrecen, 1984), 439-457, *Colloq. Math. Soc. Janos Bolyai*, 45, North-Holland, Amsterdam, 1987.
- [56] Quenez, M., Optimal portfolio in a multiple-priors model. Seminar on Stochastic Analysis, Random Fields and Applications IV, 291-321, *Progr. Probab.*, 58, Birkhäuser, Basel, 2004.
- [57] Riedel, F., Dynamic coherent risk measures. *Stochastic Process. Appl.* 112 (2004), no. 2, 185-200.
- [58] Rüschemdorf, L. Stochastically ordered distributions and monotonicity of the OC of an SPRT. *Math. Operationsforschung, Series Statistics*, Vol. 12 (1981), 327-338
- [59] Ryff, J. Orbits of L^1 -functions under doubly stochastic transformations. *Trans. Amer. Math. Soc.* 117 (1965), 92-100.
- [60] Savage, L. J., The foundations of statistics. Wiley Publ. Stat., John Wiley and Sons, New York, 1954.
- [61] A. Schied, On the Neyman-Pearson problem for law-invariant risk measures and robust utility functionals, *Ann. Appl. Probab.* 14 (2004) 1398-1423.
- [62] Schied, A., Optimal investments for robust utility functionals in complete market models. *Math. Oper. Research.* 30, No. 3, 750-764 (2005).
- [63] Schied, A. Optimal investments for risk- and ambiguity-averse preferences: a duality approach. To appear in *Finance Stoch.*
- [64] A. Schied and C.-T. Wu, Duality theory for optimal investments under model uncertainty, *Stat. Decisions* 23, No. 3, 199-217 (2005).
- [65] Schmeidler, David Subjective probability and expected utility without additivity. *Econometrica* 57 (1989), no. 3, 571-587.
- [66] Sekine, J., Dynamic minimization of worst conditional expectation of shortfall. *Math. Finance.* 14 (2004), pp.605-618.
- [67] Weber, S., Distribution-invariant risk measures, information, and dynamic consistency. *Math. Finance* 16, no. 2, 419-441 (2006).
- [68] Yaari, M., The dual theory of choice under risk. *Econometrica* 55 (1987), 95-116.