University of Zurich
Institute of Mathematics

Master's Thesis

Concave Distortions and Comonotonic Risk Measures: An Application to Distributionally Robust Insurance

---

Supervisor:
Prof. Dr. Delia Marina Coculescu

Author:
Maria Elena Zeppa

Co-Supervisor:
Sabina-Ligia Georgescu

August 2026

Acknowledgements

This thesis marks the end of my studies in mathematics, and it would not have come into being without the people who accompanied me along the way. To all of them, I would like to say thank you.

First, I would like to thank my supervisor, Prof. Dr. Delia Coculescu, for giving me the opportunity to write my Master thesis under her guidance in this fascinating area of mathematical finance. Working on this topic has been a truly enriching experience, and it has strengthened my wish to pursue a career in this field.

My sincere gratitude also goes to my co-supervisor, Sabina-Ligia Georgescu. Sabina, thank you for standing by my side throughout this project, for taking the time to answer my questions, no matter how big or small, for your thoughtful feedback that improved this thesis tremendously, and for your constant encouragement. I could not have wished for a better mentor during this time.

Finally, I want to thank my family and my friends, who have had my back not only during this semester but throughout my entire studies. You were there for me in the moments when things were most difficult, you believed in me when I doubted myself, and you never let me forget that there is a life beyond mathematics. This thesis is, in many ways, also yours.

Thank you.

Abstract

This thesis develops the theory of convex distortion risk measures and applies it to a problem of optimal insurance design under model uncertainty. Starting from the duality theory of convex analysis, we first present the axiomatic framework of monetary, convex, and coherent risk measures in the tradition of Föllmer and Schied, including their characterization via acceptance sets and their robust representations over sets of finitely additive probability measures. We then study distortion risk measures through the Choquet integral and examine the special role played by concave distortion functions. In a complementary step, we analyze comonotone additive risk measures and show how comonotonicity forces the Choquet representation. Combining both perspectives yields the central equivalence of this thesis: on an atomless probability space, a risk measure on L^∞ is convex, law-invariant, and comonotonic if and only if it is a Choquet integral with respect to a concave distortion, or equivalently a mixture of Average Value at Risk. Every such risk measure is automatically coherent.

The second part of the thesis demonstrates that these structural properties are precisely what renders a recent distributionally robust insurance problem tractable. Following Boonen and Jiang, we consider a decision maker who evaluates the retained loss by a convex distortion risk measure and hedges against all loss distributions within a Wasserstein ball around a benchmark model. We present the closed-form solution of this minimax problem, an optimal stop-loss contract together with an explicit worst-case distribution, and reproduce the accompanying numerical study for the GlueV@R risk measure, including a comparison of the Wasserstein and L^2 ambiguity sets. The results illustrate how ambiguity distorts the loss distribution only where risk is actually retained, and how the theory of comonotonic convex risk measures translates into explicit, economically interpretable insurance contracts.

Contents

1 Introduction **1**

1.1 Motivation . . . 1

1.2 Survey of the Extant Literature . . . 2

1.3 Main Objective of this Thesis . . . 3

2 Preliminaries **6**

2.1 Basic Notions of Convex Analysis . . . 7

2.1.1 Banach Spaces and Continuous Linear Functionals . . . 7

2.1.2 Convex and Concave Functions . . . 9

2.2 Topological Vector Spaces and Convex Duality Tools . . . 11

2.2.1 Locally Convex Spaces and Weak Topologies . . . 11

2.2.2 Separation and the Closure of Convex Sets . . . 13

2.3 Compactness in the Weak* Topology . . . 14

2.4 Finitely Additive Set Functions and the Dual of $\mathcal{X}$ . . . 16

2.5 Convex Duality and the Fenchel–Legendre Transform . . . 19

2.5.1 The Fenchel–Legendre Transform . . . 19

2.5.2 Synthesis: A Strategy for the Robust Representation . . . 21

3 Risk Measures **23**

3.1 Foundations . . . 23

3.1.1 Definition of Monetary Risk Measures . . . 24

3.1.2 Convex and Coherent Risk Measures . . . 26

3.1.3 Characterization via Acceptance Sets . . . 29

3.2 Robust Representation of Convex Risk Measures . . . 35

3.2.1 Representation Theorems . . . 36

3.2.2 Continuity Properties . . . 41

3.3 Convex Risk Measures on L^∞ . . . 49

3.4 Examples . . . 54

3.4.1 V@R: Value at Risk . . . 55

3.4.2 AV@R: Average Value at Risk . . . 57

3.5 Law-Invariant Risk Measures . . . 60

3.5.1 Law-Invariance and the Fatou Property . . . 60

3.5.2 Robust Representation of Law-Invariant Risk Measures 61

4 Distortion Risk Measures **66**
4.1 Distortion Functions and the Choquet Integral 66
4.2 Choquet Representation 71
4.3 Concave Distortions 75

5 Comonotone Additive Risk Measures **79**
5.1 Motivation and Comonotonicity 79
5.2 Choquet and Quantile Representation 81
5.3 Characterization of Comonotonic Convex Risk Measures 86

6 Distributionally Robust Insurance under the Wasserstein Distance **90**
6.1 Introduction and Practical Context 90
6.2 Model Setup and Problem Formulation 91
6.3 Solution Approach and Main Results 96
6.4 A Numerical Case Study 98
6.5 Discussion and Concluding Remarks 103

7 Conclusion and Outlook **106**

A Probability-Theoretic Framework **108**
A.1 Probability Spaces and Random Variables 108
A.2 Expectation and Integrable Random Variables 115
A.3 Important Integration Results 118
A.4 Change of Probability Measure 120

B Additional Material on Risk Measures **123**
B.1 Supplementary Results 123
B.2 Technical Proofs 130

C Technical Details of the Application **139**
C.1 Auxiliary Results from Optimization 139
C.1.1 The Minimax Theorem 139
C.1.2 Slater's Condition and Strong Duality 140
C.2 Implementation Code 141

References **142**

Declarations **146**

Chapter 1

Introduction

1.1 Motivation

How risky is a financial position? Any attempt to answer this question quantitatively requires a rule that assigns to each uncertain payoff a single number – its risk. Such rules, called risk measures, are of central importance in modern finance and insurance as they determine the capital a bank must hold against its trading book, the reserves an insurer must set aside against future claims, and, more generally, the acceptability of a risky position. Given their practical weight, the question of which properties a reasonable risk measure should satisfy, and what structure these properties impose, is a natural object of mathematical study.

A first answer is provided by the axiomatic theory of coherent and, more generally, convex risk measures. In this framework, a risk measure is a functional on a space of financial positions satisfying a small number of economically motivated axioms, such as monotonicity, cash invariance, and convexity, the latter reflecting the principle that diversification should not increase risk. The perhaps most striking insight of this theory is its duality. Under suitable continuity assumptions, every convex risk measure admits a robust representation as a worst-case expected loss over a family of probabilistic models, penalized according to their plausibility. A risk measure can thus be understood as the assessment of an agent who does not trust a single model but guards against an entire collection of them. Model uncertainty, or ambiguity, is in this sense built into the very structure of convex risk measures.

Within this general theory, a particular class stands out both for its analytical tractability and its practical prevalence: the distortion risk measures. These are constructed by distorting the underlying probability distribution with an increasing function ψ and computing the expected loss under the distorted law. Formally, a Choquet integral with respect to the set function $\psi \circ \mathbb{P}$. Prominent regulatory and industry standards, including Value at Risk, Average Value at Risk, and the more recent GlueV@R family, are all of this type. Distortion risk measures also possess a distinctive structural property. They are additive on comonotone positions, that is, on positions

that move together and therefore offer no mutual hedge. It is a remarkable fact that this property essentially characterizes them, and that convexity of the induced risk measure corresponds exactly to concavity of the distortion function. The interplay between convexity, law-invariance, comonotonicity, and concave distortions forms the theoretical core of this thesis.

The value of this structural theory becomes apparent when it is put to work. Optimal insurance design provides a natural testing ground, where a decision maker exposed to a random loss purchases partial coverage from an insurer, choosing the indemnity schedule that minimizes the risk of the retained position. In practice, however, the loss distribution on which such an optimization relies is itself estimated from data and hence uncertain. This difficulty is made concrete by rare but severe events that historical records capture poorly. The distributionally robust approach responds to this by seeking contracts that perform well over an entire ball of plausible loss distributions, for instance a ball in the Wasserstein distance around a benchmark model. Recently, Boonen and Jiang [11] solved this problem in closed form for decision makers using convex distortion risk measures. Their analysis is a showcase for the theory developed here: comonotone additivity, the Choquet representation, and concave distortions are precisely the ingredients that make the minimax problem tractable and its solution interpretable. This thesis develops the theory from its convex-analytic foundations up to this application.

1.2 Survey of the Extant Literature

The axiomatic study of risk measures begins with the seminal paper of Artzner, Delbaen, Eber and Heath [3], who introduced coherent risk measures on finite probability spaces and established their representation as worst-case expectations over a set of probabilistic scenarios. Delbaen [19] extended the theory to general probability spaces, characterizing the coherent risk measures on L^∞ that admit a representation in terms of σ-additive measures via the Fatou property. The generalization from coherent to convex risk measures, in which positive homogeneity and subadditivity are relaxed to convexity, was carried out independently by Föllmer and Schied [25] and by Frittelli and Rosazza Gianin [27], who obtained the robust representation with a penalty function that quantifies the plausibility of each model. A comprehensive treatment of this theory, whose presentation we largely follow in this thesis, is the monograph of Föllmer and Schied [26].

A second strand of the literature concerns law-invariant risk measures, which depend on a position only through its distribution. Kusuoka [35] proved the fundamental result that, on an atomless probability space, every law-invariant coherent risk measure that is comonotonic is a mixture of Average Value at Risk; see also Frittelli and Rosazza Gianin [28] for the convex case and Shapiro [47] for a discussion of the Kusuoka representation.

Jouini, Schachermayer and Touzi [33] showed that law-invariant convex risk measures on L^∞ automatically satisfy the Fatou property, so that the robust representation in terms of σ-additive measures comes for free in the law-invariant setting; continuity properties of such functionals were further studied by Svindland [48]. The Average Value at Risk itself, the fundamental building block of this theory, was analyzed from an optimization perspective by Rockafellar and Uryasev [41, 42], and the closely related spectral risk measures were introduced by Acerbi [1].

The third ingredient of our theory, the Choquet integral, has its origins outside of mathematical finance. Schmeidler [45, 46] established the integral representation of comonotone additive functionals with respect to non-additive set functions and used it as the foundation of a theory of decision making under ambiguity; a systematic mathematical treatment of non-additive measure theory is given by Denneberg [20] and, more recently, by Grabisch [29]. In the actuarial literature, distortion risk measures were introduced as premium principles by Wang, Young and Panjer [51], who characterized insurance prices axiomatically via comonotone additivity. The GlueV@R family employed in the numerical part of this thesis was proposed by Belles-Sampera, Guillén and Santolino [7] as a flexible parametric class interpolating between Value at Risk and Average Value at Risk.

Optimal insurance design has a long history, going back at least to the classical result that stop-loss contracts are optimal for risk-averse expected-utility maximizers; see Huberman, Mayers and Smith [32] for an early treatment. In the risk-measure framework, optimal (re)insurance with distortion risk measures and distortion premium principles was studied, among others, by Assa [4], Cheung, Chong and Yam [15], and Zhuang, Weng, Tan and Assa [52], the latter introducing the marginal indemnification function approach that also underlies the solution technique of the present application. The robustness of distortion risk measures themselves to perturbations of the underlying distribution was investigated by Bernard, Pesenti and Vanduffel [8]. Distributionally robust formulations of the insurance problem are more recent: Boonen and Jiang studied robust insurance design under the L^2 distance in [10] and under the Wasserstein distance in [11]. The Wasserstein distance, whose theory is developed in the monograph of Villani [50] and surveyed from a statistical perspective by Panaretos and Zemel [39], has become the standard choice of ambiguity metric in distributionally robust optimization, as Wasserstein balls contain discrete and continuous perturbations of the benchmark alike and lead to tractable dual reformulations.

1.3 Main Objective of this Thesis

The objective of this thesis is twofold. The first, theoretical objective is to give a self-contained development of the theory of convex distortion risk measures, culminating in a complete characterization of the law-invariant convex risk measures that are

comonotonic. The main result of the theoretical part, Theorem 5.16, establishes that on an atomless probability space the following three descriptions single out one and the same class of functionals on L^∞: (i) the convex, law-invariant, comonotonic risk measures; (ii) the Choquet integrals of the loss with respect to a concave distortion of the underlying probability; and (iii) the mixtures of Average Value at Risk. In particular, every risk measure in this class is automatically coherent. Figure 1.1 illustrates this equivalence schematically. While the individual implications are known since they combine results of Kusuoka [35], Schmeidler [45], and the treatment in Föllmer and Schied [26], the emphasis of this thesis lies on assembling them into a single coherent narrative, with complete proofs, from the convex duality underlying the robust representation up to the quantile-based Choquet calculus.

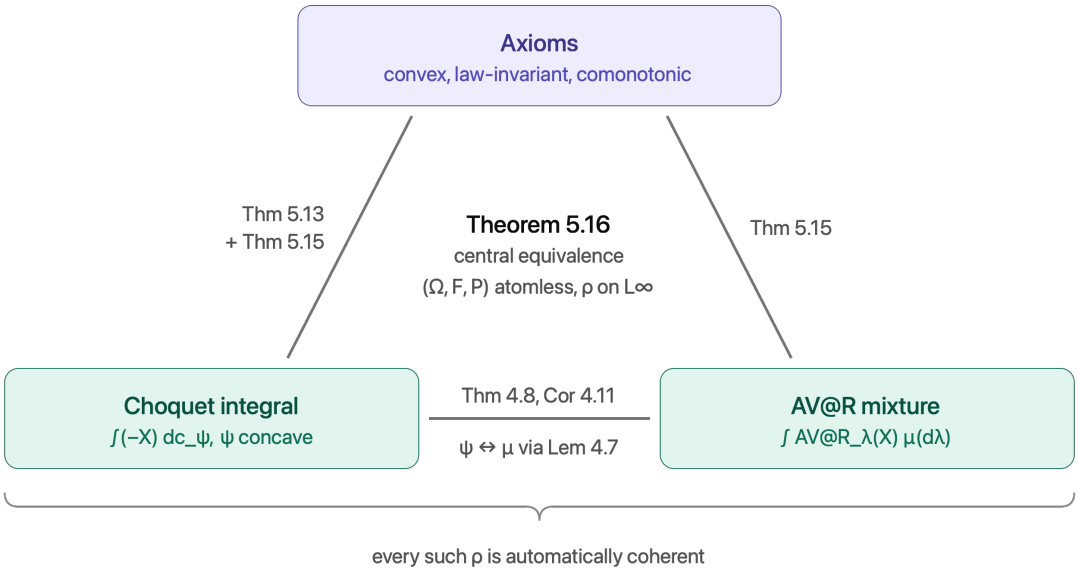

Figure 1.1: Schematic overview of the main result.

The second, applied objective is to demonstrate that this characterization is not merely an elegant structural statement but the key to solving a concrete problem of current research interest. Following Boonen and Jiang [11], we study a decision maker who evaluates his retained loss by a convex distortion risk measure and seeks the optimal indemnity function when the loss distribution is only known to lie within a Wasserstein ball around a benchmark model. We present the model, summarize the solution strategy and state the closed-form results, where the optimal contract is a stop-loss agreement, and the worst-case distribution is obtained explicitly from a constrained isotonic projection. We then reproduce the numerical study of [11] for the GlueV@R risk measure, examine how the worst-case distribution responds to the safety loading and the risk-aversion parameters, and compare the Wasserstein setting with its L^2 counterpart from [10].

The thesis is organized as follows. Chapter 2 collects the required tools from

functional and convex analysis, including Banach spaces and their duals, weak and weak* topologies, separation theorems, weak* compactness, the identification of the dual of the space of bounded measurable functions with the finitely additive set functions, and the Fenchel–Legendre transform, together with a synthesis of how these tools combine into a strategy for the robust representation. Chapter 3 develops the theory of monetary, convex, and coherent risk measures, their characterization via acceptance sets, their robust representations and continuity properties, the theory on L^∞, the examples of Value at Risk and Average Value at Risk, and the representation of law-invariant risk measures. Chapter 4 introduces distortion functions and the Choquet integral and studies the correspondence between concave distortions and mixtures of Average Value at Risk. Chapter 5 treats comonotonicity, proves the Choquet and quantile representations of comonotone additive risk measures, and establishes the central equivalence theorem. Chapter 6 is devoted to the distributionally robust insurance problem and its numerical study. Chapter 7 concludes and discusses directions for further research. Three appendices collect the probabilistic background, supplementary results and technical proofs on risk measures, and the auxiliary optimization results together with the implementation code of the numerical study.

A note on sources. The theoretical part of this thesis (Chapters 2 to 5) is an exposition of known results and follows established sources. Unless a statement carries an explicit citation, the presentation is based on the monograph *Stochastic Finance: An Introduction in Discrete Time* by Föllmer and Schied [26], which serves as the default reference throughout. Chapter 6 is based on the article of Boonen and Jiang [11].

Chapter 2

Preliminaries

In this chapter, we collect the fundamental definitions and mathematical results that are required for the development of the subsequent chapters. The material presented here serves as the theoretical foundation upon which the main results of this thesis are built.

We assume that the reader is familiar with the basic concepts of probability theory, mathematical finance, and analysis. In particular, notions such as measurable spaces, random variables, integration theory, as well as elements from functional and convex analysis are considered as known. The purpose of this chapter is therefore not to provide a comprehensive introduction, but rather to fix notation, recall key concepts, and highlight those results that are essential for the understanding of the theory developed later on. In many cases, we restrict ourselves to stating definitions and results without proofs, as they are standard and can be found in the literature.

The selection of topics is guided by the following observation. Risk measures, the central object of study in this thesis, are real-valued functionals defined on infinite-dimensional spaces of financial positions. Their fundamental properties, most notably the robust representation results developed in Chapter 3 and the Choquet representations of Chapters 4 and 5, rest on the interplay of three structures: a *Banach space structure* on the underlying space of positions, the *convexity* of the functional, and a suitable *topology* in which level sets are closed. The bridge between these structures is provided by *convex duality*, in particular by the Fenchel–Legendre transform. Accordingly, this chapter develops the necessary tools from functional and convex analysis, organized along these lines.

We begin in Section 2.1 with Banach spaces, continuous linear functionals, and basic notions of convex analysis. Section 2.2 introduces locally convex spaces, weak and weak* topologies, the Hahn–Banach separation theorems, and the fundamental relation between strong and weak closedness for convex sets. Section 2.3 addresses compactness in the weak* topology through the Banach–Alaoglu and Krein–Šmulian theorems, which will later guarantee that suprema in robust representations are actually attained. Finally, Section 2.4 then specializes to the identification of the topological

dual of the space of bounded measurable functions with the space of finitely additive set functions, a result that underlies the entire dual theory of risk measures on $\mathcal{X}$.

For the reader's convenience, a collection of fundamental results from probability theory is provided in Appendix A. This includes definitions and results concerning probability spaces, random variables, expectations, L^p-spaces, and changes of probability measure. These results are used throughout the main text and complement the functional and convex analytic material developed in the present chapter.

2.1 Basic Notions of Convex Analysis

This section introduces the basic vocabulary of functional analysis that will accompany us throughout the thesis, such as Banach spaces, continuous linear functionals, dual spaces, and the fundamental notions of convexity. The financial positions we will study in Chapter 3 are bounded random variables, and the risk measures we associate with them are convex functionals on the corresponding function space. To set up this framework precisely, we need to specify the space of positions, the topology in which it lives, and the class of admissible functionals on it. We begin with the underlying spaces.

2.1.1 Banach Spaces and Continuous Linear Functionals

Throughout this thesis, all vector spaces are understood to be defined over the field of real numbers. We recall the most basic notions from linear analysis.

Definition 2.1. [14] Let E be a vector space. A *norm* on E is a mapping $\|\cdot\| : E \to \mathbb{R}_+$ satisfying, for all $x, y \in E$ and $\lambda \in \mathbb{R}$:

(i) $\|x\| = 0$ if and only if $x = 0$;

(ii) $\|\lambda x\| = |\lambda| \, \|x\|$;

(iii) $\|x + y\| \leq \|x\| + \|y\|$.

A vector space equipped with a norm is called a *normed vector space*.

Every norm induces a metric $d(x, y) := \|x - y\|$ and hence a topology on E, under which the vector space operations are continuous [14]. A normed vector space in which every Cauchy sequence converges is called *complete* [43].

Definition 2.2. [14] A normed vector space $(E, \|\cdot\|)$ is called a *Banach space* if it is complete.

The most important example for our purposes is the following. Given a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, we denote by $\mathcal{L}^p(\Omega, \mathcal{F}, \mathbb{P})$ the set of all $\mathcal{F}$-measurable random variables Z such that $\|Z\|_p < \infty$, where

$$\|Z\|_p := \begin{cases} \mathbb{E}[|Z|^p]^{1/p}, & \text{if } 0 < p < \infty, \\ \inf\{c \geq 0 \mid \mathbb{P}[|Z| > c] = 0\}, & \text{if } p = \infty. \end{cases} \tag{2.1}$$

For $p \in [1, \infty]$, the space $L^p(\Omega, \mathcal{F}, \mathbb{P})$, obtained from $\mathcal{L}^p$ by identifying random variables that agree $\mathbb{P}$-almost surely, is a Banach space with respect to $\|\cdot\|_p$. A more detailed treatment of L^p-spaces is provided in Appendix A.

A second example, which will be essential in the development of risk measures, is the space

$$\mathcal{X} := \{F : \Omega \to \mathbb{R} \mid F \text{ is bounded and } \mathcal{F}\text{-measurable}\}, \tag{2.2}$$

equipped with the supremum norm $\|F\| := \sup_{\omega \in \Omega} |F(\omega)|$, under which $\mathcal{X}$ is a Banach space. The space $\mathcal{X}$ models bounded financial positions on a measurable space $(\Omega, \mathcal{F})$ without reference to any specific probability measure, and will serve as the default domain on which we define risk measures in Section 3.1.

We now turn to the linear functions on such spaces.

Definition 2.3. [14] Let E and F be vector spaces. A function $T : E \to F$ is called *linear* if

$$T(\lambda x + y) = \lambda T(x) + T(y) \quad \text{for all } x, y \in E,\ \lambda \in \mathbb{R}.$$

A linear function $\ell : E \to \mathbb{R}$ is called a *linear functional* on E.

In a Banach space, the relevant linear functionals are those that are compatible with the topology, that is, the continuous ones. Together, these functionals form a vector space on which a natural norm can be defined in terms of the underlying Banach space norm.

Definition 2.4. [14, 26] Let E be a Banach space. The *topological dual* of E is the vector space

$$E^* := \{\ell : E \to \mathbb{R} \mid \ell \text{ is linear and continuous}\},$$

endowed with the dual norm

$$\|\ell\|_{E^*} := \sup_{\|x\|_E \leq 1} \ell(x), \qquad \ell \in E^*. \tag{2.3}$$

The role of the dual space E^* in convex analysis is to provide the appropriate setting for the linearization of convex functionals on E, a perspective that will be made precise by the Fenchel–Legendre transform. In the context of risk measures developed in Chapter 3, elements of E^* will play the role of probabilistic models against which the risk of a financial position is evaluated.

The identification of E^* depends sensitively on the choice of E. For the spaces L^p, however, the situation is well understood. If $p \in [1, \infty)$ and q is the conjugate exponent satisfying $\frac{1}{p} + \frac{1}{q} = 1$, then the dual of L^p can be identified with L^q via the pairing

$$\ell(X) = \mathbb{E}[XY], \qquad X \in L^p, \tag{2.4}$$

for some $Y \in L^q$. The case $p = \infty$, however, is genuinely more delicate. While L^∞ is itself the dual of L^1, the dual of L^∞ is strictly larger than L^1 and is described by a space of finitely additive set functions. The analogous identification for the space $\mathcal{X}$ of bounded measurable functions is the content of Theorem 2.25 in Section 2.4: the dual $\mathcal{X}^*$ consists of finitely additive set functions of finite total variation, and the duality pairing is given by integration.

2.1.2 Convex and Concave Functions

We now turn to the second pillar of our framework: convexity. The fundamental property that distinguishes convex risk measures from arbitrary monetary risk measures is the principle that diversification should not increase risk, which is encoded mathematically in the convexity of the risk functional. The following definitions, formulated in the generality of an arbitrary real vector space, will apply both to finite-dimensional auxiliary functions such as the distortion functions of Chapter 4 and to functionals on Banach spaces of financial positions.

Definition 2.5. A subset C of a vector space E is called *convex* if $\lambda x + (1 - \lambda)y \in C$ for all $x, y \in C$ and $\lambda \in [0, 1]$.

Definition 2.6. Let C be a convex subset of a vector space E. A function $f : C \to [-\infty, +\infty]$ is called *convex* if its *epigraph*

$$\operatorname{epi} f := \{(x, \alpha) \in E \times \mathbb{R} \mid f(x) \leq \alpha\}$$

is a convex subset of $E \times \mathbb{R}$. The *effective domain* of f is

$$\operatorname{dom} f := \{x \in C \mid f(x) < +\infty\}.$$

A convex function f is called *proper* if $\operatorname{dom} f \neq \emptyset$ and f does not take the value $-\infty$. A function $u : C \to [-\infty, +\infty]$ is called *concave* if $-u$ is convex, and *proper concave* if $-u$ is proper convex.

For a function f taking values in $\mathbb{R}$ (rather than $[-\infty, +\infty]$), convexity admits the more familiar characterization

$$f(\lambda x + (1 - \lambda)y) \leq \lambda f(x) + (1 - \lambda) f(y), \qquad x, y \in C,\ \lambda \in [0, 1]. \tag{2.5}$$

The two definitions agree whenever f does not take both values $\pm\infty$.

A weaker condition, which will appear naturally in the study of distortion functions in Chapter 4, is the following.

Definition 2.7. [14] A function $f : E \to \mathbb{R} \cup [-\infty, +\infty]$ on a vector space E is called *quasi-convex* if all sublevel sets $\{x \in E \mid f(x) \leq c\}$ for $c \in \mathbb{R}$ are convex. The function f is called *quasi-concave* if $-f$ is quasi-convex.

We close this section with the notion and application of lower semicontinuity, which will play a recurring role throughout the thesis. While semicontinuity is a topological concept, we state it here for $\mathbb{R}^d$; the topological version, which we will need for risk measures on Banach spaces, is given in Section 2.2.1.

Definition 2.8. [14] A function $f : \mathbb{R}^d \to [-\infty, +\infty]$ is called *lower semicontinuous* (*lsc*) at $x_0 \in \mathbb{R}^d$ if

$$f(x_0) \leq \liminf_{x \to x_0} f(x).$$

The function f is *lsc* if it is lsc at every $x_0 \in \mathbb{R}^d$. It is called *upper semicontinuous* (*usc*) at x_0 if $-f$ is lsc at x_0, and usc if it is usc everywhere.

A function f is lsc if and only if all its sublevel sets $\{x \mid f(x) \leq c\}$, $c \in \mathbb{R}$, are closed, or equivalently, if and only if its epigraph $\operatorname{epi} f$ is a closed subset of $\mathbb{R}^d \times \mathbb{R}$ [14].

For our purposes, it will be important to understand the regularity properties of convex functions on the real line, since several key constructions in later chapters are built from one-dimensional convex or concave functions. The following classical result of convex analysis states that proper convex functions on $\mathbb{R}$ admit one-sided derivatives at every interior point of their effective domain.

Proposition 2.9. *Let f be a proper convex function on $\mathbb{R}$, and denote by D the interior of* $\operatorname{dom} f$. *Then:*

(a) f is upper semicontinuous on $\operatorname{dom} f$ *and locally Lipschitz continuous on D.*

(b) At every $y \in D$, the left- and right-hand derivatives

$$f'_-(y) := \lim_{x \uparrow y} \frac{f(x) - f(y)}{x - y}, \qquad f'_+(y) := \lim_{z \downarrow y} \frac{f(z) - f(y)}{z - y}$$

exist in $\mathbb{R}$. The functions f'_- and f'_+ are both increasing on D and satisfy $f'_-(y) \leq f'_+(y)$ for all $y \in D$.

(c) The right-hand derivative f'_+ is right-continuous, and the left-hand derivative f'_- is left-continuous.

(d) f is differentiable almost everywhere in D, and for any $x_0 \in D$,

$$f(x) = f(x_0) + \int_{x_0}^{x} f'_+(y)\, dy = f(x_0) + \int_{x_0}^{x} f'_-(y)\, dy, \qquad x \in D.$$

The existence of one-sided derivatives reflects the fact that the difference quotients of a convex function are monotone in the variable of differentiation, a property that follows directly from the convexity inequality (2.5). This monotonicity will reappear in Chapter 4 as a defining feature of distortion functions, where the concavity of the distortion will translate, via this proposition, into structural properties of the associated risk measure.

2.2 Topological Vector Spaces and Convex Duality Tools

The Banach space framework of Section 2.1 provides a metric topology generated by the norm. For the duality theory developed in Chapter 3, however, this topology is too fine because in infinite-dimensional Banach spaces, norm-bounded sets are generally not compact, which means that dual representations involving suprema over such sets need not attain their values. To recover attainment, one passes to coarser topologies in which compactness becomes available. The natural setting for such constructions is that of locally convex topological vector spaces, which we now introduce.

2.2.1 Locally Convex Spaces and Weak Topologies

We begin with the most general setting in which the interaction between linear structure and topology is well-behaved.

Definition 2.10. A vector space E equipped with a topology τ is called a *topological vector space* if every singleton $\{x\}$ for $x \in E$ is a closed set, and if the vector space operations

$$(x, y) \in E \times E \mapsto x + y \in E \quad \text{and} \quad (\alpha, x) \in \mathbb{R} \times E \mapsto \alpha x \in E$$

are continuous with respect to the product topologies on $E \times E$ and $\mathbb{R} \times E$.

Every Banach space is a topological vector space, and many of the constructions familiar from the normed setting such as continuity of functionals, closure of convex sets, and separation arguments carry over to this more general framework, often in the form most useful for the dual theory of risk measures. The crucial additional property that links topology and convexity is the following.

Definition 2.11. A topological vector space E is called *locally convex* if its topology has a base consisting of convex sets.

Since the open balls $\{y \in E \mid \|y - x\| < r\}$ form a base for the norm topology on E, and since every such ball is convex, any Banach space is in particular locally convex.

Local convexity is precisely the structural assumption needed for the Hahn–Banach separation theorems of Section 2.2.2, and hence for essentially all duality results we will employ. We note in passing that not every topological vector space arising in probability theory is locally convex. Indeed, the space $L^0(\Omega, \mathcal{F}, \mathbb{P})$ of equivalence classes of real-valued random variables, equipped with the topology of convergence in probability, fails to be locally convex on atomless probability spaces. For the spaces L^p with $p \geq 1$ and for the space $\mathcal{X}$ of bounded measurable functions, however, local convexity is automatic.

A particularly natural way to construct locally convex topologies is through families of linear functionals. This construction will allow us to define the weak and weak* topologies that govern the dual representation of risk measures.

Definition 2.12. Let E be a linear space, and let $\mathcal{F}$ be a linear class of linear functionals on E that *separates the points of* E, that is, for any two distinct points $x, y \in E$ there exists some $\ell \in \mathcal{F}$ with $\ell(x) \neq \ell(y)$. The *$\mathcal{F}$-topology* on E, denoted $\sigma(E, \mathcal{F})$, is the topology obtained by taking as a base all sets of the form

$$\{y \in E \mid |\ell_i(y) - \ell_i(x)| < r,\ i = 1, \ldots, n\}, \tag{2.6}$$

where $n \in \mathbb{N}$, $x \in E$, $\ell_1, \ldots, \ell_n \in \mathcal{F}$, and $r > 0$.

The fundamental properties of the $\mathcal{F}$-topology construction are summarized in the subsequent result.

Proposition 2.13. *Let E and $\mathcal{F}$ be as in Definition 2.12. Then:*

(a) E is a locally convex space for the $\mathcal{F}$-topology.

(b) The $\mathcal{F}$-topology is the coarsest topology on E for which every $\ell \in \mathcal{F}$ is continuous.

(c) The topological dual of E for the $\mathcal{F}$-topology is equal to $\mathcal{F}$.

The most important instances of this construction, which will accompany us throughout the thesis, are the weak and weak* topologies.

Definition 2.14. Let E be a locally convex topological vector space with topological dual E^*.

(a) The *weak topology* on E is the topology $\sigma(E, E^*)$, that is, the coarsest topology on E for which every $\ell \in E^*$ is continuous.

(b) If we view each $x \in E$ as a linear functional on E^* via the evaluation $x(\ell) := \ell(x)$, the corresponding topology $\sigma(E^*, E)$ on E^* is called the *weak* topology* on E^*.

By construction, the weak and weak* topologies are coarser than the norm topology. As a consequence, more sets are compact for these topologies than for the norm topology.

By Proposition 2.13 (c), the dual of E equipped with $\sigma(E, E^*)$ is again E^*, and the dual of E^* equipped with the weak* topology is again E.

To illustrate the relevance of these topologies, consider the Banach space $L^\infty = L^\infty(\Omega, \mathcal{F}, \mathbb{P})$. By the duality result (2.4) and its $p = 1$ analogue, L^∞ is the topological dual of L^1, but the converse is generally not true because the dual of L^∞ is strictly larger than L^1 and is instead described by finitely additive set functions, as we will see in Section 2.4. However, if we endow L^∞ with the weak* topology $\sigma(L^\infty, L^1)$, then L^1 becomes its topological dual. This observation will be central in Section 2.3, where the weak* topology on L^∞ provides the natural framework for characterizing convex risk measures with the Fatou property.

Remark 2.15. If E is infinite-dimensional, the weak and weak* topologies are typically not metrizable. Consequently, sequential arguments no longer suffice for topological assertions, and one must in general resort to nets or filters. This subtlety becomes relevant in the analysis of the Fatou property of a convex risk measure on L^∞ and its relation to weak* lower semicontinuity.

2.2.2 Separation and the Closure of Convex Sets

The Hahn–Banach theorem is the cornerstone of convex analysis in infinite dimensions and provides the key tool for the dual representation of convex risk measures. We collect here the versions of the theorem that we will use, beginning with a finite-dimensional prototype and proceeding to the topological vector space setting. Although the theorem admits a purely algebraic formulation in terms of sublinear majorants, the geometric formulation in terms of separating hyperplanes is the one we shall most frequently invoke.

Geometrically, the idea is simple. Whenever two convex subsets of $\mathbb{R}^d$ do not overlap, one can slip a hyperplane between them, realized as a level set of some nonzero linear functional. The finite-dimensional statement below makes this precise.

Proposition 2.16 (Minkowski separation)**.** *[37] Suppose $C \subseteq \mathbb{R}^d$ is a nonempty convex set with $0 \notin C$. Then there exists $\eta \in \mathbb{R}^d$ with $\eta \cdot x \geq 0$ for all $x \in C$, and $\eta \cdot x_0 > 0$ for at least one $x_0 \in C$. If, in addition, $\inf_{x \in C} |x| > 0$, then there exists $\eta \in \mathbb{R}^d$ with $\inf_{x \in C} \eta \cdot x > 0$.*

The genuine Hahn–Banach theorem extends this geometric separation to an infinite-dimensional setting. We state two versions, distinguished by the strength of the separation they yield.

Theorem 2.17. *[5, 30] In a topological vector space E, any two disjoint convex sets B and C, one of which has an interior point, can be separated by a nonzero continuous linear functional ℓ on E, that is,*

$$\ell(x) \leq \ell(y) \qquad \text{for all } x \in C \text{ and } y \in B. \tag{2.7}$$

To strictly separate two convex sets by a linear functional and obtain a strict inequality in 2.7, one uses the following Hahn–Banach Theorem. The strict separation requires a stronger hypothesis on the sets and the ambient space.

Theorem 2.18 (Hahn–Banach)**.** *[5, 30] Suppose that B and C are two nonempty, disjoint, and convex subsets of a locally convex space E. Then, if B is compact and C is closed, there exists a continuous linear functional ℓ on E such that*

$$\sup_{x\in C} \ell(x) < \inf_{y\in B} \ell(y). \tag{2.8}$$

The strict separation result of Theorem 2.18 is frequently used in this thesis. It will be invoked in the proof of the robust representation theorem for convex risk measures in Section 3.5.2, where the role of C will be played by the acceptance set of the risk measure, while B will be a suitably chosen singleton outside this set. The strict inequality (2.8) is what allows one to construct a probabilistic model relative to which the rejected position has strictly negative expected value.

A further consequence of the separation theorem concerns the interplay between closedness in the original topology and in the weak topology. The following result, central to the duality theory developed later, shows that for convex sets the two notions coincide.

Theorem 2.19 (Mazur)**.** *[36] Suppose that E is a locally convex space and that C is a convex subset of E. Then C is weakly closed if and only if C is closed in the original topology of E.*

The equivalence in Theorem 2.19 is more than a technical curiosity. It will be essential in Chapter 3, where it allows the regularity of a convex risk measure, established in the norm topology, to be transferred to the weak topology in which its dual representation is obtained.

2.3 Compactness in the Weak* Topology

The motivation for considering weak topologies in Section 2.2 was the observation that compactness, which is rare in the norm topology of an infinite-dimensional Banach space, becomes more readily available in coarser topologies. In this short section we make this idea precise through three results that will be indispensable in the dual representation of risk measures: the Banach–Alaoglu theorem, the Krein–Šmulian theorem, and a consequence of the latter that characterizes weak* closedness in L^∞. Together, these results will guarantee that the suprema appearing in the robust representations of Chapter 3 are in fact maxima, and that natural candidates for the dual variables, such as the set $\mathcal{M}_{1,f}$ of finitely additive probability measures, enjoy the compactness needed for these maxima to be attained. Of these three results, only the first is a compactness

statement in the strict sense. The Krein–Šmulian theorem and its L^∞ consequence instead provide criteria for weak* closedness; their role is to supply the missing ingredient that, together with Banach–Alaoglu, turns closedness into compactness, since a weak* closed subset of a weak* compact ball is itself weak* compact.

We begin with the most fundamental of these results. Recall that on the dual E^* of a Banach space $(E, \|\cdot\|_E)$, the dual norm is given by (2.3).

Theorem 2.20 (Banach–Alaoglu)**.** *[2] Let E be a Banach space with topological dual E^*. Then $\{\ell \in E^* \mid \|\ell\|_{E^*} \leq r\}$ is weak* compact for every $r > 0$.*

The Banach–Alaoglu theorem is the prototypical compactness statement of functional analysis. Although norm-bounded sets in E^* are typically not compact in the dual norm topology, they become compact as soon as one passes to the weaker $\sigma(E^*, E)$ topology.

The relevance of this theorem for our purposes will be the following. The space $\mathcal{X}$ of bounded measurable functions, equipped with the supremum norm, is a Banach space whose topological dual will be identified in Section 2.4 with the space $ba(\Omega, \mathcal{F})$ of finitely additive set functions of finite total variation. Within ba, the set $\mathcal{M}_{1,f}$ of finitely additive probability measures is norm-bounded and, as shown there, weak* closed; being a closed subset of a weak* compact ball, it is itself weak* compact by Theorem 2.20. This compactness will be used in Chapter 3 to upgrade the supremum in the robust representation

$$\rho(X) = \sup_{Q \in \mathcal{M}_{1,f}} \Big(\mathbb{E}_Q[-X] - \alpha^{\min}(Q) \Big)$$

to a maximum, since the upper semicontinuous functional $Q \mapsto \mathbb{E}_Q[-X] - \alpha^{\min}(Q)$ attains its supremum on a weak* compact set.

For more refined questions, the Banach–Alaoglu theorem alone is not sufficient because it characterizes weak* compactness only of norm-bounded sets, while in applications we often wish to verify weak* closedness of a convex set without assuming any norm bound a priori. The Krein–Šmulian theorem fills exactly this gap.

Theorem 2.21 (Krein–Šmulian)**.** *[34] Let E be a Banach space and suppose that C is a convex subset of the dual space E^*. Then C is weak* closed if and only if*

$$C \cap \{\ell \in E^* \mid \|\ell\|_{E^*} \leq r\}$$

is weak closed for each $r > 0$.*

For our purposes, the main payoff of the Krein–Šmulian theorem is the following criterion for weak* closedness in $L^\infty = L^\infty(\Omega, \mathcal{F}, \mathbb{P})$. Instead of testing closedness against the unwieldy topology $\sigma(L^\infty, L^1)$ directly, it suffices to verify ordinary L^1-closedness on norm-bounded balls.

Lemma 2.22. *A convex subset C of L^∞ is weak* closed (with respect to the topology $\sigma(L^\infty, L^1)$) if for every $r > 0$,*

$$C_r := C \cap \{X \in L^\infty \mid \|X\|_\infty \leq r\}$$

is closed in L^1.

Lemma 2.22 provides the link between the norm topology of L^∞, in which our risk measures will be shown to be Lipschitz continuous, and the weak* topology, in which their dual representations live. The practical payoff of all these compactness results is the same, and we isolate it in the next remark.

Remark 2.23. Compactness becomes useful in combination with semicontinuity. We record the elementary principle underlying the attainment results of later chapters. If K is a compact topological space and $f : K \to \mathbb{R} \cup \{-\infty\}$ is upper semicontinuous, then f attains its supremum on K. In Chapter 3 this will be applied to the weak* compact set $\mathcal{M}_{1,f}$ to show that the suprema in the robust representation of convex risk measures are in fact attained.

The three results will not be used equally often. Banach–Alaoglu is the one we invoke most frequently, whenever a supremum must be upgraded to a maximum or a weak* convergent subsequence extracted from a bounded family of dual variables. The Krein–Šmulian theorem and its Lemma 2.22 enter more specifically in the study of risk measures on L^∞, linking the Fatou property to the weak* closedness of acceptance sets.

With these compactness tools in hand, we now turn to the concrete identification of the dual space $\mathcal{X}^*$ that appears implicitly throughout the discussion above.

2.4 Finitely Additive Set Functions and the Dual of $\mathcal{X}$

In Section 2.1 we introduced the Banach space $\mathcal{X}$ of bounded measurable functions on a measurable space $(\Omega, \mathcal{F})$, equipped with the supremum norm. The dual representation of convex risk measures on $\mathcal{X}$ requires a concrete description of its topological dual $\mathcal{X}^*$. As already announced, this dual is not, in general, the space $\mathcal{M}_1$ of σ-additive probability measures on $(\Omega, \mathcal{F})$, but rather a strictly larger space of *finitely additive* set functions. In this section we introduce this space, develop the corresponding integration theory in the form needed for our purposes, and state the Riesz-type representation theorem that identifies $\mathcal{X}^*$ explicitly. The treatment is deliberately concise; for further details we refer to Hans Föllmer and Alexander Schied [26].

Definition 2.24. Let $(\Omega, \mathcal{F})$ be a measurable space. A mapping $\mu : \mathcal{F} \to \mathbb{R}$ is called a *finitely additive set function* if $\mu(\emptyset) = 0$, and if for any finite collection $A_1, \ldots, A_n \in \mathcal{F}$

of mutually disjoint sets,

$$\mu\Big(\bigcup_{i=1}^{n} A_i\Big) = \sum_{i=1}^{n} \mu(A_i).$$

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

$$\|\mu\|_{\mathrm{var}} := \sup\Big\{\sum_{i=1}^{n} |\mu(A_i)| \;\Big|\; A_1, \dots, A_n \text{ disjoint sets in } \mathcal{F},\, n \in \mathbb{N}\Big\}.$$

The space of all finitely additive measures μ whose total variation is finite is denoted by $ba(\Omega, \mathcal{F})$, or simply ba when no ambiguity arises. We furthermore denote by

$$\mathcal{M}_{1,f} := \mathcal{M}_{1,f}(\Omega, \mathcal{F})$$

the set of all finitely additive set functions $\mu : \mathcal{F} \to [0,1]$ that are normalized by $\mu(\Omega) = 1$.

The space ba is a vector space, and the total variation $\|\cdot\|_{\mathrm{var}}$ defines a norm on it under which ba becomes a Banach space. The set $\mathcal{M}_{1,f}$ is contained in ba and consists, intuitively, of those "probability-like" assignments that drop the requirement of countable additivity. The crucial observation is that $\mathcal{M}_{1,f}$ is strictly larger than the set $\mathcal{M}_1 := \mathcal{M}_1(\Omega, \mathcal{F})$ of all (σ-additive) probability measures on $(\Omega, \mathcal{F})$ whenever $(\Omega, \mathcal{F})$ is sufficiently rich. In particular, if $\mathcal{F}$ contains infinitely many disjoint nonempty sets, then there exist finitely additive probability measures that are not countably additive. These are precisely the objects that make the dual of $\mathcal{X}$ strictly larger than $\mathcal{M}_1$.

We now develop the corresponding integration theory, which is needed to make sense of expressions of the form $\mathbb{E}_Q[F]$ for $Q \in \mathcal{M}_{1,f}$. The construction proceeds, as in the classical case, by first defining the integral on simple functions and then extending by continuity.

Let $\mathcal{X}_0 \subseteq \mathcal{X}$ denote the linear subspace of all finitely valued step functions of the form

$$F = \sum_{i=1}^{n} \alpha_i \mathbf{1}_{A_i}$$

for some $n \in \mathbb{N}$, $\alpha_i \in \mathbb{R}$, and sets $A_1, \dots, A_n \in \mathcal{F}$. For $\mu \in ba$, we define

$$\int F \, d\mu := \sum_{i=1}^{n} \alpha_i \mu(A_i). \tag{2.9}$$

A direct computation shows that this definition is independent of the chosen representation of F, and that the inequality

$$\left|\int F \, d\mu\right| \le \|F\| \cdot \|\mu\|_{\mathrm{var}} \tag{2.10}$$

holds for all $F \in \mathcal{X}_0$. Since $\mathcal{X}_0$ is dense in $\mathcal{X}$ with respect to the supremum norm, the inequality (2.10) allows us to extend the integral as a continuous linear functional to

the entire space $\mathcal{X}$. For $Q \in \mathcal{M}_{1,f}$ we adopt the familiar notation

$$\mathbb{E}_Q[F] := \int F \, dQ, \qquad F \in \mathcal{X}.$$

This integral coincides with the usual expectation when Q happens to be σ-additive, but is well-defined for arbitrary $Q \in \mathcal{M}_{1,f}$.

We now arrive at the central result of this section, which establishes the announced identification of $\mathcal{X}^*$ with ba.

Theorem 2.25 (Riesz-type representation)**.** *The integral*

$$\ell(F) := \int F \, d\mu, \qquad F \in \mathcal{X},$$

defines a one-to-one correspondence between continuous linear functionals ℓ on $\mathcal{X}$ and finitely additive set functions $\mu \in ba$.

This identification goes back to Hildebrandt [31] and to Fichtenholz and Kantorovich [24]; we follow the presentation of Föllmer and Schied [26].

The construction of the inverse correspondence is straightforward. Given a continuous linear functional ℓ on $\mathcal{X}$, one defines a finitely additive set function μ on $(\Omega, \mathcal{F})$ by setting $\mu(A) := \ell(\mathbf{1}_A)$ for $A \in \mathcal{F}$. The continuity of ℓ ensures that μ has finite total variation, and density of simple functions in $\mathcal{X}$ guarantees that the integral against μ recovers ℓ.

Theorem 2.25 has the following important specialization, which identifies the elements of $\mathcal{M}_{1,f}$ with a distinguished subclass of continuous linear functionals.

Remark 2.26. Theorem 2.25 yields, in particular, a one-to-one correspondence between set functions $Q \in \mathcal{M}_{1,f}$ and continuous linear functionals ℓ on $\mathcal{X}$ satisfying

$$\ell(\mathbf{1}) = 1 \quad \text{and} \quad \ell(F) \geq 0 \text{ whenever } F \geq 0.$$

This characterization will be used in Chapter 3 to identify, in the dual representation of a convex risk measure, those continuous linear functionals on $\mathcal{X}$ that arise from finitely additive probability measures. These are precisely the positive functionals normalized to 1 on the constant function. The monotonicity and cash invariance of a monetary risk measure will translate, via Theorem 2.25, into exactly these two conditions on the relevant dual variables.

We close this section with the consequence for compactness that we anticipated in Section 2.3.

Corollary 2.27. *The set $\mathcal{M}_{1,f}$ is weak* compact in $\mathcal{X}^* = ba$.*

This corollary completes the picture promised at the end of Section 2.3. The natural set of dual variables for risk measures on $\mathcal{X}$ is weak* compact, so that any upper

semicontinuous functional on $\mathcal{M}_{1,f}$ attains its supremum. With the dual space $\mathcal{X}^*$ now concretely identified and its compactness properties established, we are ready to assemble the abstract duality result that ties everything together: the Fenchel–Legendre transform and the Fenchel–Moreau theorem.

2.5 Convex Duality and the Fenchel–Legendre Transform

The mathematical heart of convex analysis is the Fenchel–Legendre transform, which assigns to a function on a vector space a dual function on the dual space. This transform encodes the geometric duality between points and supporting hyperplanes that we have encountered, in various guises, throughout the previous sections, and it is the precise mechanism by which the abstract duality theory is brought to bear on concrete problems. In this section, we develop the transform and its central inversion property, the Fenchel–Moreau theorem, and conclude by showing how the tools assembled in the present chapter combine into a coherent strategy for the robust representation of convex risk measures developed in Chapter 3.

2.5.1 The Fenchel–Legendre Transform

We begin with the one-dimensional case, where the geometric content of the construction is most transparent.

Definition 2.28. The *Fenchel–Legendre transform* of a function $f : \mathbb{R} \to \mathbb{R} \cup \{+\infty\}$ is the function $f^* : \mathbb{R} \to \mathbb{R} \cup \{+\infty\}$ defined by

$$f^*(y) := \sup_{x \in \mathbb{R}} \Big(yx - f(x) \Big), \qquad y \in \mathbb{R}. \tag{2.11}$$

If f is itself a proper convex function, then f^* is also called the *conjugate function* of f.

Before describing the geometric meaning of the transform, we recall the notion that underlies it. A *supporting line* of f at a point $x \in \mathbb{R}$ is an affine function that touches the graph of f at $(x, f(x))$ and lies below it everywhere else; that is, a line of some slope α such that

$$f(y) \geq f(x) + \alpha(y - x) \qquad \text{for all } y \in \mathbb{R}. \tag{2.12}$$

The slope α is the slope of the supporting line, and the existence of such a line at x is precisely the local expression of convexity since f admits a supporting line at every point if and only if it is convex [49].

With this notion at hand, the geometric interpretation of $f^*(y)$ is the following: for each slope y, one considers all affine functions of slope y that lie beneath the graph of f, and $-f^*(y)$ is the largest possible intercept of such a minorant [49]. In this sense,

the Fenchel–Legendre transform encodes f through its supporting lines. Rather than recording the values of f pointwise, f^* records, for each slope, the best affine lower bound of f.

If $f \not\equiv +\infty$, then f^* is automatically convex and lower semicontinuous, since it is defined as the pointwise supremum of the affine functions $y \mapsto yx - f(x)$. This stability property, that the conjugate is always convex and lsc, irrespective of the original function, is what makes the Fenchel–Legendre transform an operation into the class of convex, lower semicontinuous functions.

The fundamental relation between a function and its conjugate is captured by the Fenchel–Young inequality and its associated duality.

Proposition 2.29. *[23, 40] Let f be a proper convex function on $\mathbb{R}$.*

(a) For all $x, y \in \mathbb{R}$,

$$xy \le f(x) + f^*(y), \tag{2.13}$$

with equality if x belongs to the interior of $\operatorname{dom} f$ *and if $y \in [f'_-(x), f'_+(x)]$.*

(b) If f is lower semicontinuous, then $f^{**} = f$, that is,

$$f(x) = \sup_{y \in \mathbb{R}} \Big(xy - f^*(y) \Big), \qquad x \in \mathbb{R}.$$

The Fenchel–Young inequality (2.13) expresses the duality between a convex function and its conjugate, and the case of equality identifies, for each point, the slopes of the supporting lines of f there. Equality holds exactly when y is an admissible slope at x, the range of which is delimited by the one-sided derivatives $f'_-(x)$ and $f'_+(x)$ from Proposition 2.9.

Part (b) of Proposition 2.29 is the *Fenchel–Moreau theorem in dimension one*: a proper convex function that is lower semicontinuous is recovered from its conjugate by applying the transform a second time. The underlying mechanism is a separation argument in $\mathbb{R}^2$, carried out with the help of Proposition 2.16, and it is the prototype of the more general Fenchel–Moreau theorem we are about to state [40].

We now move to the infinite-dimensional setting. The structural ingredients are the same as in dimension one, a proper convex function, lower semicontinuity, and a separation argument, but the linear functionals that played the role of slopes in $\mathbb{R}$ now live in the dual space E^*.

Definition 2.30. Let E be a locally convex space with dual E^*. The *Fenchel–Legendre transform* of a function $f : E \to \mathbb{R} \cup \{+\infty\}$ is the function $f^* : E^* \to \mathbb{R} \cup \{+\infty\}$ defined by

$$f^*(\ell) := \sup_{x \in E} \Big(\ell(x) - f(x) \Big), \qquad \ell \in E^*. \tag{2.14}$$

As in the one-dimensional case, if $f \not\equiv +\infty$, then f^* is a proper convex function, and it is lower semicontinuous with respect to the weak* topology on E^*, since it is the supremum of weak* continuous affine functions.

The central duality result is the following.

Theorem 2.31 (Fenchel–Moreau)**.** *[13, 38] Let E be a locally convex topological vector space, and let $f : E \to \mathbb{R} \cup \{+\infty\}$ be a proper convex function. If f is lower semicontinuous with respect to the weak topology $\sigma(E, E^*)$, then $f = f^{**}$, that is,*

$$f(x) = \sup_{\ell \in E^*} \Big(\ell(x) - f^*(\ell) \Big), \qquad x \in E. \tag{2.15}$$

It is worth pausing on the role of lower semicontinuity in this theorem. The following observation, which we have implicitly used several times already, makes its function transparent.

Remark 2.32. The Fenchel–Moreau theorem fails without lower semicontinuity. Indeed, the biconjugate satisfies $f^{**} \leq f$ in general, a direct consequence of the Fenchel–Young inequality, while f^{**} is always convex and lower semicontinuous. Equality $f = f^{**}$ therefore forces f itself to be convex and lsc. In fact, f^{**} is the largest convex lsc minorant of f, the so-called *convex envelope* of f. From this point of view, the Fenchel–Moreau theorem says that lower semicontinuity is precisely the regularity required for a proper convex function to be recoverable from its dual data [49].

The duality between f and f^* thus achieves something remarkable: it converts a convex object on E into a linear-style object on E^*, and back, without loss of information. The nonlinear function f is encoded in the family of its supporting affine functionals, and these supporting functionals are precisely the points $(\ell, f^*(\ell))$ at which the affine function $x \mapsto \ell(x) - f^*(\ell)$ supports f from below. This encoding is the engine of the dual representation theorems for risk measures, to which we now turn.

2.5.2 Synthesis: A Strategy for the Robust Representation

The results of this chapter were not assembled for their own sake. Each was chosen because it supplies a specific link in the derivation of the robust representation of convex risk measures, which is the subject of Chapter 3. Before turning to that theory, we pause to show how the individual pieces fit together into a single chain of reasoning. The goal of Chapter 3 will be to represent a convex risk measure ρ on $\mathcal{X}$ in the form

$$\rho(X) = \max_{Q \in \mathcal{M}_{1,f}} \Big(\mathbb{E}_Q[-X] - \alpha^{\min}(Q) \Big), \tag{2.16}$$

where $\alpha^{\min}$ is a penalty function on $\mathcal{M}_{1,f}$. The diagram below traces the path from ρ to this representation in five steps, indicating for each step the section of the present chapter that supplies the required tool.

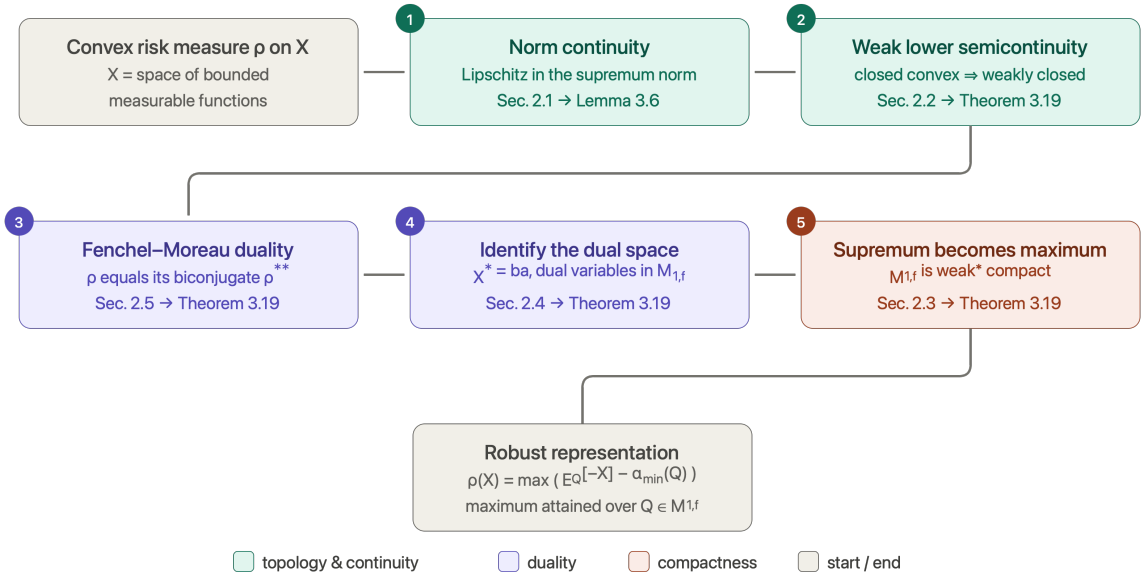

Figure 2.1: The five-step strategy behind the robust representation (2.16). Each step names the functional-analytic tool, the section of this chapter that supplies it, and the result in Chapter 3 where it enters the proof.

This chain is the abstract skeleton of the robust representation theorem. Each link rests on a result of the present chapter, and the fact that they combine so cleanly is what makes the duality theory of convex risk measures work. In the next chapter we leave the abstract setting behind and introduce risk measures formally. We will define monetary, convex, and coherent risk measures, establish the properties used informally above, and carry out the representation just outlined in full detail.

Chapter 3

Risk Measures

In the previous chapter, we introduced the probabilistic and functional analytic framework that will serve as the mathematical foundation throughout this thesis. Building on this setup, we now turn to the central object of interest in modern risk theory: risk measures. A risk measure is a functional that assigns to a financial position, represented by a random variable, a real number quantifying the associated risk. Such functionals play a fundamental role in financial mathematics, as they provide a rigorous way to determine capital requirements and to compare the riskiness of different positions.

This chapter develops the axiomatic foundations of risk measures. We begin by introducing basic definitions and structural properties, which reflect economically meaningful principles. Particular emphasis is placed on convex and coherent risk measures, which form the cornerstone of the theory.

A key aspect of this framework is the existence of dual representation results, linking risk measures to families of probability measures. These representations reveal the deep connection between risk assessment and model uncertainty, and they will play a crucial role in the subsequent analysis.

Finally, we introduce law-invariant risk measures, which depend only on the distribution of the underlying position. This property provides the bridge to the main topics of this thesis, namely distortion risk measures and comonotonic risk measures, which will be studied in the following chapters.

3.1 Foundations

This section introduces the basic notion of a risk measure and the fundamental properties that such functionals are typically required to satisfy. These attributes formalize key financial principles, such as monotonicity, cash invariance, and diversification.

3.1.1 Definition of Monetary Risk Measures

Let Ω be a fixed set of scenarios. A *financial position* is represented by a function

$$\begin{aligned} X\colon \Omega &\to \mathbb{R} \\ \omega &\mapsto X(\omega), \end{aligned}$$

where $X(\omega)$ denotes the discounted net value of the position at the end of the trading period in the scenario $\omega \in \Omega$.

The objective is to assess the risk associated with X by assigning a real number $\rho(X)$, where X is taken from a prescribed class $\mathcal{X}$ of financial positions. Throughout this section, we assume that $\mathcal{X}$ is a linear space of bounded functions that contains all constant functions. We do not specify any reference probability measure on Ω.

Definition 3.1. A mapping $\rho\colon \mathcal{X} \to \mathbb{R}$ is called a *monetary risk measure* 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 economic intuition behind monotonicity is that a position with uniformly higher payoffs is less exposed to adverse outcomes. Indeed, if $X \leq Y$, then in every scenario the payoff of Y is at least as large as the one of X, so that potential losses are reduced. As a consequence, the capital required to make the position acceptable is lower, which explains why $\rho(X) \geq \rho(Y)$.

The property of cash invariance, also referred to as translation invariance, is closely related to the interpretation of a risk measure as a capital requirement. More precisely, $\rho(X)$ can be understood as the amount of capital which needs to be added to a position X in order to render it acceptable from a regulatory perspective. Adding a deterministic amount m to the position and holding it in a risk-free asset therefore lowers the capital requirement by precisely that amount. In particular, if you set $m = \rho(X)$, cash invariance implies

$$\rho\Big(X + \rho(X)\Big) = 0. \tag{3.1}$$

This means adding the amount $\rho(X)$ to the position X renders it acceptable.

If we instead set $X = 0$, cash invariance implies

$$\rho(m) = \rho(0) - m \quad \text{for all } m \in \mathbb{R}. \tag{3.2}$$

This shows that the risk of a certain payment m depends linearly on its size.

Remark 3.2. [9] Many commonly used risk measures do not satisfy the axiom of cash invariance. A classical example is the variance. Indeed, for any constant $m \in \mathbb{R}$, one

has $\mathrm{Var}(X+m) = \mathrm{Var}(X)$, which shows that adding a risk-free amount of cash does not reduce the measured risk.

From a financial perspective, this is undesirable, since adding a sure amount of money should decrease the risk of a position in a transparent and linear way. This highlights that variance captures dispersion rather than risk in the sense of capital requirements and thus motivates the introduction of alternative risk measures satisfying more suitable axioms.

In many contexts, the assumption that a monetary risk measure is normalized can be made without loss of generality.

Definition 3.3. A monetary risk measure $\rho : \mathcal{X} \to \mathbb{R}$ is called *normalized* if it satisfies

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

Lemma 3.4. *Let ρ be a normalized monetary risk measure. Then cash invariance is equivalent to cash additivity,*

$$\rho(X+m) = \rho(X) + \rho(m) \quad \textit{for all } X \in \mathcal{X},\ m \in \mathbb{R}.$$

Proof. If $\rho(0) = 0$, then by cash invariance we have $\rho(m) = \rho(0) - m = -m$. Moreover, cash invariance yields $\rho(X+m) = \rho(X) - m$, and combining both relations gives

$$\rho(X+m) = \rho(X) + \rho(m).$$

The converse implication follows immediately. □

However, in some situations, it may be convenient not to impose normalization.

Remark 3.5. Throughout this thesis, we adopt the convention that X represents the *discounted* value of a financial position. For instance, the discount factor may be chosen as $\frac{1}{1+r}$, where r denotes the return of a risk-free investment.

Instead of working with discounted positions, one may equivalently consider the corresponding nominal value

$$\tilde{X} = (1+r)X.$$

Defining a new risk measure by $\tilde{\rho}(\tilde{X}) := \rho(X)$, monotonicity is preserved, while cash invariance takes the form

$$\tilde{\rho}(\tilde{X} + (1+r)m) = \tilde{\rho}(\tilde{X}) - m.$$

In particular, the risk is reduced by m when an additional amount m is invested in a risk-free asset. Conversely, any mapping $\tilde{\rho}$ satisfying this property and monotonicity induces a monetary risk measure via

$$\rho(X) := \tilde{\rho}((1+r)X).$$

The following lemma establishes a fundamental continuity property of monetary risk measures, which will play a key role in several results developed in the subsequent chapters.

Lemma 3.6. *Any monetary risk measure ρ is Lipschitz continuous with respect to the supremum norm $\|\cdot\|$:*

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

Proof. Let $X, Y \in \mathcal{X}$. By definition of the supremum norm, we have $X \leq Y + \|X - Y\|$. By monotonicity and cash invariance, it follows that

$$\rho(X) \geq \rho\Big(Y + \|X - Y\|\Big) = \rho(Y) - \|X - Y\|.$$

Interchanging the roles of X and Y yields $\rho(Y) \geq \rho(X) - \|X - Y\|$. Combining both inequalities gives the result. □

3.1.2 Convex and Coherent Risk Measures

From now on, we focus on monetary risk measures that satisfy an additional convexity property.

Definition 3.7. [25, 27] A monetary risk measure $\rho : \mathcal{X} \to \mathbb{R}$ is called a *convex risk measure* if it satisfies

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

Consider the set of possible future outcomes that can be generated using the resources available to an investor. One investment strategy results in a position X, while another leads to a position Y. If the investor diversifies by allocating a fraction λ of the available resources to the first strategy and the remaining proportion to the second, the resulting position is given by $\lambda X + (1 - \lambda)Y$. The convexity property then formalizes the fundamental principle that diversification should not increase risk. Especially the risk of the combined position should not exceed the weighted average of the risks of the individual positions.

This idea can be made precise by observing, for monetary risk measures, the equivalence between convexity and quasi-convexity, i.e.,

$$\rho(\lambda X + (1 - \lambda)Y) \leq \max\{\rho(X), \rho(Y)\}, \quad \text{for all } \lambda \in [0, 1].$$

A formal proof of this fact is provided in Appendix Section B.1.

A particularly important subclass of convex risk measures is obtained by imposing an additional scaling property, leading to the notion of coherent risk measures.

Definition 3.8. [3] A convex risk measure ρ is called a *coherent risk measure* if it satisfies

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

If a monetary risk measure ρ is positively homogeneous, then it is normalized in the sense that $\rho(0) = 0$. To see this, note that positive homogeneity implies

$$\rho(0) = \rho(\lambda \cdot 0) = \lambda \rho(0) \quad \text{for all } \lambda \geq 0,$$

which is only possible if $\rho(0) = 0$. Moreover, under positive homogeneity, convexity is equivalent to *subadditivity*, that is,

$$\rho(X + Y) \leq \rho(X) + \rho(Y), \qquad X, Y \in \mathcal{X}.$$

This property reflects the beneficial effect of diversification.

Subadditivity makes it possible to distribute the task of risk management across different units. If individual desks are assigned separate risk limits, the risk of the overall portfolio can be controlled by the sum of these limits, ensuring that the aggregated exposure remains bounded.

In practice, however, the relationship between position size and risk is often nonlinear, which calls into question the assumption of positive homogeneity. As positions become larger, various effects may lead to a disproportionate increase in risk. A prominent example is market liquidity risk: while small positions can typically be liquidated without affecting market prices, unwinding large positions may significantly move prices, resulting in losses that grow more than proportionally with the position size.

In addition, systemic risk factors may create strong dependencies between positions, so that adverse market movements affect multiple components simultaneously and amplify overall losses. For instance, during a financial crisis, assets that appear to be weakly correlated under normal market conditions may experience simultaneous and substantial losses. As a result, the total risk of a portfolio can increase more than linearly with its size, since diversification effects break down in stressed scenarios.

These considerations illustrate that, while subadditivity remains a desirable property reflecting diversification, the assumption of positive homogeneity may be too restrictive in realistic settings. For this reason, we do not impose positive homogeneity in general. Instead, we will primarily focus on convex risk measures, which allow for such nonlinear effects and provide a more flexible framework for modeling risk.

The previous considerations suggest that convexity, positive homogeneity, and subadditivity are closely connected. The following lemma makes this relationship precise.

Lemma 3.9. *Let ρ be a normalized monetary risk measure on $\mathcal{X}$. Then any two of the following properties imply the remaining third one:*

- *Convexity,*

- *Positive homogeneity,*
- *Subadditivity.*

Proof. We show that any two of the three properties imply the remaining third.

Step 1: Convexity and positive homogeneity imply subadditivity. Let $X, Y \in \mathcal{X}$. By convexity,

$$\rho\Big(\frac{X+Y}{2}\Big) \leq \frac{1}{2}\rho(X) + \frac{1}{2}\rho(Y).$$

By positive homogeneity,

$$\rho\Big(\frac{X+Y}{2}\Big) = \frac{1}{2}\rho(X+Y).$$

Hence,

$$\frac{1}{2}\rho(X+Y) \leq \frac{1}{2}\rho(X) + \frac{1}{2}\rho(Y),$$

and therefore multiplying by 2 results in

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

Thus, ρ is subadditive.

Step 2: Positive homogeneity and subadditivity imply convexity. Let $X, Y \in \mathcal{X}$ and $\lambda \in [0,1]$. By subadditivity,

$$\rho(\lambda X + (1-\lambda)Y) \leq \rho(\lambda X) + \rho((1-\lambda)Y).$$

Using positive homogeneity on both terms on the right-hand side, this becomes

$$\rho(\lambda X + (1-\lambda)Y) \leq \lambda\rho(X) + (1-\lambda)\rho(Y).$$

Thus, ρ is convex.

Step 3: Convexity and subadditivity imply positive homogeneity. Let $X \in \mathcal{X}$ and $\lambda \in [0,1]$. Since ρ is convex and normalized, we have

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

On the other hand, by subadditivity,

$$\rho(X) = \rho(\lambda X + (1-\lambda)X) \leq \rho(\lambda X) + \rho((1-\lambda)X).$$

Applying the first inequality to $(1-\lambda)X$, we obtain

$$\rho((1-\lambda)X) \leq (1-\lambda)\rho(X).$$

Consequently,

$$\rho(X) \leq \rho(\lambda X) + (1-\lambda)\rho(X),$$

which implies

$$\lambda\rho(X) \leq \rho(\lambda X).$$

Therefore,

$$\rho(\lambda X) = \lambda\rho(X) \qquad \text{for all } \lambda \in [0,1].$$

Now let $\lambda \geq 1$ and set $\mu = \frac{1}{\lambda} \in (0,1]$. Applying the identity just proved to the position λX, we get

$$\rho(X) = \rho(\mu(\lambda X)) = \mu\,\rho(\lambda X) = \frac{1}{\lambda}\rho(\lambda X).$$

Thus,

$$\lambda\rho(X) = \rho(\lambda X) \qquad \text{for all } \lambda \geq 1.$$

We conclude that

$$\rho(\lambda X) = \lambda\rho(X) \qquad \text{for all } \lambda \geq 0,$$

so that ρ is positively homogeneous.

Since each pair of properties implies the third, the proof is complete. □

3.1.3 Characterization via Acceptance Sets

In this section, we introduce acceptance sets as an alternative way to describe monetary risk measures. Instead of assigning a numerical value to each position, acceptance sets characterize which positions are deemed acceptable.

Definition 3.10. A monetary risk measure ρ induces the class

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

of positions which are acceptable in the sense that they do not require further capital. The class $\mathcal{A}_\rho$ is called the *acceptance set* of ρ.

The following results show how acceptance sets and monetary risk measures are related.

Proposition 3.11. *Let ρ be a monetary risk measure with acceptance set $\mathcal{A} := \mathcal{A}_\rho$.*

(a) $\mathcal{A}$ is nonempty, closed in $\mathcal{X}$ with respect to the supremum norm $\|\cdot\|$, and satisfies:

$$\inf\{m \in \mathbb{R} \mid m \in \mathcal{A}\} > -\infty \tag{3.3}$$

$$X \in \mathcal{A},\ Y \in \mathcal{X},\ Y \geq X \ \Rightarrow\ Y \in \mathcal{A} \tag{3.4}$$

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

$$\rho(X) = \inf\{m \in \mathbb{R} \mid m + X \in \mathcal{A}\} \tag{3.5}$$

(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.*

REMARK 3.12. [14] An acceptance set $\mathcal{A}$ is called a cone if it is stable under positive scaling, i.e., if $X \in \mathcal{A}$ implies $\lambda X \in \mathcal{A}$ for all $\lambda \geq 0$.

In the context of risk measures, this property corresponds to positive homogeneity. It means that if a financial position is acceptable, then any positive multiple of this position remains acceptable. Thus, acceptability is preserved under scaling.

Proof. (a) Properties (3.3) and (3.4) follow directly from the definition of a monetary risk measure. Closedness of $\mathcal{A}$ follows from Lemma 3.6.

(b) By definition of the acceptance set, we have

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

Using cash invariance, this is equal to

$$\inf\{m \in \mathbb{R} \mid \rho(X) - m \leq 0\} = \inf\{m \in \mathbb{R} \mid \rho(X) \leq m\}.$$

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

(c) $\mathcal{A}$ is clearly convex if ρ is a convex risk measure. The converse will follow from Proposition 3.13 together with (3.7).

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

Conversely, one may take a given class $\mathcal{A} \subset \mathcal{X}$ of acceptable positions as the primitive object. For any position $X \in \mathcal{X}$, the associated capital requirement can then be defined as the minimal amount m such that the shifted position $X + m$ becomes acceptable:

$$\rho_{\mathcal{A}}(X) := \inf\{m \in \mathbb{R} \mid X + m \in \mathcal{A}\}. \tag{3.6}$$

With this notation, (3.5) can be written as

$$\rho_{\mathcal{A}_\rho} = \rho. \tag{3.7}$$

Proposition 3.13. *Assume that $\mathcal{A}$ is a nonempty subset of $\mathcal{X}$ that satisfies (3.3) and (3.4). Then the functional $\rho_{\mathcal{A}}$ has the following properties:*

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

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

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

(d) $\mathcal{A}_{\rho_{\mathcal{A}}}$ is equal to the closure of $\mathcal{A}$ with respect to the supremum norm $\|\cdot\|$. In particular, $\mathcal{A}_{\rho_{\mathcal{A}}} = \mathcal{A}$ 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 nonempty set $\mathcal{A}$. For $X \in \mathcal{X}$ given, there exists a finite number m with $m + X > Y$, because 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 therefore $\rho_{\mathcal{A}}(X) \leq m < \infty$. Note that (3.3) 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 $m_1, m_2 \in \mathbb{R}$ are such that $m_1 + X_1 \in \mathcal{A}$ and $m_2 + X_2 \in \mathcal{A}$. 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 $\rho_{\mathcal{A}}$,

$$0 \geq \rho_{\mathcal{A}}\Big(\lambda(m_1 + X_1) + (1 - \lambda)(m_2 + X_2)\Big) = \rho_{\mathcal{A}}\Big(\lambda X_1 + (1 - \lambda)X_2\Big) - \Big(\lambda m_1 + (1 - \lambda)m_2\Big).$$

It follows that $\rho_{\mathcal{A}}\Big(\lambda X_1 + (1 - \lambda)X_2\Big) \leq \lambda m_1 + (1 - \lambda)m_2$. Taking $m_i \downarrow \rho_{\mathcal{A}}(X_i)$ yields the convexity of $\rho_{\mathcal{A}}$.

(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 $\lambda > 0$. Thus, $\lambda m < \rho_{\mathcal{A}}(\lambda X)$, and (c) follows.

(d) Let $\overline{\mathcal{A}}$ denote the closure of $\mathcal{A}$ with respect to $\|\cdot\|$. We first show that $\overline{\mathcal{A}} \subseteq \mathcal{A}_{\rho_{\mathcal{A}}}$. If $X \in \overline{\mathcal{A}}$, there exists a sequence (X_n) in $\mathcal{A}$ such that $\|X_n - X\|_\infty \to 0$. It follows from Lemma 3.6 that $\rho_{\mathcal{A}}(X_n) \to \rho_{\mathcal{A}}(X)$, and so $\rho_{\mathcal{A}}(X) \leq 0$. Hence, $X \in \mathcal{A}_{\rho_{\mathcal{A}}}$ and in turn $\overline{\mathcal{A}} \in \mathcal{A}_{\rho_{\mathcal{A}}}$.

For the proof of the converse inclusion, take $X \in \mathcal{A}_{\rho_{\mathcal{A}}}$. Then

$$0 \geq \rho_{\mathcal{A}}(X) = \inf\{m \in \mathbb{R} \mid m + X \in \mathcal{A}\},$$

and so there exists a sequence (m_n) such that $m_n \downarrow \rho_{\mathcal{A}}(X)$ and $m_n + X \in \mathcal{A}$. Moreover, $X_n := m_n + X - \rho_{\mathcal{A}}(X) \geq m_n + X \in \mathcal{A}$, and so $X_n \in \mathcal{A}$ by (3.4). Since $\|X_n - X\| = m_n - \rho_{\mathcal{A}}(X) \downarrow 0$, we obtain $X \in \overline{\mathcal{A}}$ and therefore $\mathcal{A}_{\rho_{\mathcal{A}}} \subseteq \overline{\mathcal{A}}$. □

The previous propositions establish a close connection between monetary risk measures and acceptance sets. Proposition 3.11 shows that every risk measure naturally induces an acceptance set, while Proposition 3.13 demonstrates that, conversely, a suit-

able acceptance set uniquely determines a monetary risk measure. This correspondence allows us to study risk measures through the geometry of their acceptance sets.

Lemma 3.14. *Let I be a nonempty index set. For each $i \in I$, let ρ_i be a monetary risk measure on $\mathcal{X}$ with associated acceptance set $\mathcal{A}_i$, and assume that $\sup_{i\in I} \rho_i(0) < \infty$. Define*

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

Then ρ is a monetary risk measure on $\mathcal{X}$ with acceptance set

$$\mathcal{A} = \bigcap_{i\in I} \mathcal{A}_i,$$

and ρ is a convex risk measure if each ρ_i is a convex risk measure.

Proof. We first show that ρ is finite-valued. For $X \in \mathcal{X} \subseteq L^\infty$, we have $X \geq -\|X\|$. By monotonicity of each ρ_i, $\rho_i(X) \leq \rho_i(-\|X\|)$. By cash invariance, $\rho_i(-\|X\|) = \rho_i(0) + \|X\|$, and hence

$$\rho_i(X) \leq \rho_i(0) + \|X\| \quad \text{for all } i \in I.$$

Taking the supremum over i yields

$$\rho(X) \leq \sup_{i\in I} \rho_i(0) + \|X\| < \infty.$$

Next, ρ is monotone since each ρ_i is monotone. Moreover, for $m \in \mathbb{R}$,

$$\rho(X+m) = \sup_{i\in I} \rho_i(X+m) = \sup_{i\in I} \Big(\rho_i(X) - m\Big) = \rho(X) - m,$$

so ρ is cash-invariant. Thus, ρ is a monetary risk measure. Let $\mathcal{A}_\rho$ denote the acceptance set of ρ. Then

$$X \in \mathcal{A}_\rho \iff \rho(X) \leq 0 \iff \sup_{i\in I} \rho_i(X) \leq 0 \iff \rho_i(X) \leq 0, \text{ for all } i \in I,$$

and hence $\mathcal{A}_\rho = \bigcap_{i\in I} \mathcal{A}_i$. Finally, if each ρ_i is convex, then for $X, Y \in \mathcal{X}$ and $\lambda \in [0,1]$,

$$\rho_i(\lambda X + (1-\lambda)Y) \leq \lambda\rho_i(X) + (1-\lambda)\rho_i(Y), \qquad \text{for all } i \in I.$$

Consequently,

$$\begin{aligned}\rho(\lambda X + (1-\lambda)Y) &= \sup_{i\in I} \rho_i(\lambda X + (1-\lambda)Y) \\ &\leq \sup_{i\in I} \Big(\lambda\rho_i(X) + (1-\lambda)\rho_i(Y)\Big) \\ &\leq \lambda \sup_{i\in I} \rho_i(X) + (1-\lambda) \sup_{i\in I} \rho_i(Y) \\ &= \lambda\rho(X) + (1-\lambda)\rho(Y).\end{aligned}$$

So ρ is convex. □

We now illustrate the concept of acceptance sets with two important examples.

In the following, we take $\mathcal{X}$ to be the linear space of all bounded measurable functions on some measurable space $(\Omega, \mathcal{F})$ and use the notation $\mathcal{M}_1$ and $\mathcal{M}_{1,f}$ from Section 2.4.

Example 3.15 (Worst-case risk measure)**.** Consider the worst-case risk measure defined by

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

This risk measure assigns to each position its largest possible loss over all scenarios and therefore corresponds to a purely pessimistic assessment of risk.

The associated acceptance set is

$$\mathcal{A} = \{X \in \mathcal{X} \mid \rho_{\max}(X) \leq 0\}.$$

Since

$$\rho_{\max}(X) \leq 0 \iff -\inf_{\omega\in\Omega} X(\omega) \leq 0 \iff \inf_{\omega\in\Omega} X(\omega) \geq 0,$$

it follows that

$$\mathcal{A} = \{X \in \mathcal{X} \mid X(\omega) \geq 0 \text{ for all } \omega \in \Omega\},$$

that is, the set of all nonnegative functions in $\mathcal{X}$. In particular, $\mathcal{A}$ is a convex cone, and therefore, by Proposition 3.11, $\rho_{\max}$ is a coherent risk measure. Moreover, $\rho_{\max}$ is the most conservative normalized monetary risk measure in the sense that every normalized monetary risk measure ρ on $\mathcal{X}$ satisfies $\rho(X) \leq \rho_{\max}(X)$. Indeed, since $X \geq \inf_{\omega\in\Omega} X(\omega)$, monotonicity yields

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

By cash-invariance and normalization, we have

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

Hence,

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

and therefore

$$\rho(X) \leq \rho_{\max}(X).$$

Finally, $\rho_{\max}$ admits the representation

$$\rho_{\max}(X) = \sup_{Q \in \mathcal{M}_1} \mathbb{E}_Q[-X],$$

where $\mathcal{M}_1$ denotes the set of all probability measures on $(\Omega, \mathcal{F})$.

Example 3.16 (Model-based acceptance sets)**.** Let $\mathcal{Q}$ be a set of probability measures on $(\Omega, \mathcal{F})$, and let $\gamma : \mathcal{Q} \to \mathbb{R}$ be a function satisfying $\sup_{Q \in \mathcal{Q}} \gamma(Q) < \infty$.

A position $X \in \mathcal{X}$ is called acceptable if

$$\mathbb{E}_Q[X] \geq \gamma(Q), \qquad \text{for all } Q \in \mathcal{Q}.$$

The corresponding acceptance set is therefore given by

$$\mathcal{A} = \{X \in \mathcal{X} \mid \mathbb{E}_Q[X] \geq \gamma(Q), \qquad \text{for all } Q \in \mathcal{Q}\}.$$

This formulation reflects a robust perspective, where acceptability is required under a whole family of probabilistic models. The function $\gamma(Q)$ can be interpreted as a model-dependent threshold that the expected payoff must exceed.

The associated monetary risk measure is given by

$$\rho(X) = \sup_{Q \in \mathcal{Q}} \Big(\gamma(Q) - \mathbb{E}_Q[X] \Big).$$

This risk measure is convex, and it quantifies the worst-case shortfall relative to the threshold $\gamma(Q)$ across all models $Q \in \mathcal{Q}$. Equivalently, it can be written in the form

$$\rho(X) = \sup_{Q \in \mathcal{M}_1} \Big(\mathbb{E}_Q[-X] - \alpha(Q) \Big),$$

where the penalty function α is defined by

$$\alpha(Q) = \begin{cases} -\gamma(Q), & \text{if } Q \in \mathcal{Q}, \\ +\infty, & \text{otherwise.} \end{cases}$$

These examples highlight how different specifications of acceptance sets lead to different classes of risk measures, ranging from worst-case assessments to model-based evaluations. Especially, the formulation in Example 3.16 anticipates the dual representation of convex risk measures, which is developed in the next subsection.

3.2 Robust Representation of Convex Risk Measures

The aim of this subchapter is to provide a deeper structural understanding of convex risk measures by linking them to families of probability measures. In particular, it will be shown that every convex risk measure admits a representation as a worst-case expected loss over a suitable set of alternative models, penalized by a function reflecting model uncertainty. This so-called robust representation highlights that convex risk measures naturally incorporate ambiguity about the underlying probability distribution and can be interpreted as evaluating risk under adverse scenarios. In this way, the results of this chapter establish a fundamental connection between risk measurement, convex duality, and the theory of probability measures.

As recalled in Section 1.2, the relaxation of coherence to convexity is due independently to Föllmer and Schied [25] and to Frittelli and Rosazza Gianin [27]. In the present chapter, this generalization takes concrete form. It is precisely the passage from the coherent representation of Proposition 3.17 to the convex representation of Theorem 3.19, in which the penalty function makes its first appearance.

In this section, we work on a fixed measurable space $(\Omega, \mathcal{F})$ and consider bounded measurable functions $X \in \mathcal{X}$, where $\mathcal{X}$ is endowed with the supremum norm. Throughout this section, all convergence statements are understood pointwise, unless stated otherwise. Rather than fixing a reference probability measure, we allow for model uncertainty in the sense of Knightian uncertainty by considering a family of probability measures on $(\Omega, \mathcal{F})$.

More precisely, we work with the classes $\mathcal{M}_1$ and $\mathcal{M}_{1,f}$ introduced in Section 2.4. The probability measures and the normalized finitely additive set functions on $(\Omega, \mathcal{F})$, respectively. Recall that each $Q \in \mathcal{M}_{1,f}$ induces a linear functional on $\mathcal{X}$ via integration,

$$X \mapsto \mathbb{E}_Q[X].$$

Such functionals can be interpreted as evaluating the payoff X under different probabilistic scenarios. This viewpoint naturally leads to the study of risk measures that admit representations in terms of expectations with respect to elements of $\mathcal{M}_{1,f}$, and specifically to worst-case evaluations over suitable classes of such measures.

Such representations can be further refined by exploiting the structural properties of convex risk measures. In particular, the combination of monotonicity, convexity, and cash invariance allows one to reinterpret risk measures in terms of suitable functionals that admit representations as infima over expectations with respect to finitely additive probability measures.

This observation provides the key step towards a dual characterization of convex risk measures, where the evaluation of a position can be expressed in terms of worst-case

expectations over a suitably chosen class of probability measures. In this way, the risk measure naturally incorporates model uncertainty by taking into account a range of alternative probabilistic scenarios, rather than relying on a single reference measure.

3.2.1 Representation Theorems

The goal of this subsection is to derive representation results for coherent and convex risk measures in terms of probability measures.

We begin with the case of coherent risk measures, where risk can be represented as the worst-case expected loss over a suitable set of probability measures.

Proposition 3.17. *[3] A functional $\rho : \mathcal{X} \to \mathbb{R}$ is a coherent risk measure if and only if there exists a subset $\mathcal{Q}$ of $\mathcal{M}_{1,f}$ such that*

$$\rho(X) = \sup_{Q \in \mathcal{Q}} \mathbb{E}_Q[-X], \quad X \in \mathcal{X}. \tag{3.8}$$

Moreover, $\mathcal{Q}$ can be chosen as a convex set for which the supremum is attained.

This representation is the foundational result of Artzner, Delbaen, Eber and Heath [3], where the coherent framework underlying this chapter was first introduced. Our goal is to extend this representation to the more general class of convex risk measures. In this setting, the representation involves an additional penalty term that reflects the plausibility of the underlying probability measure. This yields a more flexible framework that allows for a trade-off between model uncertainty and risk evaluation.

Let $\alpha : \mathcal{M}_{1,f} \to \mathbb{R} \cup \{+\infty\}$ be a functional such that

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

For each $Q \in \mathcal{M}_{1,f}$, the functional

$$X \mapsto \mathbb{E}_Q[-X] - \alpha(Q)$$

is convex, monotone, and cash invariant on $\mathcal{X}$, and these three properties are preserved when taking the supremum over $Q \in \mathcal{M}_{1,f}$. Hence,

$$\rho(X) := \sup_{Q \in \mathcal{M}_{1,f}} \Big(\mathbb{E}_Q[-X] - \alpha(Q) \Big) \tag{3.9}$$

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

$$\rho(0) = - \inf_{Q \in \mathcal{M}_{1,f}} \alpha(Q). \tag{3.10}$$

Definition 3.18. Let $\alpha : \mathcal{M}_{1,f} \to \mathbb{R} \cup \{+\infty\}$ be a functional such that

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

If a convex risk measure ρ admits the representation

$$\rho(X) = \sup_{Q \in \mathcal{M}_{1,f}} \Big(\mathbb{E}_Q[-X] - \alpha(Q) \Big),$$

then α is called a *penalty function* for ρ on $\mathcal{M}_{1,f}$.

The penalty function $\alpha(Q)$ can be interpreted as measuring the plausibility of the model Q, with larger values corresponding to less plausible scenarios. This allows us to formulate a dual representation of convex risk measures. The following theorem shows that every convex risk measure admits a representation in terms of its minimal penalty function. In this representation, the maximum is attained, which follows from compactness properties of the dual space.

Theorem 3.19. *[25, 27] Any convex risk measure ρ on $\mathcal{X}$ is of the form*

$$\rho(X) = \max_{Q \in \mathcal{M}_{1,f}} \Big(\mathbb{E}_Q[-X] - \alpha^{\min}(Q) \Big), \quad X \in \mathcal{X}, \tag{3.11}$$

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

$$\alpha^{\min}(Q) := \sup_{X \in \mathcal{X}} \Big(\mathbb{E}_Q[-X] - \rho(X) \Big), \quad Q \in \mathcal{M}_{1,f}. \tag{3.12}$$

Moreover, $\alpha^{\min}$ is the minimal penalty function that represents ρ, that is, any penalty function α for which the above representation holds satisfies

$$\alpha(Q) \geq \alpha^{\min}(Q) \quad \textit{for all } Q \in \mathcal{M}_{1,f}.$$

Proof. First, we argue that we may assume without loss of generality that $\rho(0) = 0$. Otherwise, we consider the convex risk measure $\tilde{\rho}(X) := \rho(X) - \rho(0)$. One checks that the acceptance set of $\tilde{\rho}$ satisfies

$$\mathcal{A}_{\tilde{\rho}} = \{X - \rho(0) \mid X \in \mathcal{A}_\rho\},$$

so that the minimal penalty function of $\tilde{\rho}$ is given by

$$\tilde{\alpha}^{\min}(Q) = \sup_{X \in \mathcal{A}_\rho} \mathbb{E}_Q[-X] = \sup_{X \in \mathcal{A}_\rho} \mathbb{E}_Q[-X] + \rho(0) = \alpha^{\min}(Q) + \rho(0).$$

It follows that the representation (3.11) holds for ρ if and only if it holds for $\tilde{\rho}$.

In the next step, we show that

$$\rho(X) \geq \sup_{Q \in \mathcal{M}_{1,f}} \left(\mathbb{E}_Q[-X] - \alpha^{\min}(Q) \right) \quad \text{for all } X \in \mathcal{X}.$$

To this end, recall that $X' := \rho(X) + X \in \mathcal{A}_\rho$ by (3.1). Thus, for all $Q \in \mathcal{M}_{1,f}$,

$$\alpha^{\min}(Q) \geq \mathbb{E}_Q[-X'] = \mathbb{E}_Q[-X] - \rho(X).$$

From here, the inequality follows.

For a fixed X, we now construct some $Q_X \in \mathcal{M}_{1,f}$ such that

$$\rho(X) \leq \mathbb{E}_{Q_X}[-X] - \alpha^{\min}(Q_X),$$

which, in view of the previous step, proves the representation. By cash invariance, it suffices to prove this for $X \in \mathcal{X}$ with $\rho(X) = 0$. Due to our assumption $\rho(0) = 0$, X is not contained in the nonempty convex set

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

Since $\mathcal{B}$ is open due to Lemma 3.6, we may apply the separation argument in the form of Theorem 2.17. It yields a nonzero continuous linear functional ℓ on $\mathcal{X}$ such that

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

We claim that $\ell(Y) \geq 0$ if $Y \geq 0$. Monotonicity and cash invariance of ρ imply that $1 + \lambda Y \in \mathcal{B}$ for any $\lambda > 0$. Hence,

$$\ell(X) \leq \ell(1 + \lambda Y) = \ell(1) + \lambda \ell(Y) \quad \text{for all } \lambda > 0,$$

which could not be true if $\ell(Y) < 0$.

Our next claim is that $\ell(1) > 0$. Since ℓ does not vanish identically, there must be some Y such that $0 < \ell(Y) = \ell(Y^+) - \ell(Y^-)$. We may assume without loss of generality that $\|Y\| < 1$. Positivity of ℓ implies $\ell(Y^+) > 0$ and $\ell(1 - Y^+) \geq 0$. Therefore,

$$\ell(1) = \ell(1 - Y^+) + \ell(Y^+) > 0.$$

By the two preceding steps and Theorem 2.25, we conclude that there exists some $Q_X \in \mathcal{M}_{1,f}$ such that

$$\mathbb{E}_{Q_X}[Y] = \frac{\ell(Y)}{\ell(1)} \quad \text{for all } Y \in \mathcal{X}.$$

Note that $\mathcal{B} \subset \mathcal{A}_\rho$, and so

$$\alpha^{\min}(Q_X) = \sup_{Y \in \mathcal{A}_\rho} \mathbb{E}_{Q_X}[-Y] \geq \sup_{Y \in \mathcal{B}} \mathbb{E}_{Q_X}[-Y] = -\frac{b}{\ell(1)}.$$

On the other hand, $Y + \varepsilon \in \mathcal{B}$ for any $Y \in \mathcal{A}_\rho$ and each $\varepsilon > 0$. This shows that $\alpha^{\min}(Q_X)$ is in fact equal to $-b/\ell(1)$. It follows that

$$\mathbb{E}_{Q_X}[-X] - \alpha^{\min}(Q_X) = \frac{1}{\ell(1)}\Big(b - \ell(X)\Big) \geq 0 = \rho(X).$$

Thus, Q_X is as desired, and the proof of the representation is complete.

Finally, we need to prove that $\alpha^{\min}$ is indeed the minimal penalty function. Let α be any penalty function for ρ. Then, for all $Q \in \mathcal{M}_{1,f}$ and $X \in \mathcal{X}$,

$$\rho(X) \geq \mathbb{E}_Q[-X] - \alpha(Q),$$

and hence

$$\alpha(Q) \geq \sup_{X \in \mathcal{X}} \Big(\mathbb{E}_Q[-X] - \rho(X)\Big) \geq \sup_{X \in \mathcal{A}_\rho} \Big(\mathbb{E}_Q[-X] - \rho(X)\Big) \geq \alpha^{\min}(Q). \tag{3.13}$$

Thus, α dominates $\alpha^{\min}$. □

The representation

$$\rho(X) = \sup_{Q \in \mathcal{Q}} \mathbb{E}_Q[-X], \quad X \in \mathcal{X}, \tag{3.14}$$

of a coherent risk measure ρ, as stated in Proposition 3.17, can be viewed as a special instance of the general representation result for convex risk measures. In this case, the associated penalty function is given by

$$\alpha(Q) = \begin{cases} 0, & \text{if } Q \in \mathcal{Q}, \\ +\infty, & \text{otherwise.} \end{cases}$$

The following corollary shows that the minimal penalty function of any coherent risk measure is necessarily of this form. This observation also shows that the proof of the representation result for coherent risk measures does not require a separate argument. Indeed, it follows directly from the proof of the general representation theorem for convex risk measures by specializing to penalty functions that take only the values 0 and $+\infty$.

Corollary 3.20. *The minimal penalty function $\alpha^{\min}$ of a coherent risk measure ρ takes only the values 0 and $+\infty$. In particular,*

$$\rho(X) = \max_{Q \in \mathcal{Q}_{\max}} \mathbb{E}_Q[-X], \quad X \in \mathcal{X}, \tag{3.15}$$

for the convex set

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

and $\mathcal{Q}_{\max}$ *is the largest set for which a representation of the above form* (3.15) *holds.*

Proof. Recall from Proposition 3.11 that the acceptance set $\mathcal{A}_\rho$ of a coherent risk measure is a cone. Thus, the minimal penalty function satisfies

$$\alpha^{\min}(Q) = \sup_{X \in \mathcal{A}_\rho} \mathbb{E}_Q[-X] = \sup_{\lambda X \in \mathcal{A}_\rho} \mathbb{E}_Q[-\lambda X] = \lambda \alpha^{\min}(Q)$$

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

Remark 3.21. In the coherent case, the penalty function takes only the values 0 and $+\infty$, which means that all probability measures in $\mathcal{Q}$ are considered equally plausible, while all others are excluded entirely. As a result, the risk measure reduces to a worst-case expectation over the set $\mathcal{Q}$.

From an economic perspective, this reflects a model of uncertainty without ambiguity sensitivity. This means that no preference is given to one model over another within $\mathcal{Q}$, and no gradual penalization is applied. In contrast, convex risk measures allow for more refined attitudes towards model uncertainty by assigning different penalties to different probability measures.

A more detailed analysis of the properties of these representations of coherent and convex risk measures is provided in Appendix Section B.1.

The penalty function α in the representation (3.9) of a convex risk measure is in general not unique, and the minimal penalty function does not necessarily provide the most useful representation. The definitive treatment of uniqueness for the law-invariant case is given by Shapiro [47]; see the discussion following Theorem 3.54.

The worst-case risk measure already illustrates this in Example 3.15, where the risk measure is defined as the supremum over a set of probability measures. More generally, such constructions can be understood in terms of pointwise suprema of convex risk measures, as shown in the following proposition.

Proposition 3.22. *Suppose that for every* i *in some index set* I *we are given a convex risk measure* ρ_i *on* $\mathcal{X}$ *with associated penalty function* α_i*. If* $\sup_{i \in I} \rho_i(0) < \infty$*, then*

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

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

$$\alpha(Q) := \inf_{i \in I} \alpha_i(Q), \quad Q \in \mathcal{M}_{1,f}.$$

Proof. It was shown in Lemma 3.14 that ρ is a convex risk measure. It remains to show

that it admits the stated representation. Indeed,

$$\begin{aligned}\rho(X) &= \sup_{i\in I} \rho_i(X) \\ &= \sup_{i\in I} \sup_{Q\in\mathcal{M}_{1,f}} \Big(\mathbb{E}_Q[-X] - \alpha_i(Q)\Big) \\ &= \sup_{Q\in\mathcal{M}_{1,f}} \sup_{i\in I} \Big(\mathbb{E}_Q[-X] - \alpha_i(Q)\Big) \\ &= \sup_{Q\in\mathcal{M}_{1,f}} \Big(\mathbb{E}_Q[-X] - \inf_{i\in I} \alpha_i(Q)\Big).\end{aligned}$$

This yields the desired representation. □

In the sequel, we focus on those convex risk measures that admit a representation in terms of σ-additive probability measures, that is, probability measures in $\mathcal{M}_1$. In this case, the representation takes the form

$$\rho(X) = \sup_{Q\in\mathcal{M}_1} \Big(\mathbb{E}_Q[-X] - \alpha(Q)\Big). \tag{3.16}$$

However, in contrast to the general representation over $\mathcal{M}_{1,f}$, one can no longer expect the supremum to be attained. This observation indicates that additional structural conditions are required in order to obtain more refined representation results.

In fact, representations in terms of σ-additive probability measures are closely related to continuity properties of the risk measure. We therefore turn to the study of such properties, beginning with notions of continuity based on the order structure of random variables.

3.2.2 Continuity Properties

In order to derive refined representation results, it is natural to impose suitable continuity requirements on monetary risk measures. These properties describe how the risk assessment behaves under limiting procedures and will be essential in the passage from finitely additive to σ-additive probability measures.

Definition 3.23. Let ρ be a monetary risk measure on $\mathcal{X}$.

- We say that ρ is *continuous from above* if for every sequence (X_n) in $\mathcal{X}$ such that $X_n \searrow X$ pointwise, it holds that

$$\rho(X_n) \nearrow \rho(X).$$

- We say that ρ is *continuous from below* if for every sequence (X_n) in $\mathcal{X}$ such that $X_n \nearrow X$ pointwise, it holds that

$$\rho(X_n) \searrow \rho(X).$$

These properties describe how the risk measure behaves under monotone approximations of financial positions and are particularly natural in a probabilistic setting.

In economic terms, these continuity properties express that the assessment of risk reacts consistently to gradual changes in a position. Especially, if a sequence of positions improves monotonically, then the associated risk should decrease in a stable way, while a monotone deterioration of positions should be reflected in a corresponding increase of risk in the limit.

A weaker but more flexible regularity condition is given by the Fatou property.

Definition 3.24. A monetary risk measure ρ on $\mathcal{X}$ is said to satisfy the *Fatou property* if for every uniformly bounded sequence (X_n) in $\mathcal{X}$ that converges $\mathbb{P}$-almost surely to X, one has

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

This condition ensures that the risk of a limiting position is not underestimated compared to the risks of the approximating positions.

The Fatou property is closely related to lower semicontinuity. Recall that a function $f\colon \mathcal{X} \to (-\infty, \infty]$ on a topological space is *lower semicontinuous* if all sublevel sets $\{f \leq c\}$, $c \in \mathbb{R}$, are closed. Since ρ is Lipschitz continuous with respect to the supremum norm, lower semicontinuity is only a genuine restriction with respect to coarser topologies; the relevant one in our setting is the weak* topology $\sigma(L^\infty, L^1)$. The two notions should not be conflated: the Fatou property is a sequential condition tied to the order structure of L^∞, whereas weak* lower semicontinuity is a topological condition that, since $\sigma(L^\infty, L^1)$ is not metrizable, cannot be verified by sequences alone. That the two conditions are nevertheless equivalent for convex risk measures is a theorem rather than a definitional identity.

Building on the continuity properties introduced above, which ensure the stability of risk measures under limiting procedures, we now return to the robust representation and take a closer look at the structure of the associated penalty function. In particular, the following observations provide further insight into its analytical properties and its relation to convex duality.

Remark 3.25. (a) If we take $\alpha = \alpha^{\min}$ in (3.13), then all inequalities in (3.13) become identities. Thus, we obtain the alternative representation

$$\alpha^{\min}(Q) = \sup_{X\in\mathcal{X}} \Big(\mathbb{E}_Q[-X] - \rho(X)\Big). \tag{3.17}$$

(b) The minimal penalty function $\alpha^{\min}$ is convex and lower semicontinuous with respect to the total variation distance on $\mathcal{M}_{1,f}$, since it is the supremum of affine continuous functions on $\mathcal{M}_{1,f}$.

(c) Suppose ρ is defined via $\rho := \rho_{\mathcal{A}}$ for a given acceptance set $\mathcal{A} \subset \mathcal{X}$. Then $\mathcal{A}$

determines $\alpha^{\min}$:

$$\alpha^{\min}(Q) = \sup_{X \in \mathcal{A}} \mathbb{E}_Q[-X] \qquad \text{for all } Q \in \mathcal{M}_{1,f}.$$

This follows from part (d) of Proposition 3.13, which identifies $\mathcal{A}_\rho$ as the closure of $\mathcal{A}$ with respect to the supremum norm $\|\cdot\|$, together with the $\|\cdot\|$-continuity of the mapping $X \mapsto \mathbb{E}_Q[-X]$ for $Q \in \mathcal{M}_{1,f}$.

The previous remark already suggests that the minimal penalty function admits a dual interpretation. This connection becomes explicit when viewed through the lens of convex analysis.

Remark 3.26. Equation (3.17) shows that the penalty function $\alpha^{\min}$ may be viewed as the Fenchel–Legendre transform, or conjugate function, of the convex function ρ on the Banach space $\mathcal{X}$. This is exactly the Fenchel–Legendre transform of Subsection 2.5.1, applied to the convex risk measure ρ on the Banach space $\mathcal{X}$: the penalty function $\alpha^{\min}$ is the conjugate ρ^*, and the dual representation we are about to derive is the corresponding biconjugation $\rho = \rho^{**}$. More precisely,

$$\alpha^{\min}(Q) = \rho^*(\ell_Q),$$

where the convex conjugate $\rho^* : \mathcal{X}^* \to \mathbb{R} \cup \{+\infty\}$ is defined on the dual $\mathcal{X}^*$ of $\mathcal{X}$ by

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

and where $\ell_Q \in \mathcal{X}^*$ is given by

$$\ell_Q(X) = \mathbb{E}_Q[-X] \quad \text{for } Q \in \mathcal{M}_{1,f}.$$

This viewpoint suggests an alternative proof of Theorem 3.19. First, note that $\mathcal{X}^*$ can be identified with the space $ba := ba(\Omega, \mathcal{F})$ of finitely additive set functions with finite total variation. Since ρ is lower semicontinuous with respect to the weak topology $\sigma(\mathcal{X}, \mathcal{X}^*)$, any sublevel set $\{\rho \le c\}$ being convex and strongly closed by Lemma 3.6, hence weakly closed, the Fenchel–Moreau theorem (Theorem 2.31) applies and yields

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

where ρ^{**} denotes the conjugate function of ρ^*, that is,

$$\rho(X) = \sup_{\ell \in ba} \Big(\ell(X) - \rho^*(\ell) \Big).$$

It can now be checked that by the second part of the proof of Theorem 3.19, monotonicity and cash invariance of ρ imply that $\ell \le 0$ and $\ell(1) = -1$ for any $\ell \in \mathcal{X}^* = ba$

such that $\rho^*(\ell) < \infty$. Identifying $-\ell$ with $Q \in \mathcal{M}_{1,f}$ and using $\alpha^{\min}(Q) = \rho^*(\ell_Q)$, $\rho(X) = \sup_{\ell \in ba}\Big(\ell(X) - \rho^*(\ell)\Big)$ reduces to the representation

$$\rho(X) = \sup_{Q \in \mathcal{M}_{1,f}} \Big(\mathbb{E}_Q[-X] - \alpha^{\min}(Q)\Big).$$

Moreover, the supremum is actually attained: $\mathcal{M}_{1,f}$ is weak* compact in $\mathcal{X}^* = ba$ due to the Banach–Alaoglu theorem, and so the upper semicontinuous functional

$$Q \mapsto \mathbb{E}_Q[-X] - \alpha^{\min}(Q)$$

attains its maximum on $\mathcal{M}_{1,f}$. Thus, the representation (3.11) of a convex risk measure ρ follows from the general theory of convex duality. For this reason, it is also called the dual representation of ρ.

We now turn to continuity properties of convex risk measures, which play are fundamental in the analysis of their robust representations. Explicitly continuity from above turns out to be closely related to the Fatou property and provides important insight into the structure of the dual representation.

Lemma 3.27. *A convex risk measure ρ that admits a representation (3.16) on $\mathcal{M}_1$ is continuous from above. Moreover, continuity from above is equivalent to the Fatou property of lower semicontinuity with respect to bounded pointwise convergence.*

Proof. First, we show the Fatou property under the assumption that ρ has a representation in terms of probability measures. Dominated convergence implies that

$$\mathbb{E}_Q[X_n] \to \mathbb{E}_Q[X] \qquad \text{for each } Q \in \mathcal{M}_1.$$

Hence,

$$\begin{aligned}
\rho(X) &= \sup_{Q \in \mathcal{M}_1} \Big(\mathbb{E}_Q[-X] - \alpha(Q)\Big) \\
&= \sup_{Q \in \mathcal{M}_1} \lim_{n\to\infty} \Big(\mathbb{E}_Q[-X_n] - \alpha(Q)\Big) \\
&\leq \liminf_{n\to\infty} \sup_{Q \in \mathcal{M}_1} \Big(\mathbb{E}_Q[-X_n] - \alpha(Q)\Big) \\
&= \liminf_{n\to\infty} \rho(X_n).
\end{aligned}$$

In order to show the equivalence of the Fatou property and continuity from above, let us first assume the Fatou property. By monotonicity, $\rho(X_n) \leq \rho(X)$ for each n if $X_n \searrow X$, and therefore $\rho(X_n)$ increases to $\rho(X)$.

Conversely, assume continuity from above. Let (X_n) be a bounded sequence in $\mathcal{X}$

that converges pointwise to X. Define

$$Y_n := \sup_{m \geq n} X_m \in \mathcal{X}.$$

Then $Y_n \searrow X$. Since $X_n \leq Y_n$, monotonicity yields $\rho(X_n) \geq \rho(Y_n)$. Thus,

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

□

While continuity from above follows naturally from the robust representation, continuity from below turns out to be the key property for obtaining a more refined dual representation. In fact, it can be shown that it characterizes precisely when the minimal penalty function is concentrated on σ-additive probability measures.

Theorem 3.28. *For a convex risk measure ρ on $\mathcal{X}$, the following two conditions are equivalent:*

(a) ρ is continuous from below.

(b) The minimal penalty function $\alpha^{\min}$ is concentrated on the class $\mathcal{M}_1$ of probability measures, that is, $\alpha^{\min}(Q) = \infty$ whenever Q is not σ-additive.

In particular, we then have the representation

$$\rho(X) = \max_{Q \in \mathcal{M}_1} \Big(\mathbb{E}_Q[-X] - \alpha^{\min}(Q) \Big), \quad X \in \mathcal{X},$$

whenever one of these two equivalent conditions is satisfied.

The proof of this result relies on the following auxiliary lemma, which establishes a key relation between level sets of the penalty function and uniform convergence properties.

Lemma 3.29. *Let ρ be a convex risk measure on $\mathcal{X}$ that is represented by the penalty function α on $\mathcal{M}_{1,f}$, and consider the level sets*

$$\Lambda_c := \{Q \in \mathcal{M}_{1,f} \mid \alpha(Q) \leq c\}, \qquad \text{for } c > -\rho(0) = \inf_{Q \in \mathcal{M}_{1,f}} \alpha(Q).$$

For any sequence (X_n) in $\mathcal{X}$ such that $0 \leq X_n \leq 1$, the following two conditions are equivalent:

(a) $\rho(\lambda X_n) \to \rho(\lambda)$ for each $\lambda \geq 1$,

(b) $\inf_{Q \in \Lambda_c} \mathbb{E}_Q[X_n] \to 1$ for all $c > -\rho(0)$.

Proof. (a)⇒(b): In a first step, we show that for all $Y \in \mathcal{X}$,

$$\inf_{Q\in\Lambda_c} \mathbb{E}_Q[Y] \geq -\frac{c+\rho(\lambda Y)}{\lambda} \qquad \text{for all } \lambda > 0. \tag{3.18}$$

Indeed, since α represents ρ, we have for $Q \in \Lambda_c$,

$$c \geq \alpha(Q) \geq \mathbb{E}_Q[-\lambda Y] - \rho(\lambda Y),$$

and dividing by $-\lambda$ yields (3.18).

Now consider a sequence (X_n) that satisfies (a). Then (3.18) shows that for all $\lambda \geq 1$,

$$\liminf_{n\to\infty} \inf_{Q\in\Lambda_c} \mathbb{E}_Q[X_n] \geq -\lim_{n\to\infty} \frac{c+\rho(\lambda X_n)}{\lambda} = 1 - \frac{c+\rho(0)}{\lambda}.$$

Taking $\lambda \uparrow \infty$ and assuming $X_n \leq 1$ proves (b).

(b)⇒(a): Clearly, for all $Y \in \mathcal{X}$ with $0 \leq Y \leq 1$,

$$\rho(\lambda) \leq \rho(\lambda Y) = \sup_{Q\in\mathcal{M}_{1,f}} \Big(\mathbb{E}_Q[-\lambda Y] - \alpha(Q)\Big).$$

Since $\mathbb{E}_Q[-\lambda Y] \leq 0$ for all Q, only those Q can contribute to the supremum on the right-hand side for which

$$\alpha(Q) \leq \delta - \rho(\lambda) = \delta + \lambda - \rho(0) =: c,$$

where $\delta > 0$ is an arbitrary but fixed constant. Hence,

$$\rho(\lambda Y) = \sup_{Q\in\Lambda_c} \Big(\mathbb{E}_Q[-\lambda Y] - \alpha(Q)\Big), \tag{3.19}$$

and particularly for all n,

$$\rho(\lambda X_n) = \sup_{Q\in\Lambda_c} \Big(\mathbb{E}_Q[-\lambda X_n] - \alpha(Q)\Big).$$

But condition (b) implies that $\mathbb{E}_Q[-\lambda X_n]$ converges to $-\lambda$ uniformly in $Q \in \Lambda_c$. Therefore, (3.19) yields that

$$\rho(\lambda X_n) = \sup_{Q\in\Lambda_c} \Big(\mathbb{E}_Q[-\lambda X_n] - \alpha(Q)\Big) \longrightarrow \sup_{Q\in\Lambda_c} \Big(-\lambda - \alpha(Q)\Big) = \rho(\lambda),$$

as $n \uparrow \infty$. □

The proof of Theorem 3.28 will also rely on a classical result from analysis, called *Dini's lemma*:

Lemma 3.30. *On a compact set, a sequence of continuous functions f_n increasing to a continuous function f converges even uniformly.*

Proof. For $\varepsilon > 0$, the compact sets $K_n := \{f_n \leq f - \varepsilon\}$ satisfy $\bigcap_n K_n = \emptyset$. Hence, there exists some n_0 such that $K_n = \emptyset$ for all $n \geq n_0$. □

Proof of Theorem 3.28. To prove the implication *(a)⇒(b)*, recall that $Q \in \mathcal{M}_{1,f}$ is σ-additive if and only if

$$Q[A_n] \nearrow 1$$

for any increasing sequence of events $A_n \in \mathcal{F}$ such that

$$\bigcup_n A_n = \Omega.$$

Thus, our claim is implied by the implication *(a)⇒(b)* of Lemma 3.29 if we take

$$X_n := \mathbf{1}_{A_n}.$$

We now prove the implication *(b)⇒(a)*. Suppose $X_n \nearrow X$ pointwise on Ω. We need to show that $\rho(X_n) \searrow \rho(X)$. By cash invariance, we may assume without loss of generality that $X_n \geq 0$ for all n. As in (3.19), we see that

$$\rho(X_n) = \max_{Q \in \mathcal{M}_{1,f}} \Big(\mathbb{E}_Q[-X_n] - \alpha^{\min}(Q) \Big) = \max_{Q \in \Lambda_c} \Big(\mathbb{E}_Q[-X_n] - \alpha^{\min}(Q) \Big), \tag{3.20}$$

where $c > 1 - \rho(X)$ and $\Lambda_c := \{Q \in \mathcal{M}_{1,f} \mid \alpha^{\min}(Q) \leq c\}$. We will show below that $\Lambda_c \subseteq \mathcal{M}_1$ implies that

$$\mathbb{E}_Q[-X_n] \longrightarrow \mathbb{E}_Q[-X] \qquad \text{uniformly in } Q \in \Lambda_c. \tag{3.21}$$

Together with the representation (3.20), this will imply the desired convergence

$$\rho(X_n) \to \rho(X).$$

To prove (3.21), recall first from the preliminaries that $\mathcal{M}_{1,f}$ belongs to the larger vector space

$$ba := ba(\Omega, \mathcal{F})$$

of all finitely additive set functions

$$\mu : \mathcal{F} \to \mathbb{R}$$

with finite total variation $\|\mu\|_{\text{var}}$. In fact, ba can be identified with the topological dual of the Banach space $\mathcal{X}$ with respect to $\|\cdot\|$. Since we can write

$$\mathcal{M}_{1,f} = \{\mu \in ba \mid \mu(\Omega) = 1 \text{ and } \mu(A) \geq 0 \text{ for all } A \in \mathcal{F}\},$$

it is clear that $\mathcal{M}_{1,f}$ is a bounded and weak* closed set in ba. For this reason, $\mathcal{M}_{1,f}$ is

weak* compact by the Banach–Alaoglu theorem. Since $\alpha^{\min}$ is weak* lower semicontinuous on $\mathcal{M}_{1,f}$ as supremum of the weak* continuous maps

$$Q \mapsto \mathbb{E}_Q[-Y] \qquad \text{with } Y \in \mathcal{A}_\rho,$$

the level set Λ_c is also weak* compact.

After these preparations, we can now prove (3.21). Clearly, the functions

$$\ell_n(Q) := \mathbb{E}_Q[-X_n]$$

form a decreasing sequence of weak* continuous functions on Λ_c. Moreover, if $\Lambda_c \subseteq \mathcal{M}_1$, we even have

$$\ell_n(Q) \searrow \ell(Q) := \mathbb{E}_Q[-X] \qquad \text{for each } Q \in \Lambda_c,$$

due to the monotone convergence theorem. By the established compactness of Λ_c, (3.21) thus follows from Dini's lemma. □

Remark 3.31. Let ρ be a convex risk measure that is continuous from below. Then ρ is also continuous from above. This follows by combining Theorem 3.28 with Lemma 3.27, which together show that continuity from below already guarantees that the minimal penalty function $\alpha^{\min}$ is concentrated on $\mathcal{M}_1$, and hence the representation of ρ involves only σ-additive probability measures. In particular, continuity from below is strictly stronger than continuity from above, and equivalently than the Fatou property.

Equivalent characterizations of continuity from below, including its relation to the Lebesgue property and to associated coherent risk measures, are collected in Appendix B.1, Propositions B.4 and B.5 respectively.

We conclude this section by summarizing the role of the continuity properties introduced above.

Continuity from above, continuity from below, and the Fatou property establish fundamental links between the probabilistic structure of random variables and the functional-analytic framework of risk measures. They ensure that risk evaluations behave consistently under natural approximation procedures and support robust dual representations in terms of probability measures.

More specifically, the ordering of these properties by strength, continuity from below being strictly stronger than continuity from above, equivalently the Fatou property, as noted in Remark 3.31, is the mechanism that governs, throughout the sequel, whether a robust representation is merely a supremum or an attained maximum. Continuity from above (Theorem 3.33) is exactly what makes the minimal penalty function concentrate on $\mathcal{M}_1(\mathbb{P})$, turning the representation into a genuine supremum over probability measures; continuity from below (Corollary 3.34), the strictly stronger property, is what upgrades that supremum to a maximum by forcing weak compactness of the relevant sublevel sets, while on a general measurable space the same distinction is governed by Theorem 3.28.

Every later instance in which a supremum cannot be replaced by a maximum traces back to a failure of continuity from below; we will point this out explicitly where it occurs.

3.3 Convex Risk Measures on L^∞

In this section, we restrict our attention to convex risk measures defined on the space $L^\infty(\Omega, \mathcal{F}, \mathbb{P})$ of essentially bounded random variables. This setting is particularly relevant in financial applications, as bounded positions naturally arise when modeling portfolios with limited exposure.

Moreover, working on L^∞ allows us to exploit tools from functional analysis and duality theory. In particular, risk measures can be studied via their behavior under changes of probability measure, which leads to representations in terms of expectations under alternative models. The duality theory of this section, the equivalence of the Fatou property with $\sigma(L^\infty, L^1)$-lower semicontinuity, and the weak-compactness criterion underlying Corollary 3.34, is due to Delbaen [19], who established it for coherent risk measures on general probability spaces.

Throughout the next sections, we fix a reference probability measure $\mathbb{P}$ and consider risk measures ρ that depend only on the $\mathbb{P}$-almost sure equivalence class of a position, that is,

$$\rho(X) = \rho(Y) \quad \text{whenever } X = Y \ \ \mathbb{P}\text{-almost surely.} \tag{3.22}$$

This reflects the idea that events of probability zero do not affect the assessment of risk.

Lemma 3.32. *Let ρ be a convex risk measure that satisfies (3.22) and that is represented by a penalty function α as in (3.9). Then $\alpha(Q) = +\infty$ for any $Q \in \mathcal{M}_{1,f}(\Omega, \mathcal{F})$ that is not absolutely continuous with respect to $\mathbb{P}$.*

Proof. If $Q \in \mathcal{M}_{1,f}(\Omega, \mathcal{F})$ is not absolutely continuous with respect to $\mathbb{P}$, then there exists $A \in \mathcal{F}$ such that $Q[A] > 0$ but $\mathbb{P}[A] = 0$. Take any $X \in \mathcal{A}_\rho$, and define $X_n := X - n\mathbf{1}_A$. Then $\rho(X_n) = \rho(X)$, that is, X_n is again contained in $\mathcal{A}_\rho$. Consequently,

$$\alpha(Q) \geq \alpha^{\min}(Q) \geq \mathbb{E}_Q[-X_n] = \mathbb{E}_Q[-X] + nQ[A] \to \infty$$

as $n \to \infty$. □

The previous lemma shows that, in the representation of a convex risk measure on L^∞, it is sufficient to consider probability measures that are absolutely continuous with respect to $\mathbb{P}$. Indeed, any measure Q that is not absolutely continuous with respect to $\mathbb{P}$ is assigned an infinite penalty and therefore becomes irrelevant in the supremum appearing in the dual representation. In other words, the relevant dual variables can be restricted to densities with respect to $\mathbb{P}$, which significantly simplifies the structure of the representation.

In view of (3.22), we identify $\mathcal{X}$ with the Banach space $L^\infty := L^\infty(\Omega, \mathcal{F}, \mathbb{P})$. In line with the preceding discussion, we restrict attention to

$$\mathcal{M}_1(\mathbb{P}) := \mathcal{M}_1(\Omega, \mathcal{F}, \mathbb{P}) = \{Q \in \mathcal{M}_1(\Omega, \mathcal{F}) \mid Q \ll \mathbb{P}\},$$

the class in which all representations of this section will be formulated.

The following theorem establishes the precise conditions under which a convex risk measure admits a robust representation in terms of probability measures, and relates this representation to continuity properties and the geometry of the associated acceptance set.

Theorem 3.33. *[19] Suppose $\rho : L^\infty \to \mathbb{R}$ is a convex risk measure. Then the following conditions are equivalent:*

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

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

$$\rho(X) = \sup_{Q \in \mathcal{M}_1(\mathbb{P})} \left(\mathbb{E}_Q[-X] - \alpha^{\min}(Q) \right), \quad X \in L^\infty. \tag{3.23}$$

(c) ρ is continuous from above: if $X_n \searrow X$ $\mathbb{P}$-a.s., then $\rho(X_n) \nearrow \rho(X)$.

(d) ρ has the following Fatou property: for any uniformly bounded sequence (X_n) in L^∞ that converges $\mathbb{P}$-a.s. to some X,

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

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

(f) The acceptance set $\mathcal{A}_\rho$ of ρ is weak closed in L^∞, that is, $\mathcal{A}_\rho$ is closed with respect to the topology $\sigma(L^\infty, L^1)$.*

Proof. The implications $(b) \Rightarrow (a)$ and $(e) \Rightarrow (f)$ are obvious, and $(a) \Rightarrow (c) \Leftrightarrow (d)$ follows as in Lemma 3.27, replacing pointwise convergence by $\mathbb{P}$-a.s. convergence.

$(c) \Rightarrow (e)$: 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 $\mathbb{P}$-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 Lemma 2.22 implies that $\mathcal{C} := \{\rho \leq c\}$ is weak* closed.

$(f) \Rightarrow (b)$: We fix some $X \in L^\infty$ and let

$$m := \sup_{Q \in \mathcal{M}_1(\mathbb{P})} \Big(\mathbb{E}_Q[-X] - \alpha^{\min}(Q) \Big). \tag{3.24}$$

In view of Theorem 3.19, it suffices to show that $m \geq \rho(X)$, or equivalently that $m + X \in \mathcal{A}_\rho$. Suppose by way of contradiction that $m + X \notin \mathcal{A}_\rho$. Since the nonempty convex set $\mathcal{A}_\rho$ is weak* closed by assumption, we may apply the Hahn–Banach separation theorem in the locally convex space $(L^\infty, \sigma(L^\infty, L^1))$ with $\mathcal{C} := \mathcal{A}_\rho$ and $\mathcal{B} := \{m + X\}$. We obtain a continuous linear functional ℓ on $(L^\infty, \sigma(L^\infty, L^1))$ such that

$$\beta := \inf_{Y \in \mathcal{A}_\rho} \ell(Y) > \ell(m + X) =: \gamma > -\infty. \tag{3.25}$$

Since ℓ is continuous with respect to $\sigma(L^\infty, L^1)$, it is by definition of this topology of the form $\ell(Y) = \mathbb{E}[YZ]$ for some $Z \in L^1$. We claim that $Z \geq 0$ $\mathbb{P}$-a.s. To show this, fix $Y \geq 0$ and note that $\rho(\lambda Y) \leq \rho(0)$ for $\lambda \geq 0$, by monotonicity. As a result, $\lambda Y + \rho(0) \in \mathcal{A}_\rho$ for all $\lambda \geq 0$. It follows that

$$-\infty < \gamma < \ell(\lambda Y + \rho(0)) = \lambda \ell(Y) + \ell(\rho(0)).$$

Taking $\lambda \uparrow \infty$ yields that $\ell(Y) \geq 0$ and in turn that $Z \geq 0$ $\mathbb{P}$-almost surely. Moreover, $\mathbb{P}[Z > 0] > 0$ since ℓ is nonzero.

Define

$$\frac{dQ_0}{d\mathbb{P}} := \frac{Z}{\mathbb{E}[Z]},$$

which yields a probability measure $Q_0 \in \mathcal{M}_1(\mathbb{P})$. By (3.25), we see that

$$\alpha^{\min}(Q_0) = \sup_{Y \in \mathcal{A}_\rho} \mathbb{E}_{Q_0}[-Y] = -\frac{\beta}{\mathbb{E}[Z]}.$$

However,

$$\mathbb{E}_{Q_0}[X] + m = \frac{\ell(m + X)}{\mathbb{E}[Z]} = \frac{\gamma}{\mathbb{E}[Z]} < \frac{\beta}{\mathbb{E}[Z]} = -\alpha^{\min}(Q_0),$$

in contradiction to (3.24). Hence, $m + X$ must belong to $\mathcal{A}_\rho$, and thus $m \geq \rho(X)$. □

The preceding theorem provides a comprehensive characterization of convex risk measures on L^∞ that admit a robust representation in terms of probability measures. Above all, it shows that continuity properties such as continuity from above, the Fatou property, and lower semicontinuity are equivalent to the existence of such a representation.

From a financial perspective, this result implies that the risk of a position X can be interpreted as a worst-case expected loss over a class of probability measures, adjusted by a penalty term that reflects the plausibility of each model $Q \in \mathcal{M}_1(\mathbb{P})$.

The following corollary refines the above representation by identifying conditions

under which the minimal penalty function is concentrated on probability measures that are absolutely continuous with respect to the reference measure $\mathbb{P}$. More precisely, it characterizes those convex risk measures on L^∞ that satisfy continuity from below, which is stronger than continuity from above and yields a sharper form of the dual representation.

Corollary 3.34. *[19] For a convex risk measure ρ on L^∞, the following conditions are equivalent:*

(a) ρ is continuous from below.

(b) ρ satisfies the Lebesgue property: $\rho(X_n) \to \rho(X)$ whenever (X_n) is a bounded sequence in L^∞ that converges $\mathbb{P}$-a.s. to X.

(c) The minimal penalty function $\alpha^{\min}$ is concentrated on $\mathcal{M}_1(\mathbb{P})$, that is,

$$\alpha^{\min}(Q) < \infty \;\Rightarrow\; Q \in \mathcal{M}_1(\mathbb{P}).$$

(d) For each $c \in \mathbb{R}$, the level set $\Lambda_c := \{Q \mid \alpha^{\min}(Q) \leq c\}$ is contained in $\mathcal{M}_1(\mathbb{P})$, and the corresponding set of densities,

$$\left\{ \frac{dQ}{d\mathbb{P}} \;\middle|\; Q \in \Lambda_c \right\}$$

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

Particularly, we have

$$\rho(X) = \max_{Q \in \mathcal{M}_1(\mathbb{P})} \Big(\mathbb{E}_Q[-X] - \alpha^{\min}(Q) \Big), \quad X \in L^\infty, \tag{3.26}$$

whenever one of these four equivalent conditions is satisfied.

Proof. The equivalence of conditions (a) and (b) was shown in Proposition B.4. The equivalence of conditions (a) and (c) as well as (3.26) follow from Theorem 3.28 and Lemma 3.32.

The equivalence involving condition (d) is of a more technical nature, as it relies on compactness arguments in L^1 and properties of weak convergence of densities. For completeness, a detailed proof of the equivalence with (d) is provided in Appendix B.2. □

We now restrict attention to coherent risk measures. In this case, as already observed in Corollary 3.20, the penalty function takes only the values 0 and $+\infty$, and the dual representation simplifies accordingly. The following result is therefore an immediate consequence of Theorem 3.33 and Corollary 3.20.

Corollary 3.35. *A coherent risk measure on L^∞ can be represented by a set $\mathcal{Q} \subseteq \mathcal{M}_1(\mathbb{P})$ if and only if the equivalent conditions of Theorem 3.33 are satisfied. In this case, the maximal representing subset of $\mathcal{M}_1(\mathbb{P})$ is given by*

$$\mathcal{Q}_{\max} \cap \mathcal{M}_1(\mathbb{P}) = \{Q \in \mathcal{M}_1(\mathbb{P}) \mid \alpha^{\min}(Q) = 0\}.$$

Specializing to the coherent case, the continuity-from-below criteria can be made fully explicit. The following statement results from combining the preceding corollaries.

Corollary 3.36. *For a coherent risk measure ρ on L^∞, the following properties are equivalent:*

(a) ρ is continuous from below.

(b) ρ satisfies the Lebesgue property: $\rho(X_n) \to \rho(X)$ whenever (X_n) is a bounded sequence in L^∞ that converges $\mathbb{P}$-a.s. to X.

(c) $\mathcal{Q}_{\max} \subseteq \mathcal{M}_1(\mathbb{P})$

(d) The set of densities

$$\left\{ \frac{dQ}{d\mathbb{P}} \,\middle|\, Q \in \mathcal{Q}_{\max} \cap \mathcal{M}_1(\mathbb{P}) \right\}$$

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

If these equivalent conditions hold, then the representation

$$\rho(X) = \max_{Q \in \mathcal{Q}_{\max}} \mathbb{E}_Q[-X], \quad X \in L^\infty,$$

involves only σ-additive probability measures.

We now revisit the worst-case risk measure introduced in Section 3.1, which provides a simple and representative example of a coherent risk measure within the present framework.

Example 3.37 (Worst-Case Risk Measure on L^∞)**.** Under the standing assumption (3.22), the worst-case risk measure takes the form

$$\rho_{\max}(X) := -\operatorname{ess\,inf} X = \inf\{m \in \mathbb{R} \mid X + m \geq 0 \ \ \mathbb{P}\text{-a.s.}\}. \tag{3.27}$$

One can easily check that $\rho_{\max}$ is coherent and satisfies the Fatou property. Moreover, the acceptance set of $\rho_{\max}$ is equal to the positive cone L^∞_+ in L^∞, and this implies

$$\alpha^{\min}(Q) = 0 \quad \text{for any } Q \in \mathcal{M}_1(\mathbb{P}).$$

Thus,

$$\rho_{\max}(X) = \sup_{Q \in \mathcal{M}_1(\mathbb{P})} \mathbb{E}_Q[-X]. \tag{3.28}$$

The supremum in (3.28) cannot be replaced by a maximum here, for the reason identified in Theorem 3.33 and Corollary 3.34: the worst-case measure $\rho_{\max}$ fails continuity from below, so that the distinction between supremum and maximum of Subsection 3.2.2 places it on the supremum side.

The preceding example illustrates the structure of coherent risk measures in a concrete setting. We now conclude this section with a more abstract perspective, highlighting the connection between representation results and convex duality.

Remark 3.38. In analogy to Remark 3.26, the implication (e)⇒(a) in Theorem 3.33 is again an instance of the Fenchel–Legendre transform of Subsection 2.5.1, now applied to the convex risk measure ρ on L^∞. It is the special case of the Fenchel–Moreau biconjugation of Theorem 2.31 in which the structural properties of a monetary risk measure single out the probability measures among the dual variables.

Viewed through this duality lens, the restriction to L^∞ is not essential: the same mechanism yields representation results for convex risk measures on the Banach spaces $L^p(\Omega, \mathcal{F}, \mathbb{P})$ with $1 \leq p < \infty$. More precisely, let $q \in (1, \infty]$ such that $\frac{1}{p} + \frac{1}{q} = 1$, and define

$$\mathcal{M}_1^q(\mathbb{P}) := \left\{ Q \in \mathcal{M}_1(\mathbb{P}) \;\middle|\; \frac{dQ}{d\mathbb{P}} \in L^q(\mathbb{P}) \right\}.$$

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

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

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

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

The results of this chapter establish a robust dual framework for convex risk measures. This framework will serve as the foundation for the study of more refined classes of risk measures and their structural properties in the subsequent chapters.

3.4 Examples

We now turn to two fundamental examples of risk measures, namely Value at Risk (V@R) and Average Value at Risk (AV@R). These examples illustrate how the abstract concepts introduced in the previous sections translate into concrete and widely used measures of risk. Specifically they provide important insight into the role of quantiles and distributional properties, which will be central in the subsequent analysis of law-invariant and distortion-based risk measures.

3.4.1 V@R: Value at Risk

A widely used approach to quantify the risk of a financial position is based on the distribution of its future outcomes. In particular, risk is often assessed through quantiles of the distribution under the reference probability measure $\mathbb{P}$. We refer to Appendix A for the precise definition and properties of quantiles.

Definition 3.39. Fix a level $\lambda \in (0,1)$. For a financial position X, the *Value at Risk* at level λ is defined by

$$\mathrm{V@R}_\lambda(X) := -q_X^+(\lambda) = q_{-X}^-(1-\lambda) = \inf\{m \in \mathbb{R} \mid \mathbb{P}(X+m<0) \le \lambda\}.$$

The quantity $\mathrm{V@R}_\lambda(X)$ can be interpreted as the minimal amount of capital that needs to be added to the position X in order to ensure that the probability of a loss does not exceed the level λ. Thus, V@R controls the probability of large losses, but does not take into account their magnitude beyond the specified quantile.

In economic terms, $\mathrm{V@R}_\lambda(X)$ represents the additional capital a firm must hold to reduce the probability of insolvency to the level λ. A negative value indicates that the position is sufficiently safe, so that capital could be returned to shareholders or the firm could increase its exposure to risk [18].

An important feature of $\mathrm{V@R}_\lambda$ is that it is defined on the entire space L^0, so that it can be applied to arbitrary random variables without additional integrability or boundedness assumptions. The following proposition summarizes its fundamental properties [18].

Proposition 3.40. *For each $\lambda \in (0,1]$, the functional $\mathrm{V@R}_\lambda$ defines a monetary risk measure on L^0, and it is positively homogeneous.*

Proof. Monotonicity: If $X \le Y$ almost surely, then for all $m \in \mathbb{R}$,

$$\{Y+m<0\} \subseteq \{X+m<0\},$$

hence $\mathbb{P}(Y+m<0) \le \mathbb{P}(X+m<0)$. Taking the infimum yields $\mathrm{V@R}_\lambda(X) \ge \mathrm{V@R}_\lambda(Y)$.

Cash invariance: For $m \in \mathbb{R}$,

$$\mathbb{P}(X+m'+m<0) \le \lambda \iff \mathbb{P}(X+(m'+m)<0) \le \lambda,$$

which implies $\mathrm{V@R}_\lambda(X+m) = \mathrm{V@R}_\lambda(X) - m$.

Positive homogeneity: For $\alpha \ge 0$,

$$\mathbb{P}(\alpha X + m < 0) = \mathbb{P}\Big(X + \tfrac{m}{\alpha} < 0\Big),$$

so that $\mathrm{V@R}_\lambda(\alpha X) = \alpha\,\mathrm{V@R}_\lambda(X)$. □

Despite its widespread use, V@R has important shortcomings due to its high level of generality. Most notably, as already mentioned above, it fails to capture the severity of losses beyond the chosen quantile level. That is, it does not distinguish between moderate and extreme losses once they occur with probability smaller than λ. From a financial perspective, however, the magnitude of losses beyond the chosen quantile is highly relevant, as it determines the potential severity of losses in adverse scenarios. Ignoring this information may lead to an underestimation of tail risk [18].

Moreover, V@R is in general not convex. As a consequence, diversification may increase the measured risk, which contradicts the fundamental principle that diversification should reduce risk. This phenomenon is illustrated in the following example.

Example 3.41 (Lack of convexity of V@R)**.** [18] Consider a bank that has issued a single loan of \$100 with default probability 0.008. If $\lambda = 0.01$, then $\mathrm{V@R}_\lambda(X) \leq 0$.

Now consider another bank that has issued two independent loans of \$50 each, both with default probability 0.008. In this case, the probability that at least one loan defaults is given by

$$\mathbb{P}(\text{at least one default}) = 2 \cdot 0.008 \cdot 0.992 + 0.008^2 \approx 0.0159 > 0.01,$$

so that $\mathrm{V@R}_\lambda(X)$ is \$50. Splitting the capital between the two positions has thus made the portfolio appear *riskier* in terms of V@R, the opposite of what one expects from diversification. In particular, $\mathrm{V@R}_\lambda$ violates the convexity axiom and is therefore not a convex risk measure.

The example reflects a structural feature of $\mathrm{V@R}_\lambda$: the measure reacts strongly to how *likely* a loss is, but is largely insensitive to how *severe* the loss becomes once it occurs. An agent who optimizes a portfolio under a $\mathrm{V@R}_\lambda$ constraint may therefore be drawn towards concentrated positions whose default probability falls just below the threshold λ. Positions that the measure deems acceptable even though a default, however unlikely, would result in substantial losses.

We now turn to monetary risk measures on $\mathcal{X} := L^\infty$ that, in contrast to Value at Risk, are convex or even coherent, while staying as close to $\mathrm{V@R}_\lambda$ as possible. A natural first idea is to correct the lack of convexity by passing to the least conservative convex risk measure that is continuous from above and nowhere smaller than $\mathrm{V@R}_\lambda$. The following result shows that this approach fails, since no such minimal dominating convex risk measure exists.

Proposition 3.42. *For each $X \in \mathcal{X}$ and each $\lambda \in (0,1)$,*

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

Proof. Let $q := -\mathrm{V@R}_\lambda(X) = q_X^+(\lambda)$ so that $\mathbb{P}[X < q] \leq \lambda$. If $A \in \mathcal{F}$ satisfies $\mathbb{P}[A] > \lambda$,

then $\mathbb{P}[A \cap \{X \geq q\}] > 0$. Thus, we may define a measure Q_A by

$$Q_A := \mathbb{P}[\cdot \mid A \cap \{X \geq q\}].$$

It follows that $\mathbb{E}_{Q_A}[-X] \leq -q = \mathrm{V@R}_\lambda(X)$.

Let

$$\mathcal{Q} := \{Q_A \mid A \in \mathcal{F} \text{ such that } \mathbb{P}[A] > \lambda\}$$

and use this set to define a coherent risk measure ρ via

$$\rho(Y) := \sup_{Q \in \mathcal{Q}} \mathbb{E}_Q[-Y].$$

Then $\rho(X) \leq -q = \mathrm{V@R}_\lambda(X)$. Hence, the assertion will follow if we can show that $\rho(Y) \geq \mathrm{V@R}_\lambda(Y)$ for each $Y \in \mathcal{X}$. Let $\varepsilon > 0$ and

$$A := \{Y \leq -\mathrm{V@R}_\lambda(Y) + \varepsilon\} = \{Y \leq q_Y^+(\lambda) + \varepsilon\}.$$

Then $\mathbb{P}[A] > \lambda$, and so $Q_A \in \mathcal{Q}$. Moreover, $Q_A[A] = 1$, and we obtain

$$\rho(Y) \geq \mathbb{E}_{Q_A}[-Y] \geq \mathrm{V@R}_\lambda(Y) - \varepsilon.$$

Since $\varepsilon > 0$ is arbitrary, it follows that $\rho(Y) \geq \mathrm{V@R}_\lambda(Y)$, as desired. □

The preceding discussion exposes the limitation of Value at Risk from two angles. It fails to account for the magnitude of losses beyond the specified quantile, and, as the example shows, it may even penalize diversification. Moreover, the impossibility result above demonstrates that this defect cannot be repaired from within: there is no smallest convex risk measure dominating $\mathrm{V@R}_\lambda$, so a canonical convexification of Value at Risk does not exist.

The search for a suitable alternative must therefore proceed differently. Instead of dominating $\mathrm{V@R}_\lambda$ pointwise, we consider risk measures that retain its distributional character, being defined in terms of the quantiles of the loss distribution, while satisfying convexity from the outset. This leads to Average Value at Risk, which we introduce next.

3.4.2 AV@R: Average Value at Risk

A prominent alternative that overcomes the shortcomings of Value at Risk is the Average Value at Risk. This risk measure is defined in terms of the quantiles of the distribution, similarly to V@R, but incorporates information about the magnitude of losses beyond the specified level. In contrast to V@R, AV@R satisfies the axioms of a coherent risk measure and therefore provides a more robust and consistent framework for risk assessment.

Definition 3.43. The *Average Value at Risk* at level $\lambda \in (0,1]$ of a position $X \in \mathcal{X}$ is defined by

$$\mathrm{AV@R}_\lambda(X) = \frac{1}{\lambda}\int_0^\lambda \mathrm{V@R}_\gamma(X)\, d\gamma.$$

Monotonicity, cash invariance, and positive homogeneity are all preserved when averaging over the level γ, so $\mathrm{AV@R}_\lambda$ is a positively homogeneous monetary risk measure. Note that

$$\mathrm{AV@R}_\lambda(X) = -\frac{1}{\lambda}\int_0^\lambda q_X^+(t)\, dt,$$

where q_X^+ denotes the upper quantile function of X. By Definition 3.39, in particular, the definition of $\mathrm{AV@R}_\lambda(X)$ makes sense for any $X \in L^1(\Omega, \mathcal{F}, \mathbb{P})$, and we have, in view of Lemma A.19,

$$\mathrm{AV@R}_1(X) = -\int_0^1 q_X^+(t)\, dt = \mathbb{E}[-X].$$

Lemma A.19 also implies that, in the two preceding identities, q_X^+ can be replaced by any other quantile function of X.

The following remark describes the limiting case as $\lambda \downarrow 0$.

Remark 3.44. For $X \in L^\infty$, we have

$$\lim_{\lambda \downarrow 0} \mathrm{V@R}_\lambda(X) = -\operatorname{ess\,inf} X = \inf\{m \in \mathbb{R} \mid \mathbb{P}(X + m < 0) \le 0\};$$

Therefore, it makes sense to define

$$\mathrm{AV@R}_0(X) := \mathrm{V@R}_0(X) := -\operatorname{ess\,inf} X,$$

which is the worst-case risk measure on L^∞ introduced in Example 3.37. Recall that it is continuous from above but in general not from below.

The following result provides an alternative representation of $\mathrm{AV@R}_\lambda$ that will be useful for both theoretical analysis and the study of its structural properties.

Proposition 3.45. *[41, 42] For $\lambda \in (0,1)$ and any λ-quantile q of X,*

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

Proof. Let q_X be a quantile function with $q_X(\lambda) = q$. By Lemma A.19,

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

This proves the first identity. The second one follows from Lemma A.22. It also follows from Lemma A.18, since

$$\mathbb{E}[(r - X)^+] - \lambda r = \lambda(\mathbb{E}[(r - X)^+] - r) + (1 - \lambda)\mathbb{E}[(X - r)^-]$$

$$= -\lambda\mathbb{E}[X] + \lambda\mathbb{E}[(X-r)^+] + (1-\lambda)\mathbb{E}[(X-r)^-] = -\lambda\mathbb{E}[X] + f_\lambda(r),$$

where we have used the put-call parity in the second step, and f_λ is as in (A.1). □

The second identity is the representation of Average Value at Risk due to Rockafellar and Uryasev [41, 42]. By expressing $\mathrm{AV@R}_\lambda$ as the minimum over a single real parameter r of a convex objective, it turns the minimization of Average Value at Risk into a convex optimization problem. This is what makes it tractable in practice and underlies its use in portfolio optimization.

We now establish a key structural property of $\mathrm{AV@R}_\lambda$.

Lemma 3.46. *For $\lambda \in (0,1)$, the risk measure $\mathrm{AV@R}_\lambda$ is subadditive, i.e.,*

$$\mathrm{AV@R}_\lambda(X+Y) \le \mathrm{AV@R}_\lambda(X) + \mathrm{AV@R}_\lambda(Y) \quad \textit{for all } X, Y \in \mathcal{X}.$$

Proof. The proof is based on the representation in Proposition 3.45 and an application of the inequality $(a+b)^+ \le a^+ + b^+$. For completeness, a detailed argument is provided in Appendix B.2. □

Since we have already observed that $\mathrm{AV@R}_\lambda$ is a positively homogeneous monetary risk measure, Lemma 3.46 implies that $\mathrm{AV@R}_\lambda$ is a coherent risk measure.

Alternatively, coherence also follows directly from the following representation result.

Theorem 3.47. *For $\lambda \in (0,1]$, $\mathrm{AV@R}_\lambda$ is a coherent risk measure that is continuous from below. Moreover, it admits the representation*

$$\mathrm{AV@R}_\lambda(X) = \max_{Q \in \mathcal{Q}_\lambda} \mathbb{E}_Q[-X], \quad X \in \mathcal{X},$$

where $\mathcal{Q}_\lambda$ is the set of all probability measures $Q \ll \mathbb{P}$ whose density $dQ/d\mathbb{P}$ is $\mathbb{P}$-a.s. bounded by $1/\lambda$. Furthermore, $\mathcal{Q}_\lambda$ coincides with the maximal set $\mathcal{Q}_{\max}$ of Corollary 3.36.

The proof of this theorem is given in Appendix B.2.

Remark 3.48. The dual representation of $\mathrm{AV@R}_\lambda$ in Theorem 3.47 exhibits a structural pattern that will recur throughout the remainder of this thesis. On one side stands a *distributional* description, built from the quantile function of the position and thus depending only on its law; on the other stands a *robust* description, a worst case of expected loss taken over a set of probability measures. The two are not merely two formulas for the same number but they are conjugate to one another. The quantile side records the shape of the loss distribution and the measure side records the models against which it is stress-tested. We refer to this equivalence as the *dual correspondence* between the quantile and the measure representation of a risk measure. It is the same duality that, in its most abstract form, was established through the Fenchel–Legendre transform in Subsection 2.5.1, and it will reappear, in increasingly explicit guises, for general law-invariant measures and for distortion measures in the sections that follow.

3.5 Law-Invariant Risk Measures

Many commonly used risk measures, such as Value at Risk and Average Value at Risk, depend only on the distribution of a position under a given probability measure $\mathbb{P}$, as they are defined in terms of quantiles. This observation motivates the study of law-invariant risk measures, which assign the same value to positions with the same distribution. In this section, we introduce and analyze this specific class of risk measures.

3.5.1 Law-Invariance and the Fatou Property

Definition 3.49. A monetary risk measure ρ on $\mathcal{X} = L^\infty(\Omega, \mathcal{F}, \mathbb{P})$ is called *law-invariant* if

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

whenever X and Y have the same distribution under $\mathbb{P}$.

Law-invariant risk measures are also referred to as law-determined, law-based, or law-equivariant. Throughout this section, we assume that the underlying probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is sufficiently rich in the sense that it supports a random variable with a continuous distribution. This is equivalent to saying that $(\Omega, \mathcal{F}, \mathbb{P})$ is atomless.

Under this assumption, an important structural property emerges: every law-invariant convex risk measure is automatically continuous from above. Recall from Theorem 3.33 that, for convex risk measures on L^∞, continuity from above is equivalent to $\sigma(L^\infty, L^1)$-lower semicontinuity and to the Fatou property.

Theorem 3.50. *[33] Every law-invariant convex risk measure on L^∞ has the Fatou property.*

This result is due to Jouini, Schachermayer and Touzi [33]. The point of their work is that the Fatou property need not be imposed, since it is already implied by law-invariance. They established it for standard atomless probability spaces; the extension to general atomless spaces is due to Svindland [48]. Since atomlessness is assumed throughout this section, it is precisely this hypothesis that the approximation in Proposition 3.51 below requires.

The proof of Theorem 3.50 relies on the following technical approximation result, which allows one to replace a given sequence of random variables by averages of identically distributed copies. This construction is specific to atomless probability spaces and plays a key role in establishing the Fatou property for law-invariant risk measures. As the proof is rather technical and not central to our purposes, we state the result without proof and refer to the literature [26, 33] for further details.

Proposition 3.51. *On an atomless probability space, let $X_1, X_2, \dots$ be a uniformly bounded sequence of random variables that converges in probability to the random variable*

X. Then for each $k \in \mathbb{N}$ there exist $n_k \in \mathbb{N}$ and random variables $X_{k,1}, \dots, X_{k,n_k}$ such that each $X_{k,j}$ has the same law as X_k and

$$\frac{1}{n_k} \sum_{j=1}^{n_k} X_{k,j} \longrightarrow X \quad \text{in } L^\infty \text{ as } k \uparrow \infty. \tag{3.29}$$

Once this result is available, the proof of Theorem 3.50 follows by combining law-invariance with convexity.

Proof of Theorem 3.50. Let ρ be a law-invariant convex risk measure and (X_k) in L^∞ a uniformly bounded sequence that converges $\mathbb{P}$-a.s. to $X \in L^\infty$. An application of Proposition 3.51 yields $n_k \in \mathbb{N}$ and random variables $X_{k,1}, \dots, X_{k,n_k}$ with the same law as X_k such that

$$\bar{X}_k := \frac{1}{n_k} \sum_{j=1}^{n_k} X_{k,j}$$

converges to X in L^∞. Since ρ is Lipschitz continuous on L^∞ by Lemma 3.6, we have

$$\rho(X) = \lim_{k\to\infty} \rho(\bar{X}_k) \leq \liminf_{k\to\infty} \frac{1}{n_k} \sum_{j=1}^{n_k} \rho(X_{k,j}) = \liminf_{k\to\infty} \rho(X_k),$$

where we have used the convexity of ρ in the inequality and then the law-invariance of ρ. □

3.5.2 Robust Representation of Law-Invariant Risk Measures

In the previous sections, we have established that convex risk measures admit a robust representation of the form

$$\rho(X) = \sup_{Q \in \mathcal{M}_1(\mathbb{P})} \Big(\mathbb{E}_Q[-X] - \alpha(Q) \Big),$$

where α is a suitable penalty function. This representation provides a dual perspective on risk, interpreting it as the worst-case expected loss across a family of probabilistic models.

We now specialize this representation to the class of law-invariant risk measures. Under the additional assumption that the underlying probability space is atomless, law-invariance imposes a specific structure on the robust representation, which allows for a more explicit characterization. The aim is to analyze these structural properties and to derive refined representation results for law-invariant convex risk measures.

Theorem 3.52. *Let ρ be a convex risk measure. Then ρ is law-invariant if and only if it is continuous from above and its minimal penalty function $\alpha^{\min}(Q)$ depends only on*

the law of $\varphi_Q := \frac{dQ}{d\mathbb{P}}$ *under* $\mathbb{P}$ *for* $Q \in \mathcal{M}_1(\mathbb{P})$. *In this case,* ρ *has the representation*

$$\rho(X) = \sup_{Q \in \mathcal{M}_1(\mathbb{P})} \left(\int_0^1 q_{-X}(t)\, q_{\varphi_Q}(t)\, dt - \alpha^{\min}(Q) \right). \tag{3.30}$$

Moreover, the minimal penalty function satisfies

$$\alpha^{\min}(Q) = \sup_{X \in \mathcal{A}_\rho} \int_0^1 q_{-X}(t)\, q_{\varphi_Q}(t)\, dt = \sup_{X \in L^\infty} \left(\int_0^1 q_{-X}(t)\, q_{\varphi_Q}(t)\, dt - \rho(X) \right). \tag{3.31}$$

Representation (3.30) is the quantile side of the dual correspondence (Remark 3.48); its measure-side counterpart, expressing ρ as a worst case over mixtures of Average Value at Risk, is Theorem 3.54 below.

Theorem 3.52 shows that law-invariance restricts the dual representation of a risk measure in a fundamental way. While the general robust representation depends on the choice of probability measures Q, law-invariance implies that only the distribution of the Radon–Nikodym density $\frac{dQ}{d\mathbb{P}}$ matters. In other words, the risk measure does not distinguish between different models that generate the same distributional structure.

The reduction to distributional quantities is the key step that later allows for representations in terms of quantile functions and leads to the characterization of law-invariant convex risk measures as mixtures of Average Value at Risk.

The proof of this theorem relies on the following auxiliary result, which establishes a fundamental relation between quantile functions and expectations. The argument is based on rearrangement inequalities, specifically the Hardy–Littlewood inequality, and exploits the atomless structure of the underlying probability space. For completeness, a detailed proof of this auxiliary result is provided in Appendix B.2.

Proposition 3.53. *For two random variables* X *and* Y,

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

provided that all occurring integrals are well-defined and where we write $\tilde{X} \sim X$ *to indicate that the two random variables* $\tilde{X}$ *and* X *have the same law.*

Proof of Theorem 3.52. Suppose first that ρ is law-invariant. Then ρ is continuous from above by Theorem 3.50. Moreover, $X \in \mathcal{A}_\rho$ implies that $\tilde{X} \in \mathcal{A}_\rho$ for all $\tilde{X} \sim X$. Hence,

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

where we have used Proposition 3.53 in the final step. It follows that $\alpha^{\min}(Q)$ depends only on the law of φ_Q. We have thus also established the first identity in (3.31). For the second one, just note that $\tilde{X} := X + \rho(X)$ belongs to $\mathcal{A}_\rho$ for any $X \in L^\infty$ and that $q_{-X} - \rho(X)$ is a quantile function for $-\tilde{X}$.

Conversely, let us assume that $\alpha^{\min}(Q)$ depends only on the law of φ_Q. For the moment, let us write $\tilde{Q} \sim Q$ to indicate that φ_Q and $\varphi_{\tilde{Q}}$ have the same law. Then Proposition 3.53 yields

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

We thus have established (3.30), and the law-invariance of ρ follows. □

If $\lambda \in (0,1)$ and $X \in L^\infty$, then the Average Value at Risk admits the equivalent representations

$$\mathrm{AV@R}_\lambda(X) = \sup\left\{\mathbb{E}_Q[-X] \,\middle|\, Q \in \mathcal{M}_1(\mathbb{P}),\ \frac{dQ}{d\mathbb{P}} \le \frac{1}{\lambda}\ \mathbb{P}\text{-a.s.}\right\} = \frac{1}{\lambda}\int_0^\lambda \mathrm{V@R}_\gamma(X)\, d\gamma.$$

This result highlights the fundamental dual nature of $\mathrm{AV@R}_\lambda$. On the one hand, it is defined in terms of quantiles of the loss distribution, averaging over the worst λ-tail. On the other hand, it admits a representation as a worst-case expected loss over a class of probability measures with bounded densities.

This equivalence provides a direct link between a distribution-based and a model-based perspective on risk, and explains why $\mathrm{AV@R}_\lambda$ is particularly well suited for robust risk assessment. The proof of this result is given in Appendix B.1.

More generally, law-invariant convex risk measures can be represented in terms of mixtures of $\mathrm{AV@R}_\lambda$, leading to the below theorem.

Theorem 3.54 (Kusuoka-Representation)**.** *[28, 35] A convex risk measure ρ is law-invariant if and only if*

$$\rho(X) = \sup_{\mu\in\mathcal{M}_1((0,1))}\left(\int_{(0,1]} \mathrm{AV@R}_\lambda(X)\, \mu(d\lambda) - \beta^{\min}(\mu)\right), \tag{3.32}$$

where

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

This so-called Kusuoka-type mixture representation is the measure side of the dual correspondence (Remark 3.48), conjugate to the quantile-integral form of Theorem 3.52.

Proof. Clearly, the right-hand side of (3.32) defines a law-invariant convex risk measure that is continuous from above. Conversely, let ρ be law-invariant. Then ρ is continuous from above by Theorem 3.52. We will show that for $Q \in \mathcal{M}_1(\mathbb{P})$ there exists a probability

measure $\mu \in \mathcal{M}_1((0,1])$ such that

$$\int_0^1 q_{-X}(t)q_\varphi(t)\,dt = \int_{(0,1]} \mathrm{AV@R}_s(X)\,\mu(ds),$$

where $\varphi := \varphi_Q = \dfrac{dQ}{d\mathbb{P}}$. Then the assertion will follow from Theorem 3.52. Since $q_{-X}(t) = \mathrm{V@R}_{1-t}(X)$ and $q_\varphi(t) = q_\varphi^+(t)$ for a.e. $t \in (0,1)$, we obtain

$$\int_0^1 q_{-X}(t)q_\varphi(t)\,dt = \int_0^1 \mathrm{V@R}_t(X)\,q_\varphi^+(1-t)\,dt. \tag{3.33}$$

Since q_φ^+ is nonnegative, increasing, and right-continuous, we can write

$$q_\varphi^+(t) = v([1-t,1])$$

for some positive Radon measure v on $(0,1]$. Moreover, the measure μ given by

$$\mu(dt) = t\,v(dt)$$

is a probability measure on $(0,1]$. Indeed, applying Tonelli leads to:

$$\int_{(0,1]} \mu(dt) = \int_{(0,1]} t\,v(dt) = \int_{(0,1]}\int_0^1 \mathbf{1}_{\{s\le t\}}\,ds\,v(dt) = \int_0^1\int_{(0,1]} \mathbf{1}_{\{s\le t\}}\,v(dt)\,ds$$

$$= \int_0^1 v([s,1])\,ds = \int_0^1 q_\varphi^+(s)\,ds = \mathbb{E}[\varphi] = 1.$$

Thus, from (3.33),

$$\int_0^1 q_{-X}(t)q_\varphi(t)\,dt = \int_0^1 \mathrm{V@R}_t(X)\left(\int_{(t,1]} \frac{1}{s}\,\mu(ds)\right)dt$$

$$= \int_{(0,1]} \left(\frac{1}{s}\int_0^s \mathrm{V@R}_t(X)\,dt\right)\mu(ds) = \int_{(0,1]} \mathrm{AV@R}_s(X)\,\mu(ds). \tag{3.34}$$

Conversely, for any probability measure μ on $(0,1]$, the function q defined by

$$q(t) := \int_{(1-t,1]} \frac{1}{s}\,\mu(ds)$$

can be viewed as a quantile function of the density $\varphi := q(U)$ of a measure $Q \in \mathcal{M}_1(\mathbb{P})$, where U has a uniform distribution on $(0,1)$. Indeed, φ is nonnegative, and

$$\mathbb{E}[\varphi] = \int_0^1 q(t)\,dt = \int_0^1 \int_{(1-t,1]} \frac{1}{s}\,\mu(ds)\,dt = \int_{(0,1]}\int_0^1 \mathbf{1}_{\{t>1-s\}}\,dt\,\frac{1}{s}\,\mu(ds) = \mu((0,1]) = 1.$$

Altogether, we obtain a one-to-one correspondence between laws of densities φ and probability measures μ on $(0,1]$. □

The representation of Theorem 3.54 is in its original form due to Kusuoka [35], who established it for law-invariant *coherent* risk measures. The extension to the *convex* case stated here is due to Frittelli and Rosazza Gianin [28].

The Kusuoka representation is not unique. The definitive analysis of this question is due to Shapiro [47]. The integral representation $\rho(X) = \int_{(0,1]} \mathrm{AV@R}_\lambda(X)\, \mu(\mathrm{d}\lambda)$ is unique precisely when the dual set is generated by a single one of its elements, whereas the maximum-over-integrals form of Theorem 3.54 is in general not unique. Theorem 3.54 therefore leaves open exactly this latter, generic kind of non-uniqueness.

For coherent risk measures the Kusuoka representation takes the following form.

Corollary 3.55. *A coherent risk measure ρ is law-invariant if and only if*

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

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

Corollary 3.55 follows immediately from Theorem 3.54 by noting that for coherent risk measures the minimal penalty function reduces to the indicator function of a suitable set of probability measures, as already seen in Corollary 3.20.

The results of this section show that law-invariance imposes a strong structural restriction on convex risk measures, reducing their dependence to distributional characteristics of the underlying positions. The robust representation can be expressed in terms of quantile functions, thereby shifting the focus from probability measures to distributional properties.

This perspective provides the key link to the subsequent chapters. In Chapter 4, we will study distortion risk measures, which are constructed explicitly from distribution functions and can be interpreted as weighted averages of quantiles. In Chapter 5, we will turn to comonotone risk measures, where the role of quantiles and their interaction becomes even more pronounced. Ultimately, these developments lead to representation results that connect law-invariant convex risk measures with mixtures of Average Value at Risk.

Chapter 4

Distortion Risk Measures

In the previous chapters, we introduced convex and coherent risk measures and studied their dual representations. In particular, we identified law-invariance as a natural and important property, ensuring that the risk of a financial position depends only on its distribution. We now refine this perspective by focusing on a class of law-invariant convex risk measures that admits a particularly explicit and tractable representation.

More precisely, we are interested in coherent risk measures that can be written in the form

$$\rho_\mu(X) := \int \mathrm{AV@R}_\lambda(X)\, \mu(d\lambda),$$

for some probability measure μ on $\mathcal{M}_1((0,1])$. This representation expresses the risk of a position as a weighted average of Average Value at Risk at different confidence levels. It provides a flexible way to model risk preferences by assigning different importance to various parts of the loss distribution, in particular to extreme losses.

The aim of this chapter is to characterize this class of risk measures in terms of transformations of probabilities. To this end, we replace the classical linear expectation by a nonlinear functional that distorts probabilities in a systematic way. This leads to a generalized notion of integration, which will be introduced in the sequel, and which allows us to evaluate random variables with respect to suitably transformed set functions.

One of the main results of this chapter shows that these distorted evaluations are equivalent to the representation given above. In this way, distortion risk measures provide a natural link between probability distortions and mixtures of Average Value at Risk, thereby complementing the representation results developed earlier.

4.1 Distortion Functions and the Choquet Integral

In this section, we present the fundamental building blocks for the construction of distortion risk measures. We define distortion functions and the associated set functions that replace classical probability measures. This naturally leads to a generalized notion

of integration, namely the Choquet integral, which will serve as the key tool for the representation results developed later in this chapter.

The standard reference for the Choquet integral and the theory of non-additive measures is the monograph of Denneberg [20], to which we refer for the results of this section and for the lemma-level material used in the present and the following chapter; our notation and terminology, however, follow Föllmer and Schied [26].

We begin by introducing the class of functions that will be used to transform probabilities and thereby define distortion risk measures. These functions provide a flexible way to model deviations from the classical linear evaluation of risk.

Definition 4.1. Let $\psi : [0,1] \to [0,1]$ be an increasing function such that $\psi(0) = 0$ and $\psi(1) = 1$. The set function

$$c_\psi(A) := \psi(\mathbb{P}[A]), \quad A \in \mathcal{F},$$

is called the *distortion* of the probability measure $\mathbb{P}$ with respect to the distortion function ψ.

The function c_ψ is, in general, no longer a probability measure, since additivity is not preserved. Instead, it belongs to a larger class of set functions, which we introduce next.

Definition 4.2. A set function $c : \mathcal{F} \to [0,1]$ is called *monotone* if

$$c(A) \leq c(B) \quad \text{for } A \subseteq B,$$

and *normalized* if

$$c(\emptyset) = 0 \quad \text{and} \quad c(\Omega) = 1.$$

A monotone set function is called *submodular* or *strongly subadditive* if

$$c(A \cup B) + c(A \cap B) \leq c(A) + c(B).$$

Submodularity reflects a form of diminishing marginal contributions: the incremental impact of enlarging a set decreases as the set becomes larger. In an economic interpretation, this property is closely related to diversification effects, as the joint contribution of two sets is smaller than the sum of their individual contributions. This feature will play an important role in the context of risk aggregation and the Choquet integral [20].

It is easy to see that any distortion c_ψ is normalized and monotone. Such monotone and normalized set functions are commonly referred to as *capacities* in the literature. More generally, a capacity is understood as a set function that is monotone, but not necessarily additive. Capacities play a fundamental role in the definition of the Choquet integral [20, 29].

Before turning to the notion of the Choquet integral, we briefly establish an important connection between the shape of the distortion function ψ and a structural property of the induced set function c_ψ.

Proposition 4.3. *Let c_ψ be the distortion of $\mathbb{P}$ with respect to the distortion function ψ. Then ψ is concave if and only if c_ψ is submodular.*

Proof. Suppose first that ψ is concave. Take $A, B \in \mathcal{F}$ with $\mathbb{P}[A] \leq \mathbb{P}[B]$. We must show that $c := c_\psi$ satisfies

$$c(A) - c(A \cap B) \geq c(A \cup B) - c(B).$$

This is trivial if $r = 0$, where

$$r := \mathbb{P}[A] - \mathbb{P}[A \cap B] = \mathbb{P}[A \cup B] - \mathbb{P}[B].$$

For $r > 0$ the concavity of ψ yields via Proposition 2.9 that

$$\frac{c(A) - c(A \cap B)}{\mathbb{P}[A] - \mathbb{P}[A \cap B]} \geq \frac{c(A \cup B) - c(B)}{\mathbb{P}[A \cup B] - \mathbb{P}[B]}.$$

Multiplying both sides with r gives the result.

Now suppose that $c = c_\psi$ is submodular and assume that $(\Omega, \mathcal{F}, \mathbb{P})$ is atomless. We have to show that

$$\psi(y) \geq \frac{\psi(x) + \psi(z)}{2}$$

whenever $0 \leq x \leq z \leq 1$ and

$$y = \frac{x + z}{2}.$$

To this end, we will construct two sets $A, B \in \mathcal{F}$ such that

$$\mathbb{P}[A] = \mathbb{P}[B] = y, \qquad \mathbb{P}[A \cap B] = x, \qquad \mathbb{P}[A \cup B] = z.$$

Submodularity then gives

$$\psi(x) + \psi(z) \leq 2\psi(y),$$

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 by Proposition A.23. Then

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

are as desired. □

This result will play a crucial role in the sequel, as it links concavity of the distortion function to a form of subadditivity of the induced set function, which will later translate

into convexity properties of the associated risk measures.

In order to construct risk measures based on distorted probabilities, we require a notion of integration that extends the classical expectation beyond the setting of probability measures. Since distorted probabilities are in general not σ-additive, the standard Lebesgue integral is no longer applicable.

The Choquet integral provides such an extension by allowing integration with respect to capacities. In contrast to the classical expectation, which is linear, the Choquet integral captures nonlinear aggregation effects arising from non-additivity, such as ambiguity or interaction between events. This makes it a fundamental tool in the theory of risk measures and decision under uncertainty [29].

Definition 4.4. [20] Let $c : \mathcal{F} \to [0,1]$ be a set function that is normalized and monotone. The *Choquet integral* of a bounded measurable function X on $(\Omega, \mathcal{F})$ with respect to c is defined as

$$\int X \, dc := \int_{-\infty}^{0} \Big(c(X > x) - 1 \Big) \, dx + \int_{0}^{\infty} c(X > x) \, dx.$$

The following example illustrates how the Choquet integral can be constructed by first considering nonnegative random variables and then extending the definition to the general case.

Example 4.5 (The Choquet Integral and its Induced Risk Measure)**.** Let $c : \mathcal{F} \to [0,1]$ be any set function that is normalized and monotone. For instance, one may consider a distorted probability of the form

$$c(A) := \psi(\mathbb{P}[A]),$$

where $\mathbb{P}$ is a probability measure and $\psi : [0,1] \to [0,1]$ is increasing with $\psi(0) = 0$ and $\psi(1) = 1$.

For a bounded measurable function $X \geq 0$, the Choquet integral of X with respect to c is defined by

$$\int X \, dc := \int_{0}^{\infty} c(X > x) \, dx.$$

If c is a probability measure, then the Choquet integral coincides with the usual integral. In general, however, the Choquet integral is a nonlinear functional of X. Nevertheless, for constants $\lambda, m \geq 0$, it still satisfies

$$\int \lambda X \, dc = \lambda \int X \, dc \qquad \text{and} \qquad \int (X + m) \, dc = \int X \, dc + m.$$

If X is an arbitrary bounded measurable function, choose $m \in \mathbb{R}$ such that $X + m \geq 0$. Then

$$\int (X + m) \, dc - m = \int_{-m}^{0} \Big(c(X > x) - 1 \Big) \, dx + \int_{0}^{\infty} c(X > x) \, dx.$$

Since the right-hand side does not depend on the particular choice of $m \geq -\inf X$, this motivates the extension of the Choquet integral to arbitrary bounded measurable functions by setting

$$\int X\,dc := \int_{-\infty}^{0} \Big(c(X > x) - 1\Big)\,dx + \int_{0}^{\infty} c(X > x)\,dx.$$

for all $X \in \mathcal{X}$. It follows that

$$\int \lambda X\,dc = \lambda \int X\,dc \quad \text{and} \quad \int (X + m)\,dc = \int X\,dc + m$$

for all $\lambda \geq 0$ and $m \in \mathbb{R}$. Moreover, we have

$$\int Y\,dc \geq \int X\,dc \quad \text{for } Y \geq X.$$

Thus, the Choquet integral of the loss

$$\rho(X) := \int (-X)\,dc,$$

is a positively homogeneous monetary risk measure on $\mathcal{X}$. In Chapter 5, we will characterize these risk measures in terms of a property called "comonotonicity." We will also show that ρ is convex, and hence coherent, if and only if c is submodular or strongly subadditive, that is,

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

In this case, ρ admits the representation

$$\rho(X) = \max_{Q \in \mathcal{Q}_c} \mathbb{E}_Q[-X],$$

where $\mathcal{Q}_c$ is the core of c, defined as the class of all finitely additive and normalized set functions $Q : \mathcal{F} \to [0, 1]$ such that $Q[A] \leq c(A)$ for all $A \in \mathcal{F}$; see Theorem 5.17.

Remark 4.6. The Choquet integral of a simple random variable

$$X = \sum_{i=1}^{n} x_i \mathbf{1}_{A_i}, \qquad \text{with } x_1 \geq \cdots \geq x_n \geq x_{n+1} := 0,$$

and disjoint sets $A_1, \ldots, A_n$ can be computed as

$$\int X\,dc = \sum_{i=1}^{n} (x_i - x_{i+1})\,c(B_i) = \sum_{i=1}^{n} x_i \Big(c(B_i) - c(B_{i-1})\Big),$$

where $B_0 := \emptyset$ and

$$B_i := \bigcup_{k=1}^{i} A_k, \qquad i = 1, \ldots, n.$$

This representation shows that the Choquet integral aggregates the values of X in a stepwise manner. Each level x_i is weighted by the increment $c(B_i) - c(B_{i-1})$, which can be interpreted as the additional contribution of the set A_i given the previously accumulated sets.

An equivalent interpretation can be obtained from a geometric perspective. For nonnegative simple functions, the Choquet integral corresponds to the area under a staircase function associated with the capacity c, where the heights are given by the values of X and the widths by the increments of the capacity. This highlights the nonlinear nature of the Choquet integral, in contrast to the classical Lebesgue integral, where the weights are determined by a probability measure [29].

4.2 Choquet Representation

This section is devoted to deriving a Choquet integral representation for coherent risk measures of the form

$$\rho_\mu(X) := \int \mathrm{AV@R}_\lambda(X)\, \mu(d\lambda).$$

More precisely, we show that ρ_μ can be expressed as the Choquet integral of the loss $-X$ with respect to a suitable capacity.

Throughout this section, we assume that the underlying probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is atomless. Coherence, continuity from below, and law-invariance are stable under mixing: as each $\mathrm{AV@R}_\lambda$ enjoys these properties, so does every ρ_μ.

According to Remark 3.44, we may define

$$\mathrm{AV@R}_0(X) := -\operatorname{ess\,inf} X,$$

which allows us to extend the definition of ρ_μ to probability measures μ on the closed interval $[0, 1]$. However, if $\mu(\{0\}) > 0$, then ρ_μ is in general only continuous from above and not from below, since $\mathrm{AV@R}_0$ fails to be continuous from below.

The following lemma establishes the correspondence between probability measures on $[0, 1]$ and concave distortion functions, which will be crucial for deriving the Choquet integral representation of the risk measures ρ_μ in the subsequent theorem.

Recall that every concave function $\psi : [0, 1] \to [0, 1]$ admits a right-continuous right-hand derivative ψ'_+ and satisfies

$$\psi(t) = \psi(0+) + \int_0^t \psi'_+(s)\, ds, \quad t \in (0, 1].$$

Lemma 4.7. *The identity*

$$\psi'_+(t) = \int_{(t,1]} s^{-1}\, \mu(ds), \quad 0 < t < 1,$$

defines a one-to-one correspondence between probability measures μ on $[0,1]$ and increasing concave functions $\psi : [0,1] \to [0,1]$ with $\psi(0) = 0$ and $\psi(1) = 1$. Moreover, we have $\psi(0+) = \mu(\{0\})$.

Proof. Suppose first that μ is given and that ψ is defined by $\psi(1) = 1$ and

$$\psi'(t) = \int_{(t,1]} s^{-1}\, \mu(ds), \qquad 0 < t < 1.$$

Then ψ is increasing and concave on $(0,1]$. Moreover,

$$1 - \psi(0+) = \int_0^1 \psi'(t)\, dt = \int_0^1 \int_{(t,1]} s^{-1}\, \mu(ds)\, dt.$$

By Fubini's theorem,

$$\int_0^1 \int_{(t,1]} s^{-1}\, \mu(ds)\, dt = \int_{(0,1]} \frac{1}{s} \int_0^1 \mathbf{1}_{\{t \le s\}}\, dt\, \mu(ds) = \int_{(0,1]} \frac{1}{s}\, s\, \mu(ds) = \mu((0,1]) \le 1.$$

Hence, we may set $\psi(0) := 0$ and obtain an increasing concave function on $[0,1]$. Furthermore, $\psi(0+) = 1 - \mu((0,1]) = \mu(\{0\})$.

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

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

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

Theorem 4.8. *Let μ be a probability measure on $[0,1]$, and let ψ be the concave function associated with μ via Lemma 4.7. Then, for all $X \in \mathcal{X}$,*

$$\rho_\mu(-X) = \psi(0+)\, \mathrm{AV@R}_0(-X) + \int_0^1 q_X(t)\, \psi'(1-t)\, dt,$$

where q_X denotes the quantile function of X.

In particular, recalling the definition of the Choquet integral from the previous section, this representation shows that ρ_μ coincides with the Choquet integral of the loss $-X$ with respect to the capacity

$$c_\psi(A) := \psi(\mathbb{P}[A]), \qquad A \in \mathcal{F},$$

that is,

$$\rho_\mu(-X) = \int_{-\infty}^0 \big(\psi(\mathbb{P}[X > x]) - 1\big)\, dx + \int_0^\infty \psi(\mathbb{P}[X > x])\, dx = \int (-X)\, dc_\psi.$$

Proof. Using the fact that $\mathrm{V@R}_\lambda(-X) = q_X^-(1-\lambda)$, we get as in (3.34) that

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

In consequence, we obtain the first identity. For the second one, we assume first $X \geq 0$. Then

$$q_X^+(t) = \sup\{x \geq 0 \mid F_X(x) \leq t\} = \int_0^\infty \mathbf{1}_{\{F_X(x)\leq t\}}\,dx,$$

where F_X is the cumulative distribution function of X. Using Fubini's theorem and the fact that

$$\int_0^y \psi'(t)\,dt = (\psi(y) - \psi(0+))\,\mathbf{1}_{\{y>0\}},$$

we obtain

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

This proves the second identity for $X \geq 0$, since $\psi(0+) = \mu(\{0\})$ and $\operatorname{ess\,sup} X = \mathrm{AV@R}_0(-X)$. If $X \in L^\infty$ is arbitrary, we consider $X + C$, where $C := -\operatorname{ess\,inf} X$. The cash invariance of ρ_μ yields

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

□

This representation reveals that distortion risk measures arise naturally as Choquet integrals, where the distortion function ψ modifies the underlying probability measure and thereby captures attitudes towards risk. This allows us to identify classical risk measures as Choquet integrals corresponding to specific choices of the underlying measure μ.

Example 4.9 (AV@R as a Choquet integral)**.** [9, 26] Clearly, the risk measure $\mathrm{AV@R}_\lambda$ is itself of the form ρ_μ for a suitable choice of μ. More precisely, let $\mu = \delta_\lambda$ be the Dirac measure at λ, defined by

$$\delta_\lambda(A) := \begin{cases} 1, & \text{if } \lambda \in A, \\ 0, & \text{otherwise,} \end{cases} \qquad A \subset [0,1].$$

Then, by definition of ρ_μ, we have

$$\rho_\mu(X) = \int_{(0,1]} \mathrm{AV@R}_\alpha(X)\, \delta_\lambda(d\alpha) = \mathrm{AV@R}_\lambda(X),$$

so that $\mathrm{AV@R}_\lambda$ is a special case of the class of risk measures ρ_μ.

For $\lambda > 0$, the corresponding concave distortion function ψ is given by

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

Inserting this distortion into the equation from Theorem 4.8, we obtain another representation of $\mathrm{AV@R}_\lambda$

$$\mathrm{AV@R}_\lambda(-X) = \frac{1}{\lambda}\int_0^\infty \Big(\mathbb{P}[X > x] \wedge \lambda\Big)\, dx, \qquad X \in L^\infty_+.$$

The above Choquet representation admits a particularly transparent form in the case where the measure μ does not charge the point 0, which can be seen in the below corollary. In this situation, the representation can be expressed directly in terms of the quantile function, which can be particularly useful in applications.

Corollary 4.10. *If $\mu(\{0\}) = 0$ in Theorem 4.8, then*

$$\rho_\mu(X) = -\int_0^1 q_X(\varphi(t))\, dt,$$

where φ is an inverse function of ψ, taken in the sense of Definition A.12.

Proof. Due to Lemma A.13, the distribution of φ under the Lebesgue measure is the concave distortion function ψ. By Proposition 2.9(d), we have $\psi(t) = \psi(0) + \int_0^t \psi'(s)\, ds$, $t \in [0,1]$. Hence, the law of φ under the Lebesgue measure has density ψ'. Therefore,

$$\int_0^1 q_X(\varphi(t))\, dt = \int_0^1 q_X(t)\psi'(t)\, dt = -\int_0^1 q_{-X}(1-t)\psi'(t)\, dt,$$

where we have used Lemma A.21 in the last step. An application of Theorem 4.8 concludes the proof. □

This representation highlights that, in the absence of an atom at 0, the risk measure ρ_μ can be interpreted as a weighted average of quantiles of X, where the transformation φ distorts the probability levels. The distortion acts directly on the quantile scale, providing an alternative perspective to the Choquet integral representation in terms of tail probabilities.

The previous section highlights the central role of concave functions ψ in the construction of the risk measures ρ_μ. We now study these functions in more detail and interpret their effect on risk.

4.3 Concave Distortions

In this part of the thesis, we take a closer look at the class of risk measures introduced in the previous section through their Choquet integral representation. We want to analyse how the concave function ψ modifies the underlying probability measure and thereby shapes the assessment of risk.

As shown in Section 4.2, any risk measure of the form

$$\rho_\mu(X) := \int \mathrm{AV@R}_\lambda(X)\, \mu(d\lambda)$$

can be represented as a Choquet integral of the loss $-X$ with respect to the capacity $c_\psi(A) := \psi(\mathbb{P}[A])$. This representation reveals that the function ψ acts as a distortion of the probability measure $\mathbb{P}$, assigning different weights to different regions of the distribution.

In contrast to the classical expectation, where probabilities enter linearly, the use of a concave distortion function ψ leads to an overweighting of extreme outcomes, reflecting a preference for robustness with respect to adverse scenarios.

The coherent risk measures studied here are, under another name, the *spectral risk measures* introduced by Acerbi [1]. The overweighting of extreme outcomes described informally above is precisely what Acerbi formalises through a *risk-aversion spectrum*: a weight function attached to the quantiles of the loss, with larger weight placed on the more adverse ones. In the notation of Theorem 4.8, this spectrum is the weight $\psi'(1-\cdot)$ multiplying the quantile q_{-X}, that is, the derivative of the concave distortion function ψ. Acerbi's coherence condition that the spectrum be *decreasing* is, in this language, exactly the concavity of ψ established in Proposition 4.3. Acerbi denotes this spectrum by a function he calls the risk spectrum. We caution that this is not the Radon–Nikodym density $\varphi = \mathrm{d}Q/\mathrm{d}\mathbb{P}$ for which the symbol φ is used in Section 3.5, although the two are closely related: combining the representation (3.30) and Theorem 4.8, Acerbi's spectrum is the quantile function q_φ of that density.

In this chapter we will study distortion functions more closely and show that they provide an equivalent characterization of law-invariant coherent risk measures. In particular, we will see that distortion risk measures admit alternative representations in terms of quantiles and can be interpreted within the framework of decision theory.

Corollary 4.11. *Let μ be a probability measure on $[0,1]$, let ψ be the concave distortion function associated with μ as in Lemma 4.7, and define the set function*

$$c_\psi(A) := \psi(\mathbb{P}[A]), \qquad A \in \mathcal{F},$$

as the distortion of $\mathbb{P}$ with respect to ψ. Then, for all $X \in L^\infty$,

$$\rho_\mu(X) = \int (-X)\, dc_\psi.$$

Proof. The result follows directly from the Choquet integral representation established in Theorem 4.8 together with the definition of the Choquet integral. □

Combining Corollary 4.11 with the representation result in Theorem 4.8, we obtain the following characterization of law-invariant convex risk measures in terms of concave distortions.

Corollary 4.12. *A convex risk measure ρ is law-invariant and continuous from above if and only if it admits a representation of the form*

$$\rho(X) = \sup_{\psi} \left(\int (-X)\, dc_\psi - \gamma^{\min}(\psi) \right),$$

where the supremum is taken over the class of all concave distortion functions ψ, and

$$\gamma^{\min}(\psi) := \sup_{X \in \mathcal{A}_\rho} \int (-X)\, dc_\psi.$$

For completeness, a related identity for the Choquet integral involving dual set functions is provided in Appendix B.1.

Theorem 4.8 has a second payoff. It allows us to describe explicitly the maximal set $\mathcal{Q}_\mu \subseteq \mathcal{M}_1(\mathbb{P})$ representing the risk measure ρ_μ.

Theorem 4.13. *Let μ be a probability measure on $[0,1]$, and let ψ be the corresponding concave function defined in Lemma 4.7. Then the risk measure ρ_μ admits the representation*

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

where the set $\mathcal{Q}_\mu$ is given by

$$\mathcal{Q}_\mu := \left\{ Q \in \mathcal{M}_1(\mathbb{P}) \;\middle|\; Z := \frac{dQ}{d\mathbb{P}} \text{ satisfies } \int_t^1 q_Z(s)\, ds \le \psi(1-t) \text{ for } t \in (0,1) \right\}.$$

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

This core representation is the measure side of the dual correspondence (Remark 3.48): it is conjugate to the Choquet-integral description of ρ_μ in Theorem 4.8 and Corollary 4.11, which form its quantile, distribution-based side.

Proof. Since $\mathrm{AV@R}_\lambda$ is coherent, continuous from below, and law-invariant for every $\lambda \in (0,1]$, these properties are inherited by the mixture ρ_μ. In particular, ρ_μ is a coherent risk measure that is continuous from above. Hence, by Corollary 3.35, it admits a representation

$$\rho_\mu(X) = \sup_{Q \in \mathcal{Q}_{\max}} \mathbb{E}_Q[-X],$$

where $\mathcal{Q}_{\max} = \{Q \in \mathcal{M}_1(\mathbb{P}) \mid \alpha^{\min}(Q) = 0\}$. Using (3.31) and Theorem 4.8, we obtain that a measure $Q \in \mathcal{M}_1(\mathbb{P})$ with density $Z = dQ/d\mathbb{P}$ belongs to $\mathcal{Q}_{\max}$ if and only if

$$\int_0^1 q_X(s) q_Z(s)\, ds \le \rho_\mu(-X)$$

for all $X \in L^\infty$, where

$$\rho_\mu(-X) = \psi(0+)\,\mathrm{AV@R}_0(-X) + \int_0^1 q_X(s)\psi'(1-s)\, ds. \tag{4.1}$$

To show that $\mathcal{Q}_{\max} \subseteq \mathcal{Q}_\mu$, let X be Bernoulli distributed with parameter $1-t$, i.e., the two values 0 and 1 and $\mathbb{P}[X=1] = 1-t$. Then $q_X = \mathbf{1}_{[t,1]}$ a.e., and hence

$$\int_t^1 q_Z(s)\, ds \le \psi(0+) + \int_t^1 \psi'(1-s)\, ds = \psi(1-t),$$

which shows that first inclusion.

For the proof of the converse inclusion, we show that the density Z of a fixed measure $Q \in \mathcal{Q}_\mu$ satisfies (4.1) for any given $X \in L^\infty$. We may assume without loss of generality that $X \ge 0$. Let ν be the positive finite measure on $[0,1]$ such that $q_X^+(s) = \nu([0,s])$.Then, in particular,

$$\nu((0,1]) = q_X(1-) = \operatorname{ess\,sup} X = \mathrm{AV@R}_0(-X).$$

Using Fubini's theorem and the definition of $\mathcal{Q}_\mu$, we get

$$\begin{aligned}
\int_0^1 q_X(s)\, q_Z(s)\, ds &= \int_0^1 q_Z(s) \int_{[0,1]} \mathbf{1}_{\{t\le s\}}\, \nu(dt)\, ds = \int_{[0,1]} \int_t^1 q_Z(s)\, ds\, \nu(dt) \\
&\le \int_{[0,1]} \psi(1-t)\, \nu(dt) = \int_{[0,1]} \left(\psi(0+) + \int_0^{1-t} \psi'(s)\, ds\right) \nu(dt) \\
&= \psi(0+)\nu((0,1]) + \int_{[0,1]} \int_0^1 \mathbf{1}_{\{s\le 1-t\}}\, \psi'(s)\, ds\, \nu(dt) \\
&= \psi'(0+)\,\mathrm{AV@R}_0(-X) + \int_0^1 \psi'(1-s) \int_{[0,s]} \nu(dt)\, ds,
\end{aligned}$$

which coincides with the right-hand side of (4.1). □

Corollary 4.14. *In the context of Theorem 4.13, the following conditions are equivalent:*

(a) ρ_μ *is continuous from below.*

(b) $\mu(\{0\}) = 0$.

Proof of Corollary 4.14. If (b) holds, then (a) follows from Theorem 3.47 and monotone convergence. For the proof of the converse implication, we suppose by way of contradiction that $\delta := \mu(\{0\}) > 0$. In this case, we can write

$$\rho_\mu = \delta\,\mathrm{AV@R}_0 + (1-\delta)\rho_{\mu'},$$

where $\mu' := \mu(\cdot|(0,1])$. Then $\rho_{\mu'}$ is continuous from below since $\mu'(\{0\}) = 0$, but on our atomless probability space, $\mathrm{AV@R}_0$ is not, and so ρ_μ does not satisfy (a); see Remark 3.44. □

This result provides a precise characterization of continuity from below in terms of the representing measure. In particular, it shows that the presence of an atom at zero leads to a singular $\mathrm{AV@R}_0$ component, which violates this regularity property. This is the same mechanism as in Example 3.37: the atom at zero breaks continuity from below, and so, by the supremum-versus-maximum distinction of Theorem 3.33 and Corollary 3.34, the supremum representing ρ_μ in Theorem 4.13 is not in general attained as a maximum.

In this chapter, we have focused on the representation of law-invariant risk measures in terms of distortion functions and Choquet integrals. Starting from the introduction of distortion functions, we have shown that the associated risk measures can be expressed as Choquet integrals with respect to the distribution of the underlying position.

This representation provides a direct link between risk measures and the distributional properties of random variables, emphasizing that law-invariant risk measures depend only on the distribution function F_X. Moreover, we have seen that concave distortion functions play a central role, as they give rise to coherent and convex risk measures and admit a representation in terms of mixtures of Average Value at Risk.

The dual representation further characterizes these risk measures in terms of suitable sets of probability measures, where the admissible densities are constrained via stochastic order relations. In particular, this highlights how distortion functions control the allocation of mass to adverse outcomes.

In the next chapter, we turn to the class of comonotone additive risk measures. The aim will be to characterize these risk measures and to show that comonotone additivity is intimately linked to representations in terms of Choquet integrals, thereby providing a deeper structural understanding of distortion-based risk measures.

Chapter 5

Comonotone Additive Risk Measures

The previous chapter introduced distortion risk measures via the Choquet integral and established their fundamental properties. In particular, we have seen that concave distortion functions give rise to convex and law-invariant risk measures, together with a representation in terms of quantile functions.

The aim of this chapter is to further investigate the structure of such risk measures. In particular, we identify a key structural property that allows for a complete characterization of distortion risk measures within the class of convex risk measures.

The chapter is organized as follows. In Section 5.1, we introduce the notion of comonotonicity and discuss its relevance in the context of risk measurement. Section 5.2 recalls the Choquet integral representation and its formulation in terms of quantile functions. Finally, in Section 5.3, we present the main characterization result, which shows that distortion risk measures induced by concave distortion functions coincide with the class of law-invariant convex risk measures satisfying this structural property.

5.1 Motivation and Comonotonicity

Before introducing the formal definition, let us briefly discuss the intuition behind comonotonicity. In many situations, the risk of a combined position $X + Y$ is strictly smaller than the sum of the individual risks, reflecting the effect of diversification. This typically occurs when one position partially offsets adverse outcomes of the other, that is, when some form of hedging is present.

Comonotonicity describes the opposite situation. Two positions are said to be comonotone if their outcomes move in the same direction, so that no such hedging effect is possible. In this case, combining the positions does not reduce risk, and it is therefore natural to expect that the total risk simply equals the sum of the individual risks. It should be stressed that comonotonicity is a property of the *coupling* of X and Y, that

is, of their joint law, and not of their correlation: it is the pathwise, order-theoretic requirement that the two positions be increasing functions of a common source of randomness, a condition strictly stronger than, and not to be confused with, positive correlation.

This observation motivates the following definition, which formalizes the idea of perfectly positively dependent risks. For the whole of this chapter, $\mathcal{X}$ again stands for the space of bounded measurable functions on a measurable space $(\Omega, \mathcal{F})$, as in Chapter 3.

Definition 5.1. Two measurable functions X and Y on $(\Omega, \mathcal{F})$ are called *comonotone* if

$$\Big(X(\omega) - X(\omega')\Big)\Big(Y(\omega) - Y(\omega')\Big) \geq 0$$

for all $(\omega, \omega') \in \Omega \times \Omega$.

Definition 5.2. A monetary risk measure ρ on $\mathcal{X}$ is called *comonotonic* if

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

whenever $X, Y \in \mathcal{X}$ are comonotone.

The defining feature of comonotonic risk measures is the additivity property on comonotone positions. While convex risk measures are generally subadditive, reflecting diversification effects, this inequality becomes an equality in the absence of hedging possibilities. Thus, comonotonicity identifies a class of positions for which the risk measure behaves linearly.

This property is pivotal in the subsequent characterization of distortion risk measures.

Comonotonicity can be illustrated by simple examples.

Example 5.3 (Comonotonicity)**.** Let $A \in \mathcal{F}$ be an event and consider the random variables

$$X := \mathbf{1}_A, \qquad Y := 2\,\mathbf{1}_A.$$

Then X and Y are comonotone, since both variables increase and decrease simultaneously depending on whether the event A occurs or not.

In contrast, if we define $Y := \mathbf{1}_{A^c}$, then X and Y are not comonotone. Indeed, an increase in X corresponds to a decrease in Y, and there exist $\omega \in A$ and $\omega' \in A^c$ such that

$$(X(\omega) - X(\omega'))(Y(\omega) - Y(\omega')) < 0.$$

We begin by noting a first immediate consequence of comonotonicity for risk measures. The following lemma shows that comonotonicity already implies a basic structural property.

Lemma 5.4. *If ρ is a comonotonic monetary risk measure on $\mathcal{X}$, then ρ is positively homogeneous.*

Proof. Note that (X, X) is a comonotone pair. Hence, $\rho(2X) = \rho(X) + \rho(X) = 2\rho(X)$. 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 3.6. □

To gain a deeper understanding of comonotone random variables themselves, we next provide an equivalent characterization that reveals their underlying structure.

Lemma 5.5. *For two measurable functions X and Y on $(\Omega, \mathcal{F})$, the following conditions are equivalent:*

(a) X and Y are comonotone.

(b) There exists a measurable function Z on $(\Omega, \mathcal{F})$ and increasing functions f and g on $\mathbb{R}$ such that

$$X = f(Z) \quad \text{and} \quad Y = g(Z).$$

(c) There exist increasing and Lipschitz continuous functions f and g on $\mathbb{R}$ such that

$$X = f(X + Y) \quad \text{and} \quad Y = g(X + Y).$$

The proof is deferred to Appendix B.2.

5.2 Choquet and Quantile Representation

In this section, we recall the Choquet integral representation of distortion risk measures and establish an alternative formulation in terms of quantile functions. This representation will be crucial for linking comonotonicity with the structural properties of such risk measures.

To this end, let $c : \mathcal{F} \to [0, 1]$ be a normalized and monotone set function. Recall from Chapter 4 that the Choquet integral of $X \in \mathcal{X}$ with respect to c is defined by

$$\int X \, dc := \int_{-\infty}^{0} \big(c(X > x) - 1\big) \, dx + \int_{0}^{\infty} c(X > x) \, dx.$$

The following proposition recalls that the Choquet integral of the loss defines a monetary risk measure. Its proof was already provided in Example 4.5.

Proposition 5.6. *The Choquet integral of the loss,*

$$\rho(X) := \int (-X) \, dc,$$

is a monetary risk measure on $\mathcal{X}$ that is positively homogeneous.

In order to obtain a more tractable representation of the Choquet integral, it is useful to express it in terms of quantile functions. To this end, we recall the notion of a quantile function with respect to the set function c.

Definition 5.7. [20] Let X be a measurable function on $(\Omega, \mathcal{F})$. An inverse function $r_X : (0,1) \to \mathbb{R}$ of the increasing function $G_X(x) := 1 - c(X > x)$, taken in the sense of Definition A.12, is called a *quantile function* for X with respect to c.

If c is a probability measure, then $G_X(x) = c(X \leq x)$, and the above definition coincides with the classical notion of a quantile function.

Choquet integrals can equivalently be computed on the quantile scale; the following proposition makes this precise.

Proposition 5.8. *[20] Let r_X be a quantile function with respect to c for $X \in \mathcal{X}$. Then*

$$\int X \, dc = \int_0^1 r_X(t) \, dt.$$

Proof. We have $\int (X + m) \, dc = \int X \, dc + m$, and one easily checks that $r_{X+m} = r_X + m$ a.e. for all $m \in \mathbb{R}$ and each quantile function r_{X+m} of $X + m$. Thus, we may assume without loss of generality that $X \geq 0$. In this case, Remark A.15 and Lemma A.13 imply that the largest quantile function r_X^+ is given by

$$r_X^+(t) = \sup\{x \geq 0 : G_X(x) \leq t\} = \int_0^\infty \mathbf{1}_{\{G_X(x) \leq t\}} \, dx.$$

Since $r_X = r_X^+$ a.e. on $(0,1)$, Fubini's theorem implies

$$\int_0^1 r_X(t) \, dt = \int_0^1 \int_0^\infty \mathbf{1}_{\{G_X(x) \leq t\}} \, dx \, dt = \int_0^\infty \int_0^1 \mathbf{1}_{\{G_X(x) \leq t\}} \, dt \, dx.$$

For fixed x, the inner integral satisfies

$$\int_0^1 \mathbf{1}_{\{G_X(x) \leq t\}} \, dt = 1 - G_X(x).$$

Hence,

$$\int_0^1 r_X(t) \, dt = \int_0^\infty (1 - G_X(x)) \, dx.$$

Since $G_X(x) = 1 - c(X > x)$, it follows that $1 - G_X(x) = c(X > x)$, and therefore

$$\int_0^1 r_X(t) \, dt = \int_0^\infty c(X > x) \, dx = \int X \, dc.$$

□

Lemma 5.9. *If $X, Y \in \mathcal{X}$ is a pair of comonotone functions, and r_X, r_Y, r_{X+Y} are quantile functions with respect to c, then*

$$r_{X+Y}(t) = r_X(t) + r_Y(t)$$

for a.e. t.

Proof. By way of Lemma 5.5(c), we may write $Z = X + Y$, $X = f(Z)$, and $Y = g(Z)$ for two increasing and continuous functions f and g. Using the continuity of f and g, the same argument as in the proof of Lemma A.21 shows that $f(r_Z)$ and $g(r_Z)$ are quantile functions for X and Y under c if r_Z is a quantile function for Z. Here, the continuity of f and g is needed, because we did not make any continuity assumptions on the capacity c. An identical argument applied to the continuous and increasing function $h := f + g$ shows that $h(r_Z) = f(r_Z) + g(r_Z)$ is a quantile function for $X + Y$. The assertion now follows from the fact that all quantile functions of a random variable coincide a.e., due to Lemma A.13. □

Remark 5.10. [6] The comonotonicity of Definition 5.1 has a precise order-theoretic meaning. A comonotone pair (X, Y) realizes the *upper Fréchet bound* of its marginals: among all couplings of the laws of X and Y, the comonotone one is the maximal element in the supermodular (concordance) order, the joint law with the strongest possible positive dependence compatible with the given marginals. Equivalently, by Lemma 5.5 both positions are increasing functions of a common variable, which is exactly the defining feature of that coupling. In this language, Lemma 5.9, applied to a quantile function with respect to a probability measure, is precisely the statement that quantile functions are *additive* under the upper Fréchet coupling. This is the order-theoretic structure underlying the comonotonic additivity of Definition 5.2, and it is the point of contact with the broader theory of stochastic orders.

Remark 5.11. Applied to the special case of a quantile function with respect to a probability measure, the preceding lemma implies that $\mathrm{V@R}_\lambda$ and $\mathrm{AV@R}_\lambda$ are comonotonic.

Indeed, in this case, quantile functions coincide with the usual quantile functions of a random variable. By Lemma 5.9, for comonotone random variables X and Y, we have

$$r_{X+Y}(t) = r_X(t) + r_Y(t)$$

for almost every $t \in (0, 1)$. For $\lambda \in (0, 1)$, the Value at Risk is given by

$$\mathrm{V@R}_\lambda(X) = r_{-X}(\lambda),$$

and hence

$$\mathrm{V@R}_\lambda(X + Y) = r_{-(X+Y)}(\lambda) = r_{-X}(\lambda) + r_{-Y}(\lambda) = \mathrm{V@R}_\lambda(X) + \mathrm{V@R}_\lambda(Y),$$

where we used that the quantile functions are additive for comonotonic random variables.

Similarly, the Average Value at Risk admits the representation

$$\mathrm{AV@R}_\lambda(X) = \frac{1}{\lambda} \int_0^\lambda r_{-X}(t)\, dt.$$

Using again the additivity of quantile functions for comonotonic random variables, we obtain

$$\begin{aligned}\mathrm{AV@R}_\lambda(X+Y) &= \frac{1}{\lambda}\int_0^\lambda r_{-(X+Y)}(t)\,dt \\ &= \frac{1}{\lambda}\int_0^\lambda \Big(r_{-X}(t)+r_{-Y}(t)\Big)\,dt \\ &= \mathrm{AV@R}_\lambda(X)+\mathrm{AV@R}_\lambda(Y).\end{aligned}$$

Applying Proposition 5.8 to a continuous distortion of a probability measure in the sense of Definition 4.1 yields the following generalization of Corollary 4.10. It provides a representation of the Choquet integral in terms of quantile functions under distorted probabilities.

Corollary 5.12. *Let $c_\psi(A)=\psi(\mathbb{P}[A])$ be the distortion of the probability measure $\mathbb{P}$ with respect to the continuous distortion function ψ. If φ is an inverse function for the increasing function ψ in the sense of Definition A.12, then the Choquet integral with respect to c_ψ satisfies*

$$\int X\,dc_\psi = \int_0^1 q_X\Big(1-\varphi(t)\Big)\,dt,$$

where q_X is a quantile function for $X\in\mathcal{X}$, taken with respect to $\mathbb{P}$.

Proof. Due to the continuity of ψ, we have

$$\psi(a)\le t \quad\Longleftrightarrow\quad a\le \varphi^+(t)=\inf\{x\mid \psi(x)>t\}.$$

Thus, we can compute the lower quantile function of X with respect to c_ψ:

$$\begin{aligned}r_X^-(t) &= \inf\{x\in\mathbb{R}\mid 1-c_\psi(X>x)\ge t\} \\ &= \inf\{x\in\mathbb{R}\mid \psi(\mathbb{P}[X>x])\le 1-t\} \\ &= \inf\{x\in\mathbb{R}\mid \mathbb{P}[X>x]\le \varphi^+(1-t)\} \\ &= q_X^-\Big(1-\varphi^+(1-t)\Big).\end{aligned}$$

Next, note that $\varphi^+(t)=\varphi(t)$ for a.e. t. Moreover, φ has the continuous cumulative distribution function ψ under the Lebesgue measure, and so we can replace q_X^- by the arbitrary quantile function q_X. □

We are now in a position to state the main structural result of this section, which characterizes comonotonic monetary risk measures in terms of Choquet integrals.

Theorem 5.13. *[45] A monetary risk measure ρ on $\mathcal{X}$ is comonotonic if and only if there exists a normalized monotone set function c on $(\Omega,\mathcal{F})$ such that*

$$\rho(X)=\int(-X)\,dc,\qquad X\in\mathcal{X}.$$

In this case, c is given by

$$c(A) = \rho(-\mathbf{1}_A).$$

The preceding theorem shows that comonotonicity is precisely the structural property that links monetary risk measures to Choquet integrals. This representation is due to Schmeidler [45]. In the companion paper [46], he made comonotonicity a central notion in decision theory under uncertainty, where the additivity of preferences is required only across comonotone acts. In view of Corollary 4.11, this implies in particular that mixtures of Average Value at Risk are comonotonic. As will be shown in the next section, these mixtures in fact exhaust the class of all law-invariant convex risk measures that are comonotonic.

Proof. We already know from Proposition 5.6 that the Choquet integral of the loss is a monetary risk measure. Comonotonicity follows by combining Proposition 5.8 with Lemma 5.9.

Conversely, suppose now that ρ is comonotonic. Then ρ is positively homogeneous according to Lemma 5.4. In particular, we have $\rho(-m) = \rho(0)+m = m$ for $m \geq 0$. Thus, we obtain a normalized monotone set function by letting $c(A) := \rho(-\mathbf{1}_A)$. Moreover, $\rho_c(X) := \int(-X)\,dc$ is a comonotonic monetary risk measure on $\mathcal{X}$ that coincides with ρ on indicator functions: $\rho(-\mathbf{1}_A) = c(A) = \rho_c(-\mathbf{1}_A)$. Let us now show that ρ and ρ_c coincide on simple random variables of the form

$$X = \sum_{i=1}^{n} x_i \mathbf{1}_{A_i}, \qquad x_i \in \mathbb{R},\ A_i \in \mathcal{F}.$$

Since these random variables are dense in L^∞, Lemma 3.6 will then imply that $\rho = \rho_c$. In order to show that $\rho_c(X) = \rho(X)$ for X as above, we may assume without loss of generality that $x_1 \geq x_2 \geq \cdots \geq x_n$ and that the sets A_i are disjoint. By cash invariance, we may also assume $X \geq 0$, that is, $x_n \geq 0$. Thus, we can write

$$X = \sum_{i=1}^{n} b_i \mathbf{1}_{B_i},$$

where

$$b_i := x_i - x_{i+1} \geq 0, \quad x_{n+1} := 0, \quad B_i := \bigcup_{k=1}^{i} A_k.$$

Note that $b_i\mathbf{1}_{B_i}$ and $b_k\mathbf{1}_{B_k}$ are comonotone for all i, k. Therefore, also $\sum_{i=1}^{k-1} b_i\mathbf{1}_{B_i}$ and $b_k\mathbf{1}_{B_k}$ are comonotone, and we obtain inductively

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

Since $\rho(-\mathbf{1}_{B_i}) = c(B_i) = \rho_c(-\mathbf{1}_{B_i})$, we conclude that

$$\rho(-X) = \sum_{i=1}^{n} b_i \rho_c(-\mathbf{1}_{B_i}) = \rho_c(-X).$$

This shows that $\rho = \rho_c$, and the proof is complete. □

Remark 5.14. The argument at the end of the preceding proof provides a derivation of the explicit formula for the Choquet integral of simple random variables already stated in Remark 4.6. This confirms that the representation introduced earlier is consistent with the structural characterization obtained here.

5.3 Characterization of Comonotonic Convex Risk Measures

In the previous sections, we have shown that comonotonic monetary risk measures can be represented as Choquet integrals with respect to normalized monotone set functions. We now refine this perspective by focusing on those risk measures that additionally satisfy convexity and law-invariance.

Our goal in this section is to obtain a complete characterization of this class. To this end, we restrict attention to atomless probability spaces and exploit the fact that law-invariance allows us to express the underlying set function in terms of the distribution of the position. This leads to a representation in terms of mixtures of Average Value at Risk.

We now turn to the following characterization.

Theorem 5.15. *[35] On an atomless probability space, the class of risk measures*

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

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

Proof. Comonotonicity of ρ_μ follows from Corollary 4.11 and Theorem 5.13.

Conversely, let us assume that ρ is a law-invariant convex risk measure that is also comonotonic. By Theorem 5.13, $\rho(X) = \int (-X)\, dc$ for $c(A) := \rho(-\mathbf{1}_A)$. The law-invariance of ρ implies that $c(A)$ is a function of the probability $\mathbb{P}[A]$, that is, there exists an increasing function ψ on $[0,1]$ such that

$$\psi(0) = 0, \qquad \psi(1) = 1, \qquad c(A) = \psi(\mathbb{P}[A]).$$

Note that $\mathbf{1}_{A\cup B}$ and $\mathbf{1}_{A\cap B}$ form a pair of comonotone functions for all $A, B \in \mathcal{F}$. Hence, comonotonicity and subadditivity of ρ imply

$$\begin{aligned} c(A\cap B) + c(A\cup B) &= \rho(-\mathbf{1}_{A\cap B}) + \rho(-\mathbf{1}_{A\cup B}) \\ &= \rho(-\mathbf{1}_{A\cap B} - \mathbf{1}_{A\cup B}) \\ &= \rho(-\mathbf{1}_A - \mathbf{1}_B) \\ &\leq c(A) + c(B). \end{aligned} \tag{5.1}$$

Proposition 4.3 thus implies that ψ is concave. Corollary 4.11 finally shows that the Choquet integral with respect to c can be identified with a risk measure ρ_μ, where μ is obtained from ψ via Lemma 4.7. □

The two preceding chapters have, from opposite directions, established the various faces of a single class of risk measures. Chapter 4 proceeded top-down: beginning with the mixtures of Average Value at Risk, Theorem 4.8 and Corollary 4.11 identified them as Choquet integrals against a concave distortion. Chapter 5 proceeded in reverse: Theorem 5.13 showed that comonotonicity alone forces the Choquet form, and Theorem 5.15 showed that convexity and law-invariance then single out exactly the concave distortions. Collecting these results, we obtain the equivalence at the heart of this thesis.

Theorem 5.16 (Central Equivalence)**.** *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be atomless and let ρ be a risk measure on L^∞. The following are equivalent:*

1. *ρ is convex, law-invariant and comonotonic.*
2. *$\rho(X) = \int(-X)\, dc_\psi$ is the Choquet integral of the loss against the distortion $c_\psi = \psi \circ \mathbb{P}$ of a concave distortion function ψ.*
3. *$\rho(X) = \int_{[0,1]} \mathrm{AV@R}_\lambda(X)\, \mu(d\lambda)$ is a mixture of Average Value at Risk for some $\mu \in \mathcal{M}_1([0,1])$.*

In particular, every such ρ is automatically coherent.

Proof. The equivalence *1.*⇔*3.* is Theorem 5.15. For *2.*⇔*3.*, Theorem 4.8 and Corollary 4.11 identify the Average-Value-at-Risk mixtures with the Choquet integrals against a concave distortion, the correspondence $\psi \leftrightarrow \mu$ being given by Lemma 4.7. The concavity of ψ in (*2.*), which is what makes the corresponding set function submodular and consequently ρ convex, is precisely the content of inequality (5.1) in the proof of Theorem 5.15. Coherence is part of Theorem 5.15. □

Theorem 5.15 shows that law-invariant convex risk measures that are comonotonic admit a representation in terms of Choquet integrals associated with distortion functions.

This naturally raises the question which structural properties of the underlying set function c correspond to convexity of the induced risk measure.

To address this, we now turn to a characterization of those normalized monotone set functions that generate convex (and coherent) risk measures via the Choquet integral.

Theorem 5.17. *For the Choquet integral with respect to a normalized monotone set function c, the following conditions are equivalent:*

(a) $\rho(X) := \int (-X)\, dc$ *is a convex risk measure on* $\mathcal{X}$.

(b) $\rho(X) := \int (-X)\, dc$ *is a coherent risk measure on* $\mathcal{X}$.

(c) For $\mathcal{Q}_c := \{Q \in \mathcal{M}_{1,f} \mid Q[A] \leq c(A) \text{ for all } A \in \mathcal{F}\}$,

$$\int X\, dc = \max_{Q \in \mathcal{Q}_c} \mathbb{E}_Q[X] \qquad \text{for } X \in \mathcal{X}.$$

(d) The set function c is submodular:

$$c(A \cap B) + c(A \cup B) \leq c(A) + c(B).$$

In this case, $\mathcal{Q}_c$ is equal to the maximal representing set $\mathcal{Q}_{\max}$ for ρ.

Condition (c) is once more the measure side of the dual correspondence (Remark 3.48): the representation of the Choquet integral as a maximum over the core $\mathcal{Q}_c$ is conjugate to its distributional, quantile-based form.

The equivalence of the conditions follows by relating the convexity of $\rho(X) = \int (-X)\, dc$ to structural properties of the set function c. In particular, convexity corresponds to submodularity of c, while the dual representation arises via finitely additive probability measures in $\mathcal{M}_{1,f}$. For completeness, the full argument is provided in Appendix B.2, where the necessary auxiliary results are established.

As a direct consequence of Theorem 5.13 and Theorem 5.17, we obtain the following complete characterization of comonotonic convex risk measures.

Corollary 5.18. *A convex risk measure on $\mathcal{X}$ is comonotonic if and only if it arises as the Choquet integral of the loss with respect to a submodular, normalized, and monotone set function c. In this case, c is given by $c(A) = \rho(-\mathbf{1}_A)$, and ρ has the representation*

$$\rho(X) = \max_{Q \in \mathcal{Q}_c} \mathbb{E}_Q[-X],$$

where $\mathcal{Q}_c = \{Q \in \mathcal{M}_{1,f} \mid Q[A] \leq c(A) \text{ for all } A \in \mathcal{F}\}$ is equal to the maximal representing set $\mathcal{Q}_{\max}$.

Proof. Theorems 5.13 and 5.17 state that $\rho(X) := \int (-X)\, dc$ is a comonotonic coherent risk measure that can be represented as in the assertion, as soon as c is a submodular, normalized, and monotone set function.

Conversely, any comonotonic convex risk measure ρ is coherent and arises as the Choquet integral of $c(A) := \rho(-\mathbf{1}_A)$, due to Theorem 5.13. Theorem 5.17 then gives the submodularity of c. □

The previous results not only provide a characterization of comonotonic convex risk measures, but also yield additional insight into the structure of the underlying set function and its representing set of measures. In this context, the role of the core $\mathcal{Q}_c$, which collects all probability measures compatible with the set function c, becomes apparent. This interpretation is summarized in the following remark.

Remark 5.19. Let c be a normalized, monotone, submodular set function. Theorem 5.17 implies in particular that the core $\mathcal{Q}_c$ of c is nonempty. Moreover, c can be recovered from $\mathcal{Q}_c$:

$$c(A) = \max_{Q \in \mathcal{Q}_c} Q[A] \qquad \text{for all } A \in \mathcal{F}.$$

If c has the additional continuity property that $c(A_n) \to 0$ for any decreasing sequence (A_n) of events such that $\bigcap_n A_n = \emptyset$, then this property is shared by any $Q \in \mathcal{Q}_c$, and it follows that Q is σ-additive. Thus, the corresponding coherent risk measure $\rho(X) = \int (-X)\, dc$ admits a representation in terms of σ-additive probability measures. It follows by Lemma 3.27 that ρ is continuous from above.

In this chapter, we have studied comonotonic risk measures and identified comonotonicity as the key structural property underlying the Choquet integral representation introduced in the previous chapter.

Building on the theory of distortion risk measures, we have shown that comonotonic monetary risk measures admit a representation in terms of Choquet integrals with respect to normalized monotone set functions. Moreover, we have seen that convexity of the induced risk measure corresponds to submodularity of the underlying set function.

Combining these insights, we obtained a complete characterization of law-invariant convex comonotonic risk measures on atomless probability spaces. In particular, this class coincides with mixtures of Average Value at Risk, thereby linking the Choquet integral representation to a more classical quantile-based formulation.

Overall, this chapter clarifies the structural role of comonotonicity in risk measurement and provides the foundation for understanding distortion-based risk measures as a natural and tractable subclass of convex risk measures.

Chapter 6

Distributionally Robust Insurance under the Wasserstein Distance

6.1 Introduction and Practical Context

In the preceding chapters, we developed the theory of convex distortion risk measures from the ground up. We studied their axiomatic foundations as convex and coherent risk measures, their representation through the Choquet integral and concave distortion functions, as well as their characterization as the comonotone additive members of this class. The aim of this chapter is to apply this theory to a concrete and economically relevant problem, namely the design of optimal insurance contracts. We follow the recent work of Boonen and Jiang [11], whose framework rests precisely on the structural properties we have established, and we summarize their main results before reproducing their numerical analysis in Section 6.4.

Insurance is one of the most important tools for managing risk. Individuals and firms are regularly exposed to the possibility of financial losses. A house may be damaged by a storm or a fire, a car may be involved in an accident, or a business may face unexpected liabilities. Since most people and companies are unwilling or unable to bear these losses themselves, insurance provides a way to reduce their financial risk. Insurance addresses this by allowing the decision maker to transfer part of the risk to an insurer. in exchange for a fixed payment made in advance, the insurer agrees to cover some portion of the loss should it occur. As a result, an uncertain and potentially large loss is replaced by a smaller, certain cost, which is precisely the kind of trade a risk-averse decision maker values.

We now make this precise. An insurance contract has two main components. A decision maker (DM) faces an uncertain, non-negative loss X and may transfer part of it to an insurer. The contract is described by the indemnity function I, where $I(x)$ is the amount reimbursed when the realized loss is x, together with the associated premium $\pi(I)$, which denotes the fixed payment made by the decision maker to the

insurer in exchange for the coverage. After purchasing such a contract, the DM is left with the retained loss $X - I(X) + \pi(I)$, consisting of the original loss, reduced by the reimbursement and increased by the premium paid. Since the premium is a fixed function of the indemnity, the central question of optimal insurance design reduces to the choice of the indemnity function alone. The DM seeks the contract that is, in a sense to be made precise, least risky.

The importance of this problem goes beyond its mathematical formulation. In practice, insurers and reinsurers do not know the true loss distribution. Instead, they rely on historical data, which may not fully reflect future developments or rare catastrophic events. Consequently, a contract that is optimal under one estimated distribution may perform poorly in reality. The distributionally robust approach addresses this problem by seeking contracts that perform well across a range of plausible loss distributions instead of relying on a single model.

The main contribution of Boonen and Jiang [11], who study the distributionally robust insurance under the Wasserstein distance, is to solve this problem explicitly. They derive the optimal contract, which turns out to take the familiar form of a stop-loss agreement. This is a contract in which the insurer covers the entire loss above a fixed threshold, while the decision maker retains everything below it. They also characterize the worst-case loss distribution against which the decision maker hedges. What makes their results particularly valuable is that both objects are obtained in closed form rather than only numerically, and that the method rests precisely on the structural properties of convex distortion risk measures developed in the preceding chapters. This makes their work a natural application of the theory of this thesis, and it is the reason we examine it in detail.

The remainder of this chapter is organized as follows. Section 6.2 introduces the model and formulates the distributionally robust insurance problem. Section 6.3 presents the solution approach and the main results, namely the optimal indemnity function and the worst-case distribution. Section 6.4 reproduces the numerical study of Boonen and Jiang [11] for the case in which the DM employs the GlueV@R risk measure, replicating their analysis of how the worst-case distribution responds to the model parameters. Section 6.5 concludes with a short discussion of the findings.

6.2 Model Setup and Problem Formulation

The optimal insurance literature studies the contracting problem from several perspectives, including that of the insurer, that of both parties jointly, giving rise to Pareto-optimal contracts, or that of the decision maker alone. Boonen and Jiang [11] adopt the last of these. The decision maker (DM) is free to purchase any admissible indemnity function at a price fixed by a given premium principle, and he does so by minimizing a risk measure of his retained loss, rather than by maximizing expected

utility. The present section makes this setting precise. Throughout, we use the notation $x \wedge y = \min\{x, y\}$, $x \vee y = \max\{x, y\}$ and $(x)_+ = \max\{x, 0\}$, and we write $\mathbf{1}_A$ for the indicator function of a set A.

The loss and its distribution

We fix a one-period economy on an atomless probability space $(\Omega, \mathcal{F}, \mathbb{P})$. The DM faces an insurable, non-negative loss modelled by a random variable X with support contained in $[0, M]$ for some $M > 0$. Following the convention of Boonen and Jiang [11], we denote by $\mathcal{P}$ the collection of all probability measures P on $(\Omega, \mathcal{F})$ under which X takes values in $[0, M]$ almost surely; each $P \in \mathcal{P}$ represents a candidate model for the loss. For $P \in \mathcal{P}$, we write $F_P(x) := P(X \leq x)$ for the cumulative distribution function of X under P and

$$S_P(x) := 1 - F_P(x) = P(X > x), \qquad x \in [0, M], \tag{6.1}$$

for the associated survival function. We assume that X and $(\Omega, \mathcal{F})$ are rich enough that every probability distribution on $[0, M]$ arises as the law of X under some $P \in \mathcal{P}$. Since all functionals considered in this chapter depend on P only through F_P, we may thus identify $\mathcal{P}$ with the set of possible loss distributions; this identification underlies the optimization over survival functions in Section 6.3. The boundedness of the support guarantees that all quantities introduced below are finite.

The risk measure

To assess the riskiness of a position, the DM employs a convex distortion risk measure. These are exactly the risk measures constructed in Chapter 4. Starting from a distortion function $g : [0, 1] \to [0, 1]$ that is increasing with $g(0) = 0$ and $g(1) = 1$, one forms the distorted set function $c_g(A) = g(P[A])$ and evaluates a loss through the Choquet integral with respect to c_g (Definition 4.4 and Example 4.5). Following the convention of Boonen and Jiang [11], who work directly with the loss variable Z, this functional reads

$$\rho_g^P(Z) = \int_0^\infty g\Big(P(Z > z)\Big)\, dz + \int_{-\infty}^0 \Big[g\Big(P(Z > z)\Big) - 1\Big]\, dz, \tag{6.2}$$

where the explicit superscript P records the dependence on the reference measure.

In what follows we assume that g is, in addition, concave. By Proposition 4.3, this is equivalent to the submodularity of the capacity c_g, which in turn is precisely the condition under which the induced risk measure is convex, and hence coherent in the sense of Definition 3.8. The risk measure ρ_g^P therefore inherits all the structural properties established earlier; in particular, for a fixed P it is [21, 51]:

- *comonotone additive*: $\rho_g^P(Z + Y) = \rho_g^P(Z) + \rho_g^P(Y)$ whenever Z and Y are comonotone;

- *subadditive*: $\rho_g^P(Z+Y) \le \rho_g^P(Z) + \rho_g^P(Y)$ for arbitrary Z and Y.

Comonotone additivity together with $\rho_g^P(1) = 1$ yields translation invariance, $\rho_g^P(Z+c) = \rho_g^P(Z) + c$ for all $c \in \mathbb{R}$. This class is rich and it contains the Value at Risk and the Average Value at Risk of Section 3.4 as special cases, the latter being a genuine convex distortion risk measure. As we shall see, comonotone additivity, translation invariance, and the concavity of g are exactly the properties that drive the explicit solution of Section 6.3.

The insurance contract

An insurance contract is a pair $(I, \pi(I))$, where I is the indemnity function and $\pi(I)$ the corresponding premium. If the DM purchases $(I, \pi(I))$, his end-of-period loss is $X - I(X) + \pi(I)$. To rule out moral hazard, we impose the *incentive compatibility* condition of Huberman et al. [32], which requires both the loss retained by the DM and the loss ceded to the insurer to be increasing in X. The admissible indemnity functions therefore form the set

$$\mathcal{I} := \Big\{ I : [0,M] \to [0,M] \;\Big|\; I(0) = 0, \quad 0 \le I(x_2) - I(x_1) \le x_2 - x_1, \quad \forall\, 0 \le x_1 \le x_2 \le M \Big\}. \tag{6.3}$$

A crucial consequence is that, for $I \in \mathcal{I}$, the retained loss $X - I(X)$ and the ceded loss $I(X)$ are comonotone in the sense of Chapter 5. This is what makes the comonotone additivity of ρ_g^P applicable to the decomposition of the total loss, and it is the reason that the theory of comonotone risk measures is the right tool for this problem. The set $\mathcal{I}$ is large and contains many familiar contracts, including the stop-loss function $I(x) = (x-d)_+$, the quota-share function $I(x) = cx$, and the truncated stop-loss function $I(x) = (x-d_1)_+ \wedge d_2$. Moreover, every $I \in \mathcal{I}$ is 1-Lipschitz continuous and admits the integral representation

$$I(x) = \int_0^x \eta(t)\, dt, \qquad x \in [0,M], \tag{6.4}$$

where η, taking values in $[0,1]$, is the *marginal indemnity function* [4, 52]; it specifies the rate at which coverage increases with the loss.

The premium principle

The insurer prices contracts under a selected probability measure $Q \in \mathcal{P}$. We adopt the *expected-value premium principle*,

$$\pi(I) = (1+\theta)\,\mathbb{E}_Q[I(X)] = (1+\theta)\int_0^M I(x)\,dF_Q(x), \tag{6.5}$$

where $\theta \geq 0$ is the *safety loading factor*, an additional margin charged by the insurer on top of the expected indemnity to cover expenses and profit.

If the DM knew the loss distribution P exactly, his problem would be the classical one of minimizing the risk of the retained loss,

$$\min_{I \in \mathcal{I}} \rho_g^P\Big(X - I(X) + \pi(I)\Big), \tag{6.6}$$

which is well understood in the literature. The distinctive feature of the present model is that P is *not* known exactly.

Ambiguity and the benchmark distribution

In practice, the DM does not have perfect knowledge of the distribution of X. His information comes from limited historical data and incomplete market information, from which he can form a single best estimate but not a fully reliable model. We capture this by assuming that the DM has a *benchmark distribution* $F_{\mathbb{B}}$, with $\mathbb{B} \in \mathcal{P}$, that represents his central guess for the loss distribution, while acknowledging that the true distribution may deviate from it. Rather than committing to $F_{\mathbb{B}}$ alone, the DM considers all distributions that are sufficiently close to it, and protects himself against the entire resulting set. This is the notion of *ambiguity*, and the set of admissible distributions is called the *ambiguity set*. It remains to specify how closeness between distributions is measured.

The Wasserstein distance

Boonen and Jiang [11] measure the distance between distributions by the Wasserstein distance, a central object in the theory of optimal transport [50]. Intuitively, the Wasserstein distance quantifies the minimal "effort" required to transport the probability mass of one distribution onto another, where effort is measured as the average distance the mass is moved. Among its attractive features are that it is a genuine metric on distributions, that it is robust against small perturbations, and that, unlike divergence-based measures, it can compare distributions with differing supports.

Definition 6.1. For $p \in \mathbb{Z}_+$, the *p-Wasserstein distance* between F_{P_1}, F_{P_2} with $P_1, P_2 \in$

$\mathcal{P}$ is

$$W_p(F_{P_1}, F_{P_2}) = \left(\inf_{X \sim F_{P_1}, Y \sim F_{P_2}} \mathbb{E}\big[\, |X - Y|^p \,\big] \right)^{1/p},$$

where $X \sim F_P$ means that X has cumulative distribution function F_P and the infimum is taken over all joint distributions with the prescribed marginals.

On the real line, the Wasserstein distance admits the closed-form quantile representation

$$W_p(F_{P_1}, F_{P_2}) = \left(\int_0^1 \left| F_{P_1}^{-1}(t) - F_{P_2}^{-1}(t) \right|^p dt \right)^{1/p}, \tag{6.7}$$

where $F_P^{-1}(t) = \inf\{x \in \mathrm{dom}(F_P) : F_P(x) \geq t\}$ is the left-continuous inverse [39]. The ambiguity set is then taken to be the Wasserstein ball of radius $\epsilon^{1/p}$ centered at the benchmark,

$$\mathcal{P}_\epsilon := \Big\{ P \in \mathcal{P} : W_p(F_P, F_{\mathbb{B}}) \leq \epsilon^{1/p} \Big\}. \tag{6.8}$$

The radius parameter $\epsilon > 0$ controls the size of the ambiguity set and hence the degree of robustness. In other words, a larger ϵ admits more distant distributions and yields a more conservative contract. The choice $\epsilon = 0$ collapses $\mathcal{P}_0$ to the singleton $\{\mathbb{B}\}$ and recovers the classical problem (6.6), which is why we take $\epsilon > 0$ throughout. It is worth noting that for $p = 2$ the constraint $W_2(F_P, F_{\mathbb{B}}) \leq \epsilon^{1/2}$ implies

$$\epsilon \geq (\mu_P - \mu_{\mathbb{B}})^2 + (\sigma_P - \sigma_{\mathbb{B}})^2, \tag{6.9}$$

where μ and σ denote mean and standard deviation under the respective measures [8]. The Wasserstein ball thus simultaneously constrains the first two moments of X, connecting the distance-based formulation to moment-based approaches.

The robust problem

Combining these ingredients, the DM seeks the indemnity function that minimizes his risk under the most adverse distribution in the ambiguity set. This is the central problem of the chapter.

Problem 6.2. *For a given $\epsilon > 0$, solve*

$$\inf_{I \in \mathcal{I}} \sup_{P \in \mathcal{P}_\epsilon} \rho_g^P\Big(X - I(X) + \pi(I)\Big).$$

The inner supremum identifies, for each candidate indemnity, the worst-case distribution within the Wasserstein ball, while the outer infimum selects the indemnity that is best protected against this worst case. Solving Problem 6.2 therefore amounts to determining two objects, namely the optimal indemnity function I^*, which describes the robust contract, and the worst-case distribution P^*, which identifies the scenario the DM hedges against. Their explicit characterization is the subject of Section 6.3.

6.3 Solution Approach and Main Results

This section summarizes the analytical solution of Problem 6.2 obtained by Boonen and Jiang [11]. In keeping with the scope of this chapter, we state the results without proof and concentrate on their structure and economic interpretation, before turning to the numerical replication in Section 6.4. The argument proceeds in three steps. The inf–sup problem is first turned into a sup–inf problem by means of the minimax theorem. Next, the inner minimization over indemnity functions is then solved explicitly. Finally, the outer maximization over distributions is reduced, via a survival-function representation of the Wasserstein constraint, to a tractable problem whose solution is the worst-case survival function. Throughout, the explicit computations can be found in the original paper.

Interchanging infimum and supremum

The first and decisive step is that the order of optimization in Problem 6.2 may be reversed. The admissible set $\mathcal{I}$ is convex and compact, the objective $\rho_g^P(X-I(X)+\pi(I))$ is linear in I, due to the translation invariance and comonotone additivity of ρ_g^P, and concave in F_P, owing to the concavity of g. The ambiguity set $\mathcal{P}_\epsilon$ is convex. These are exactly the hypotheses of the minimax theorem, stated as Theorem C.1, which therefore applies and yields the equivalent formulation

$$\sup_{P\in\mathcal{P}_\epsilon}\ \inf_{I\in\mathcal{I}}\ \rho_g^P\Big(X-I(X)+\pi(I)\Big). \tag{6.10}$$

The benefit is that the inner problem now optimizes over indemnity functions for a *fixed* distribution P, which is a classical and well-understood problem.

The optimal indemnity function

For a fixed distribution, the inner minimization is solved by a marginal analysis in terms of the marginal indemnity function $\eta = I'$ of (6.4).

Lemma 6.3. *For a fixed $P\in\mathcal{P}_\epsilon$, the optimal indemnity function to the inner problem of (6.10) is $I^*(x;P)=\int_0^x \eta^*(t;P)\,dt$, where*

$$\eta^*(t;P)=\mathbf{1}_{\{t:(1+\theta)S_Q(t)<g(S_P(t))\}}(t)+\gamma(t)\,\mathbf{1}_{\{t:(1+\theta)S_Q(t)=g(S_P(t))\}}(t),$$

and γ is an arbitrary $[0,1]$-valued, Lebesgue-measurable function.

The quantity

$$(1+\theta)S_Q(x)-g\Big(S_P(x)\Big) \tag{6.11}$$

can be read as the *net price* of marginal coverage at loss level x. The term $(1+\theta)S_Q(x)$ is the marginal premium charged by the insurer, while $g(S_P(x))$ is the marginal reduction

in risk perceived by the DM. Lemma 6.3 states that the DM buys full marginal coverage ($I'(x) = 1$) wherever this net price is negative and no coverage ($I'(x) = 0$) wherever it is positive. On the set where it vanishes he is indifferent, which is the source of the non-uniqueness encoded by γ. In particular, once the worst-case survival function S_{P^*} is known, $I^*(\cdot; P^*)$ solves Problem 6.2, and the remaining task is to identify S_{P^*}.

Reduction of the outer problem

Substituting the optimal indemnity into the objective reduces the outer problem in (6.10) to a maximization over survival functions,

$$\sup_{S_P} \int_0^M \Big(g(S_P(x)) \wedge (1+\theta) S_Q(x)\Big)\, dx \qquad \text{subject to} \quad W_p(F_P, F_{\mathbb{B}})^p \le \epsilon. \tag{6.12}$$

Since the objective is increasing in S_P, attention may be restricted to distributions whose survival function dominates the benchmark. The key device of the paper is that, on this restricted set, the p-Wasserstein constraint, which is originally stated in terms of quantile functions, can be rewritten entirely in terms of survival functions, and in this form it is convex in S_P. We do not reproduce the resulting expression here; it is obtained through a survival-function representation of the moments of X, and the details are given in Boonen and Jiang [11]. What matters for the sequel is that the objective in (6.12) is concave in S_P, the rewritten constraint is convex, and the benchmark itself is strictly feasible since $\epsilon > 0$. Slater's condition therefore holds, strong duality applies [12], and the problem can be solved through its Lagrangian dual, in which a multiplier $\beta \ge 0$ enforces the Wasserstein constraint. The value of β distinguishes the two cases that follow, with $\beta = 0$ when the constraint is slack and $\beta > 0$ when it binds.

The worst-case survival function

To state the results, define

$$x_0 := \sup\Big\{x \in [0, M) : S_Q(x) \ge \tfrac{1}{1+\theta}\Big\}, \tag{6.13}$$

together with the sets

$$\mathcal{A} := \Big\{x \in [x_0, M] : (1+\theta) S_Q(x) \ge g(S_{\mathbb{B}}(x))\Big\}, \qquad \mathcal{B} := [x_0, M] \setminus \mathcal{A}.$$

When the constraint is slack, the worst-case survival function is given in closed form.

Theorem 6.4. *Let x_0, $\mathcal{A}$ and $\mathcal{B}$ be as above and set $t_0 := g^{-1}(1)$. If the survival function*

$$\widetilde{S}^*(x) = \Big(t_0 \vee S_{\mathbb{B}}(x)\Big)\mathbf{1}_{[0,x_0)}(x) + g^{-1}\Big((1+\theta) S_Q(x)\Big)\mathbf{1}_{\mathcal{A}}(x) + S_{\mathbb{B}}(x)\mathbf{1}_{\mathcal{B}}(x) \tag{6.14}$$

satisfies the Wasserstein constraint, then it is the worst-case survival function solving (6.12)*, that is,* $S_{P^*} = \widetilde{S}^*$.

When the constraint binds, the solution is described through the function

$$K(t;x) = g(t) - \beta\left(x^{p-1}t - \int_0^M (p-1)(x-y)^{p-2}\, t \wedge S_{\mathbb{B}}(y)\, dy\right), \tag{6.15}$$

which collects the distortion term and the dual contribution of the Wasserstein constraint.

Theorem 6.5. *Let* x_0, $\mathcal{A}$ *and* $\mathcal{B}$ *be as in Theorem* 6.4*. If* $\beta > 0$*, the worst-case survival function is*

$$S_{P^*}(x;\beta) = \widehat{S}(x;\beta)\mathbf{1}_{[0,x_0)}(x) + \Big(\widehat{S}(x;\beta) \wedge g^{-1}\Big((1+\theta)S_Q(x)\Big)\Big)\mathbf{1}_{\mathcal{A}}(x) + S_{\mathbb{B}}(x)\mathbf{1}_{\mathcal{B}}(x), \tag{6.16}$$

where $\widehat{S}(x;\beta) = \inf\{t \in [S_{\mathbb{B}}(x), 1] : K'(t;x) \le 0\}$*, with* $\inf\emptyset$ *taken as the right endpoint of the interval, and* β *is chosen so that the Wasserstein constraint holds with equality.*

Interpretation

The two theorems share a common structure that admits a clear economic reading. On the set $\mathcal{B}$ the worst-case survival function coincides with the benchmark, $S_{P^*}(x) = S_{\mathbb{B}}(x)$. There the net price (6.11) is negative, so these losses are fully ceded to the insurer, and the DM has no incentive to distort a distribution for risk he does not retain. On $[0, x_0)$ and on $\mathcal{A}$, by contrast, he retains exposure and therefore shifts probability mass towards larger losses, so that S_{P^*} dominates the benchmark. The multiplier β measures how tightly the ambiguity budget ϵ binds. When ϵ is large the constraint is slack and $\beta = 0$, recovering Theorem 6.4; when ϵ is small it binds and $\beta > 0$ rescales the worst-case function so as to exhaust exactly the available budget.

Two further remarks situate these results. First, the parameter p controls the size of the ambiguity set. A larger p yields a smaller Wasserstein ball and hence a less adverse worst-case distribution, while the case $p = 1$ is degenerate, as the constraint then reduces to a bound on the first moment alone. Second, when the benchmark coincides with the insurer's measure, $\mathbb{B} = Q$, the net price is positive on $[0, x_0)$ and negative on the upper tail, so that the optimal contract reduces to a classical stop-loss function $I^*(x) = (x - x_1)_+$ with a deductible x_1 determined by the model parameters. This is precisely the configuration we study numerically, for the GlueV@R risk measure, in Section 6.4.

6.4 A Numerical Case Study

The analytical results of Section 6.3 describe the worst-case survival function abstractly, for any convex distortion risk measure. To see the theory at work, and to develop

intuition for how the worst-case distribution depends on the model parameters, we now specialise to a concrete setting. Following Boonen and Jiang [11], we let the decision maker assess risk through the *GlueV@R* family and we reproduce their sensitivity analysis under the 2-Wasserstein distance. The purpose of this section is twofold. First, we want to make the closed-form solution of Section 6.3 tangible on a fully specified example and second, to isolate the two economic levers that govern the robust contract, the insurer's safety loading θ and the DM's level of risk aversion, and to read off their effect on the worst-case scenario.

The GlueV@R risk measure

The GlueV@R risk measures were introduced by Belles-Sampera, Guillén and Santolino [7] as a flexible generalization of the two risk measures most widely used in banking and insurance, the Value at Risk and the Average Value at Risk of Section 3.4. Like these two, GlueV@R is a distortion risk measure. What distinguishes it is that its distortion function is a piecewise-linear curve with two adjustable "kinks", which lets the DM place tunable weight on different portions of the loss tail. Every member of the family is obtained by choosing four parameters,

$$0 \le \alpha \le \gamma \le 1, \qquad r_1 \in [0,1], \qquad r_2 \in [r_1, 1],$$

and setting

$$g_{\alpha,\gamma}^{r_1,r_2}(x) = \begin{cases} \dfrac{r_1}{1-\gamma}\, x, & 0 \le x < 1-\gamma, \\ r_1 + \dfrac{r_2 - r_1}{\gamma - \alpha}\Big[x - (1-\gamma)\Big], & 1-\gamma \le x < 1-\alpha, \\ 1, & 1-\alpha \le x \le 1. \end{cases} \tag{6.17}$$

The two levels $1-\gamma$ and $1-\alpha$ mark the confidence levels at which the slope of the distortion changes, while r_1 and r_2 fix the heights of the distortion at those two points. The construction is rich enough to recover the classical risk measures as boundary cases: one obtains V@R at level α by taking $g_{\alpha,\alpha}^{0,0}$, AV@R at level α by taking $g_{\alpha,\alpha}^{1,1}$, and the Range Value at Risk by taking $g_{\alpha,\gamma}^{0,1}$.

Recall from Section 6.3 that the entire solution machinery, the minimax argument, the marginal analysis of the indemnity, and the reduction of the outer problem, rests on the *concavity* of the distortion function g. Not every GlueV@R is concave, so we must restrict the parameters to the region in which (6.17) is concave on $[0,1]$. Following Boonen and Jiang [11], we set $r_2 = 1$ and impose

$$\frac{1-r_1}{\gamma-\alpha} < \frac{r_1}{1-\gamma}, \tag{6.18}$$

which forces the slope on the middle segment to be smaller than the slope on the first segment, so that the piecewise-linear curve bends downward at each kink and is therefore concave. We further exclude the degenerate value $r_1 = 1$, for which the GlueV@R collapses to the AV@R already treated by Boonen and Jiang [10] for the L^2 and L^1 balls, and we assume $r_1 < 1$ throughout. For the p-Wasserstein distance we take $p \geq 2$, the case $p = 1$ being degenerate as noted in Section 6.3.

The data of the experiment

To reproduce the numerical study we fix the following specification, which is exactly that of Boonen and Jiang [11].

The loss distribution. The insurable loss X follows a truncated exponential law, with cumulative distribution function

$$F_X(x) = \begin{cases} 1 - e^{-x/100}, & x \in [0, 5000], \\ 1, & x \in (5000, \infty). \end{cases} \tag{6.19}$$

The truncation at $M = 5000$ keeps the support bounded, as required by the model of Section 6.2, while the level is chosen far enough into the tail that it has negligible effect on the computations. Throughout the study the DM adopts the insurer's pricing measure as his benchmark, $\mathbb{B} = Q$, so that the net price of marginal coverage is positive on the body of the distribution and negative on the upper tail. As observed in the previous sections, the optimal contract is then the classical stop-loss $I^*(x) = (x - x_1)_+$.

The risk measure. The DM's GlueV@R preference is specified by

$$r_1 = 0.6, \qquad r_2 = 1, \qquad \alpha = 0.05, \qquad \gamma = 0.7,$$

a parameter combination that satisfies the concavity condition (6.18). The level $\alpha = 0.05$ fixes the point $1 - \alpha = 0.95$ beyond which the distortion is saturated at 1, so that the 5% worst outcomes receive full weight, while $\gamma = 0.7$ and $r_1 = 0.6$ locate and shape the intermediate kink. When we later vary the DM's risk aversion, we do so by increasing r_1. A larger r_1 makes the distortion function more sharply concave, and consequently corresponds to a more risk-averse DM.

The ambiguity set. We measure the distance between distributions with the 2-Wasserstein distance throughout, and fix the radius of the Wasserstein ball at $\epsilon^{1/2} = 10$. This value is chosen so that the ambiguity budget is large enough to leave the constraint slack in some of the scenarios below and binding in others, which lets us observe both

regimes, the $\beta = 0$ case of Theorem 6.4 and the $\beta > 0$ case of Theorem 6.5, within a single experiment.

The sensitivity design. Two one-at-a-time sensitivity analyses are carried out. In the first we vary the safety loading over $\theta \in \{0.1, 0.5, 0.9\}$ while holding the GlueV@R parameters fixed. In the second we vary the risk-aversion parameter over $r_1 \in \{0.6, 0.7, 0.8\}$ while holding $\theta = 0.5$ fixed. For each setting we report the resulting worst-case survival function together with the induced net price of marginal coverage, $(1+\theta)S_Q(x) - g(S_{P^*}(x))$.

Results and economic interpretation

Effect of the safety loading. Figure 6.1 displays the worst-case survival functions (left) and the corresponding net prices (right) for the three values of the safety loading. With $\epsilon^{1/2} = 10$ the constraint is slack for $\theta = 0.1$ and binding for $\theta = 0.5$ and $\theta = 0.9$. The pattern is economically intuitive. A larger safety loading makes insurance more expensive, so the DM chooses to retain a greater share of his losses. Retaining more exposure gives him a larger stake in the tail, and the theory then pushes more probability mass upward, producing a riskier worst-case survival function. Thus the worst-case curves for $\theta = 0.5$ and $\theta = 0.9$ dominate that for $\theta = 0.1$. Because the constraint binds in the latter two cases, the entire ambiguity budget is exhausted and the two survival functions can no longer be ordered pointwise. They can, however, be compared through their means: writing X_1, X_2, X_3 for the worst-case losses under $\theta = 0.1, 0.5, 0.9$, one computes

$$\int_0^{5000} S_2^*(x)\,dx = 106.59 < 107.78 = \int_0^{5000} S_3^*(x)\,dx,$$

so that, in the terminology of Cheung, Chong and Yam [15],

$$X_1 \leq_{\mathrm{st}} X_2 \leq_{\mathrm{icx}} X_3,$$

where $\leq_{\mathrm{st}}$ and $\leq_{\mathrm{icx}}$ denote first-order stochastic dominance and the increasing convex order, respectively. The right panel shows the counterpart for the net price. A larger loading raises the net price of marginal coverage, which discourages the purchase of insurance, even as the DM's subjectively worst-case distribution grows more adverse.

Effect of risk aversion. Figure 6.2 repeats the analysis with the loading fixed at $\theta = 0.5$ and the risk-aversion parameter taking the values $r_1 \in \{0.6, 0.7, 0.8\}$. Under the common radius $\epsilon^{1/2} = 10$ the constraint binds in all three cases. As before the worst-case survival functions cannot be ordered pointwise, but their means now decrease

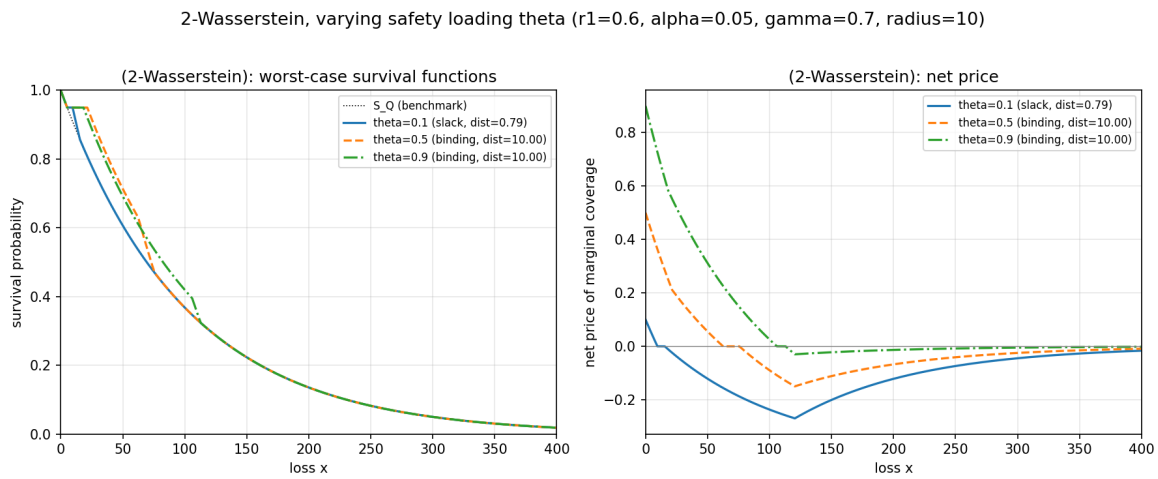

Figure 6.1: Under the 2-Wasserstein distance: (left) the worst-case survival functions for different values of the safety loading θ; (right) the corresponding net price of marginal coverage.

in r_1:

$$\int_0^{5000} S_1^*(x)\,dx = 106.59 > \int_0^{5000} S_2^*(x)\,dx = 105.99 > \int_0^{5000} S_3^*(x)\,dx = 105.53,$$

which by the same result of Cheung, Chong and Yam [15] gives

$$X_3 \leq_{\text{icx}} X_2 \leq_{\text{icx}} X_1.$$

At first sight this is counter-intuitive because a more risk-averse DM ends up with a *less* adverse worst-case distribution. The explanation lies in the interaction between risk aversion and the insurance decision. A larger r_1 lowers the net price of marginal coverage in the tail (right panel), so a more risk-averse DM demands more insurance and cedes more of his tail risk to the insurer. But the DM has no incentive to distort the distribution of losses he has already ceded, precisely those losses on the set $\mathcal{B}$ where the worst-case survival function coincides with the benchmark. Transferring more tail risk therefore leaves less exposure over which to shift probability mass, and the worst-case distribution becomes correspondingly milder. In short, the heavier hedging of a risk-averse DM more than offsets his stronger distaste for risk, and the tail is over-weighted *less*, not more.

Comparison with the L^2 distance

It is instructive to contrast the Wasserstein ball with the ball built from the L^2 distance between distribution functions, the metric studied by Boonen and Jiang [10]. For every

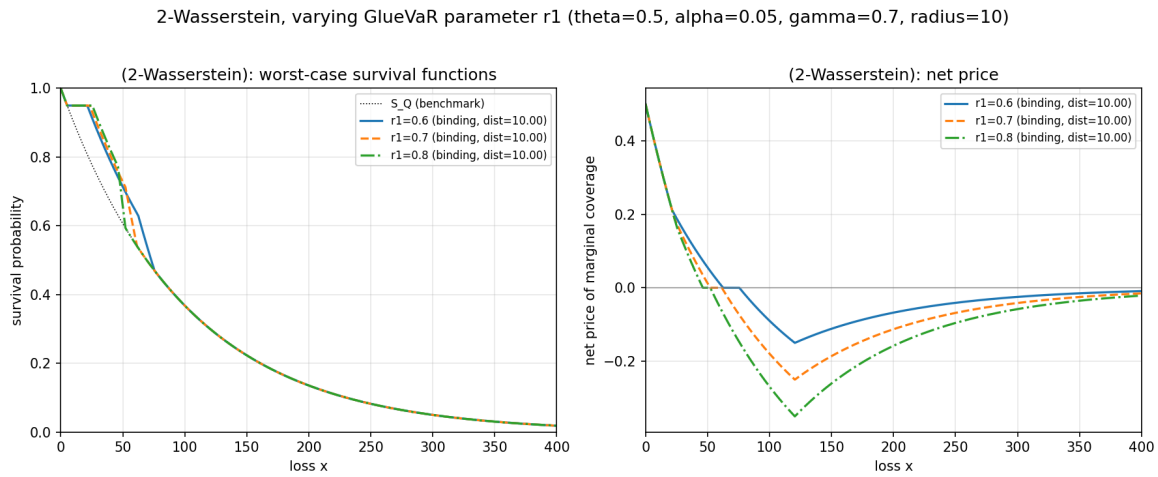

Figure 6.2: Under the 2-Wasserstein distance: (left) the worst-case survival functions for different values of the risk-aversion parameter r_1; (right) the corresponding net price of marginal coverage.

parameter combination considered above one verifies that

$$\max_{\substack{\theta\in\{0.1,0.5,0.9\}\\ r_1\in\{0.6,0.7,0.8\}}} \left(\int_0^M \left(\tilde{S}^*(x) - S_Q(x) \right)^2 dx \right)^{1/2} = 2.5779 < 10,$$

so that the slack-constraint solution $\tilde{S}^*$ of Theorem 6.4 is feasible for the L^2 ball in every case, and the constraint never binds there. The reason is structural: as shown by Boonen and Jiang [10], the L^2 distance between two distribution functions is never larger than the 2-Wasserstein distance, so for a common radius ϵ the L^2 ball is the larger of the two and leaves the DM more freedom in his choice of loss distribution. Consequently the DM selects a *riskier* worst-case survival function under the L^2 distance than under the 2-Wasserstein distance, and the net price of marginal coverage is correspondingly lower. The worst-case survival functions and net prices under the L^2 distance are shown in Figures 6.3 and 6.4. We stress, however, that this difference concerns only the worst-case distribution and the pricing it induces, because the DM buys no insurance on the region where he is indifferent between insuring and not. Consequently the optimal indemnity function itself, the stop-loss contract $I^*(x) = (x - x_1)_+$, is unchanged when one passes from the 2-Wasserstein distance to the L^2 distance.

6.5 Discussion and Concluding Remarks

This chapter has presented and numerically replicated the distributionally robust insurance model of Boonen and Jiang [11]. Starting from a decision maker who is uncertain about his loss distribution and hedges against a Wasserstein ball of candidate distributions, we summarized the analytical solution and reproduced the numerical

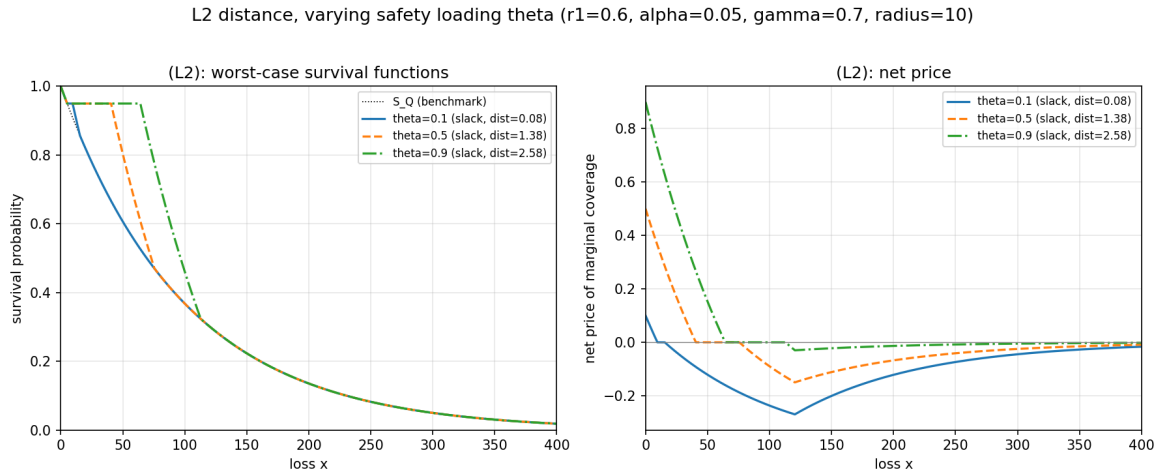

Figure 6.3: Under the L^2 distance: (left) the worst-case survival functions for different values of the safety loading θ; (right) the corresponding net price of marginal coverage.

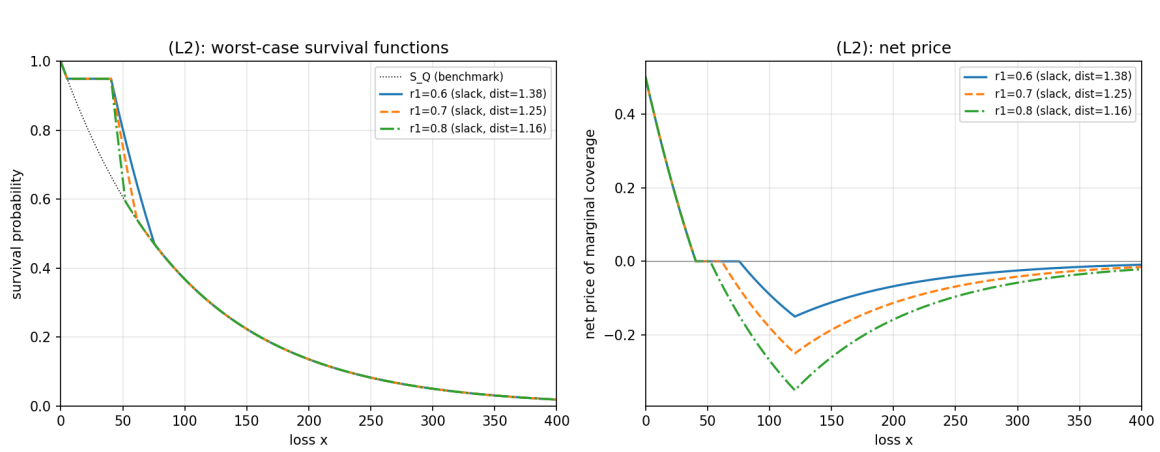

Figure 6.4: Under the L^2 distance: (left) the worst-case survival functions for different values of the risk-aversion parameter r_1; (right) the corresponding net price of marginal coverage.

study for the GlueV@R risk measure. The exercise confirmed the qualitative behavior reported in the paper and, at the same time, demonstrated that the theory developed earlier in this thesis is not merely a collection of convenient assumptions. Rather, the comonotone additivity and the Choquet structure of convex distortion risk measures are precisely the properties that make the problem tractable and admit an explicit solution. In this sense, the chapter serves as a concrete application of the preceding theory rather than as an isolated case study.

Beyond the technical results, the analysis yields an economically intelligible picture of how ambiguity shapes the optimal contract. The decision maker distorts the loss distribution only where he retains risk. For losses that are fully ceded to the insurer, the worst-case distribution coincides with the benchmark, since there is no incentive to guard against risk one does not bear. This mechanism also explains the seemingly

counter-intuitive observations from the numerical study, such as a more risk-averse decision maker over-weighting the tail *less*. Since a more risk-averse decision maker cedes more of the tail, less retained tail risk remains to be distorted. The worst-case distribution is thus best understood not in isolation but jointly with the optimal contract that determines which losses the decision maker keeps.

The framework is not without limitations, and these point to natural directions for further work. The analysis relies on the *convexity* of the distortion risk measure, which is what makes the minimax argument and the duality-based characterization of the worst case available. The practically important but non-convex Value at Risk falls outside its scope and would require different techniques. A second limitation concerns the treatment of ambiguity as exogenous. The radius of the Wasserstein ball is fixed in advance, whereas in reality the decision maker might reduce his ambiguity by investing in data collection, at a cost. Incorporating such endogenous ambiguity, and extending the results to non-convex preferences, are promising avenues that the original authors themselves identify as open problems.

Chapter 7

Conclusion and Outlook

This thesis followed a class of risk measures from its theoretical foundations to a practical application in insurance. Building on the convex-analytic framework developed in Chapter 2, we introduced monetary, convex, and coherent risk measures, their characterization through acceptance sets, and their robust representation as penalized worst-case expectations over families of probability measures. We then studied distortion risk measures via the Choquet integral and, from the opposite perspective, comonotone additive risk measures. These viewpoints culminated in the main theoretical result: on an atomless probability space, convex, law-invariant, comonotonic risk measures on L^∞, Choquet integrals with respect to concave distortions, and mixtures of Average Value at Risk are equivalent descriptions of the same class of coherent risk measures (Theorem 5.16).

Chapter 6 showed that this equivalence is not merely of theoretical interest. In the distributionally robust insurance model of Boonen and Jiang [11], each representation contributes to the solution. Comonotone additivity enables the minimax interchange, the Choquet and quantile representations transform the problem into a tractable convex optimization problem, and the concavity of the distortion ensures the convexity required for the dual characterization of the worst case. This yields a closed-form stop-loss contract together with an explicit worst-case distribution. Moreover, the solution has a clear economic interpretation because the decision maker distorts the loss distribution only over the retained part of the risk. Our numerical study for the GlueV@R risk measure illustrated this behavior across different safety loadings and risk-aversion parameters. More generally, the thesis demonstrates how abstract representation theorems can provide the mathematical foundation for solving and interpreting practical optimization problems in finance and insurance.

Several directions for future research remain. Within the application, two natural extensions arise from Chapter 6. The first is the study of non-convex risk measures such as Value at Risk, where the duality techniques developed in this thesis no longer apply. The second is the endogenization of the ambiguity radius, allowing the decision maker to reduce model uncertainty at a cost. On the theoretical side, it would be

natural to extend the analysis to spaces larger than L^∞, where the relationship between law-invariance, continuity, and robust representations becomes more subtle, as well as to dynamic, multi-period settings in which time consistency introduces additional structure. Finally, the comparison between the Wasserstein and L^2 ambiguity sets in the numerical study suggests a broader question: how does the geometry of the ambiguity set influence the worst-case distribution and the resulting insurance contract? This connects the present work to the rapidly growing literature on distributionally robust optimization and represents a promising direction for future research.

Appendix A

Probability-Theoretic Framework

This appendix provides a brief overview of the fundamental concepts and definitions from probability theory that are used throughout this thesis. In particular, we introduce the notions of distributions and expectations, discuss integrability and L^p spaces, and recall key integration results as well as changes of probability measures.

The results presented here are standard and mainly serve to fix notation and ensure consistency. For the sake of brevity, most statements are given without proof, as they can be found in classical textbooks on probability theory.

This chapter is intended as a reference section to support the main text, where these concepts are applied in the study of risk measures, Choquet integrals, and their representations.

A.1 Probability Spaces and Random Variables

This section recalls the basic building blocks of probability theory. We begin with measurable spaces and probability measures and then introduce random variables as measurable mappings between such spaces.

Definition A.1. [16] Let Ω be a set. A *σ-algebra* on Ω is a family $\mathcal{F} \subseteq \mathcal{P}(\Omega)$ of subsets of Ω satisfying the following properties:

- $\Omega \in \mathcal{F}$.
- If $A \in \mathcal{F}$, then $A^c \in \mathcal{F}$.
- If $(A_n)_{n\geq 1}$ is a countable family of elements of $\mathcal{F}$, then $\bigcup_{n\geq 1} A_n \in \mathcal{F}$.

The pair $(\Omega, \mathcal{F})$ is called a *measurable space*, and the elements of $\mathcal{F}$ are called *events*.

If $\mathcal{E}$ is an arbitrary family of subsets of Ω, the smallest σ-algebra containing $\mathcal{E}$ is called the *σ-algebra generated by $\mathcal{E}$*. In particular, the σ-algebra generated by the open intervals of $\mathbb{R}$ is called the *Borel σ-algebra* of $\mathbb{R}$, denoted $\mathcal{B}(\mathbb{R})$[16].

Having introduced measurable spaces, we now equip them with a probability measure in order to quantify uncertainty.

Definition A.2. [9] Let $(\Omega, \mathcal{F})$ be a measurable space. A *probability measure* on $(\Omega, \mathcal{F})$ is a map $\mathbb{P}\colon \mathcal{F} \to [0,1]$ satisfying the following axioms:

- $\mathbb{P}(\Omega) = 1$.
- If $A, A' \in \mathcal{F}$ are disjoint, then $\mathbb{P}(A \cup A') = \mathbb{P}(A) + \mathbb{P}(A')$.
- If $(A_n)_{n\in\mathbb{N}}$ is an increasing sequence of events in $\mathcal{F}$, then

$$\mathbb{P}\left(\bigcup_n A_n\right) = \lim_{n\to\infty} \mathbb{P}(A_n).$$

Note that the third axiom can equivalently be stated for decreasing sequences: if $(A_n)_{n\in\mathbb{N}}$ is a decreasing sequence of events in $\mathcal{F}$, then $\mathbb{P}(\bigcap_n A_n) = \lim_{n\to\infty} \mathbb{P}(A_n)$.

Combining these concepts yields the notion of a probability space.

Definition A.3. [16] A *probability space* is a triple $(\Omega, \mathcal{F}, \mathbb{P})$, where $(\Omega, \mathcal{F})$ is a measurable space and $\mathbb{P}$ is a probability measure on $(\Omega, \mathcal{F})$.

We say that an event $A \in \mathcal{F}$ holds *almost surely* (*a.s.*) if $\mathbb{P}(A^c) = 0$, i.e., if A occurs with probability one [9].

On a given probability space, random variables serve as the fundamental objects used to model uncertain quantities.

Definition A.4. [9, 16] Let $(\Omega, \mathcal{F})$ and $(E, \mathcal{E})$ be two measurable spaces. A function $X\colon \Omega \to E$ is called a *random variable* (from $(\Omega, \mathcal{F})$ to $(E, \mathcal{E})$) if it is *$\mathcal{F}/\mathcal{E}$-measurable*, i.e., if for every $A \in \mathcal{E}$,

$$X^{-1}(A) \;:=\; \{\omega \in \Omega : X(\omega) \in A\} \;\in\; \mathcal{F}.$$

A *real-valued random variable* on $(\Omega, \mathcal{F})$ is a random variable from $(\Omega, \mathcal{F})$ to $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. Equivalently, $X\colon \Omega \to \mathbb{R}$ is a real-valued random variable if and only if $\{\omega \in \Omega : X(\omega) \leq a\} \in \mathcal{F}$ for all $a \in \mathbb{R}$.

Definition A.5. [9] Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and let X, Y be two real-valued random variables on $(\Omega, \mathcal{F})$. We say that X and Y are *equal $\mathbb{P}$-almost surely*, written

$$X = Y \quad \mathbb{P}\text{-a.s.},$$

if

$$\mathbb{P}(\{\omega \in \Omega : X(\omega) \neq Y(\omega)\}) = 0.$$

In other words, X and Y agree outside a set of probability zero.

Definition A.6. [16] Let $(\Omega, \mathcal{F})$ be a measurable space and let $A \in \mathcal{F}$. The *indicator function* (or *characteristic function*) of A is the random variable $\mathbf{1}_A \colon \Omega \to \{0, 1\}$ defined by

$$\mathbf{1}_A(\omega) \;:=\; \begin{cases} 1 & \text{if } \omega \in A, \\ 0 & \text{if } \omega \notin A. \end{cases}$$

Note that $\mathbf{1}_A$ is $\mathcal{F}$-measurable whenever $A \in \mathcal{F}$, and that $\mathbb{P}(A) = \mathbb{E}[\mathbf{1}_A]$ for any probability measure $\mathbb{P}$ on $(\Omega, \mathcal{F})$.

Definition A.7. [9, 16] Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, $(E, \mathcal{E})$ a measurable space, and $X \colon (\Omega, \mathcal{F}) \to (E, \mathcal{E})$ a random variable. The *distribution* (or *law*) of X is the probability measure $\mathbb{P}_X$ on $(E, \mathcal{E})$ defined by

$$\mathbb{P}_X(A) \;=\; \mathbb{P}\Big(\{\omega \in \Omega : X(\omega) \in A\}\Big) \;=\; \mathbb{P}(X \in A), \qquad A \in \mathcal{E}.$$

One can verify directly that $\mathbb{P}_X$ is indeed a probability measure on $(E, \mathcal{E})$.

Remark A.8. The distribution $\mathbb{P}_X$ is also called the *image measure* of $\mathbb{P}$ under X, sometimes written $\mathbb{P} \circ X^{-1}$. It describes the probabilistic behavior of X independently of the underlying probability space $(\Omega, \mathcal{F}, \mathbb{P})$.

Definition A.9. [9] Let X be a real-valued random variable on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. The *cumulative distribution function* (CDF) of X is the function $F_X \colon \mathbb{R} \to [0, 1]$ defined by

$$F_X(x) \;=\; \mathbb{P}_X\Big((-\infty, x]\Big) \;=\; \mathbb{P}(X \le x), \qquad x \in \mathbb{R}.$$

Proposition A.10. *The cumulative distribution function F_X satisfies the following properties:*

- *F_X is increasing, with $\lim_{x \to -\infty} F_X(x) = 0$ and $\lim_{x \to +\infty} F_X(x) = 1$.*
- *F_X is right-continuous; its left-hand limit at any point x satisfies*

$$F_X(x^-) \;:=\; \lim_{y \nearrow x} F_X(y) \;=\; \mathbb{P}(X < x).$$

- *For every $x \in \mathbb{R}$,*

$$\mathbb{P}(X = x) \;=\; F_X(x) - F_X(x^-).$$

  In particular, F_X characterizes the distribution of X.

Remark A.11. Conversely, by Lemma A.19 below, every normalized increasing right-continuous function $F \colon \mathbb{R} \to [0, 1]$ is the CDF of some real-valued random variable. This gives a one-to-one correspondence between Borel probability measures on $\mathbb{R}$ and normalized increasing right-continuous functions.

We now develop the theory of quantile functions, following Föllmer and Schied.

Let $(a,b) \subset \mathbb{R}$ be a nonempty open interval and let $F : (a,b) \to \mathbb{R}$ be an increasing function. We define

$$c := \lim_{x \downarrow a} F(x) \in [-\infty, \infty), \qquad d := \lim_{x \uparrow b} F(x) \in (-\infty, \infty].$$

We seek a function $q\colon (c,d) \to (a,b)$ acting as a generalized inverse of F — generalized because F may have jumps and flat stretches, so that a genuine inverse need not exist.

A particularly important special case arises when $F : \mathbb{R} \to [0,1]$ is increasing and right-continuous. Such an F is nothing but the cumulative distribution function of a real-valued random variable X on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$, i.e., $F(x) = \mathbb{P}(X \le x)$. This probabilistic setting will be assumed throughout. We write $F(x-) := \lim_{y \uparrow x} F(y)$.

Conversely, every cumulative distribution function of a real-valued random variable satisfies these properties.

Definition A.12. A function $q\colon (c,d) \to (a,b)$ is called an *inverse function* for F if

$$F(q(s)-) \;\le\; s \;\le\; F(q(s)+) \qquad \text{for all } s \in (c,d).$$

The *lower* and *upper quantile functions* are defined respectively by

$$q^-(s) \;:=\; \sup\Big\{x \in (a,b) \mid F(x) < s\Big\} \qquad \text{and} \qquad q^+(s) \;:=\; \inf\Big\{x \in (a,b) \mid F(x) > s\Big\}.$$

Lemma A.13. *A function $q\colon (c,d) \to (a,b)$ is an inverse function for F if and only if*

$$q^-(s) \;\le\; q(s) \;\le\; q^+(s) \qquad \textit{for all } s \in (c,d).$$

In particular, q^- and q^+ are themselves inverse functions. Moreover, q^- is left-continuous, q^+ is right-continuous, and every inverse function q is increasing and satisfies

$$q(s^-) = q^-(s) \qquad \textit{and} \qquad q(s^+) = q^+(s) \qquad \textit{for all } s \in (c,d).$$

In particular, all inverse functions coincide except on at most countably many points of (c,d).

Lemma A.14. *Every inverse function q satisfies*

$$q(s) \;\searrow a \quad \textit{as } s \downarrow c, \qquad \textit{and} \qquad q(s) \;\nearrow b \quad \textit{as } s \uparrow d.$$

Remark A.15. The left- and right-continuous inverse functions can also be expressed as

$$q^-(s) = \inf\{x \in (a,b) \mid F(x) \ge s\} \quad \text{and} \quad q^+(s) = \sup\{x \in (a,b) \mid F(x) \le s\}.$$

To justify these representations, note that $q^-(s)$ is bounded above by the corresponding infimum on the right-hand side. Conversely, for any $\varepsilon > 0$ such that $q^-(s) + \varepsilon < b$, one has $F(q^-(s) + \varepsilon) > s$. Taking the infimum over all $x \in (a, b)$ with $F(x) \geq s$ and letting $\varepsilon \downarrow 0$ yields the first identity. The argument for q^+ is analogous.

Lemma A.16. *A function q is an inverse function for F if and only if it is an inverse function for any other function $\tilde{F}\colon (a, b) \to (c, d)$ satisfying*

$$F(x^-) \;\leq\; \tilde{F}(x) \;\leq\; F(x^+) \qquad \text{for all } x \in \mathbb{R}.$$

In particular, the lower quantile function q^- admits the following equivalent representations:

$$\begin{aligned} q^-(t) &= \sup\Big\{x \in (a,b) \mid F(x^-) < t\Big\} = \sup\Big\{x \in (a,b) \mid F(x^+) < t\Big\} \\ &= \inf\Big\{x \in (a,b) \mid F(x^-) \geq t\Big\} = \inf\Big\{x \in (a,b) \mid F(x^+) \geq t\Big\}. \end{aligned}$$

Definition A.17. Let F be a cumulative distribution function. An inverse function $q : (0, 1) \to \mathbb{R}$ of F is called a *quantile function*. That is, q is a function such that

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

The left- and right-continuous inverses,

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

are called the *lower* and *upper quantile functions*.

Lemma A.18. *Let $X \in L^1(\Omega, \mathcal{F}, \mathbb{P})$ and $\lambda \in (0, 1)$. Define the convex function $f_\lambda : \mathbb{R} \to \mathbb{R}$ by*

$$f_\lambda(c) \;:=\; \lambda\,\mathbb{E}\Big[(X - c)^+\Big] \;+\; (1 - \lambda)\,\mathbb{E}\Big[(X - c)^-\Big]. \tag{A.1}$$

Then a real number $q \in \mathbb{R}$ is a λ-quantile of X if and only if q is a minimizer of f_λ.

For a detailed discussion of the expectation operator, we refer to the next section.

Proof. Throughout the proof we use the identities $(x)^+ = \max(x, 0)$ and $(x)^- = \max(-x, 0) = (-x)^+$, together with $x = (x)^+ - (x)^-$.

Step 1: f_λ is finite, convex, and admits one-sided derivatives. Since $|(X - c)^\pm| \leq |X| + |c|$ and $X \in L^1$, the function f_λ is finite on $\mathbb{R}$. Both $c \mapsto (X - c)^+$ and $c \mapsto (X - c)^-$ are convex, and convexity is preserved by taking expectations and non-negative linear combinations; hence f_λ is convex on $\mathbb{R}$. Convex functions on $\mathbb{R}$ admit left- and right-hand derivatives at every interior point, which we denote by $(f_\lambda)'_-(c)$ and $(f_\lambda)'_+(c)$.

Step 2: Computation of the one-sided derivatives. For fixed $\omega \in \Omega$, the maps $c \mapsto (X(\omega) - c)^+$ and $c \mapsto (X(\omega) - c)^-$ are piecewise linear with kink at $c = X(\omega)$, and their one-sided derivatives are

$$\frac{d}{dc^+}(X-c)^+ = -\mathbf{1}_{\{X>c\}}, \qquad \frac{d}{dc^-}(X-c)^+ = -\mathbf{1}_{\{X\geq c\}},$$

$$\frac{d}{dc^+}(X-c)^- = \mathbf{1}_{\{X\leq c\}}, \qquad \frac{d}{dc^-}(X-c)^- = \mathbf{1}_{\{X<c\}}.$$

Since the difference quotients are uniformly bounded by 1 in absolute value and $X \in L^1$, dominated convergence allows differentiation under the expectation, yielding

$$(f_\lambda)'_+(c) = -\lambda\,\mathbb{P}[X > c] + (1-\lambda)\,\mathbb{P}[X \leq c] \;=\; \mathbb{P}[X \leq c] - \lambda, \tag{A.2}$$

$$(f_\lambda)'_-(c) = -\lambda\,\mathbb{P}[X \geq c] + (1-\lambda)\,\mathbb{P}[X < c] \;=\; \mathbb{P}[X < c] - \lambda, \tag{A.3}$$

where in each line we used $\mathbb{P}[X > c] = 1-\mathbb{P}[X \leq c]$ respectively $\mathbb{P}[X \geq c] = 1-\mathbb{P}[X < c]$ to simplify.

Step 3: Characterization of minimizers via subdifferential. By convexity of f_λ, a point $q \in \mathbb{R}$ is a minimizer if and only if 0 belongs to the subdifferential of f_λ at q, which on $\mathbb{R}$ amounts to

$$(f_\lambda)'_-(q) \leq 0 \leq (f_\lambda)'_+(q). \tag{A.4}$$

Substituting (A.2) and (A.3) into (A.4), q is a minimizer of f_λ if and only if

$$\mathbb{P}[X < q] \;\leq\; \lambda \;\leq\; \mathbb{P}[X \leq q]. \tag{A.5}$$

Step 4: Conclusion. The two inequalities in (A.5) are precisely the defining conditions of a λ-quantile of X: indeed, $\mathbb{P}[X < q] \leq \lambda$ is equivalent to $\mathbb{P}[X \geq q] \geq 1 - \lambda$, and $\mathbb{P}[X \leq q] \geq \lambda$ is the standard lower-tail condition. Hence q minimizes f_λ if and only if q is a λ-quantile of X. □

Lemma A.19. *Let U be a random variable uniformly distributed on $(0,1)$, that is, $\mathbb{P}(U \leq s) = s$ for all $s \in (0,1)$. If q is a quantile function for a normalized increasing right-continuous function $F\colon \mathbb{R} \to [0,1]$, then the random variable $X := q(U)$ has CDF F.*

In particular, this establishes a one-to-one correspondence between Borel probability measures μ on $\mathbb{R}$ and normalized increasing right-continuous functions $F_\mu(x) = \mu((-\infty, x])$.

Remark A.20. Lemma A.19 yields a one-to-one correspondence

$$F(x) = \mu\Big((-\infty, x]\Big)$$

between all Borel probability measures μ on $\mathbb{R}$ and all normalized increasing right-

continuous functions $F\colon \mathbb{R} \to [0,1]$. This correspondence extends naturally to the setting of positive Radon measures: every increasing right-continuous function on a nonempty open interval $I \subseteq \mathbb{R}$ gives rise to a corresponding positive Radon measure on I.

Lemma A.21. *Let $X = f(Y)$ for an increasing function f and let q_Y be a quantile function for Y. Then $f(q_Y(t))$ is a quantile function for X. In particular, for any quantile function q_X of X,*

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

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 increasing, then

$$\begin{aligned} F_X(f(q_Y(t))) &= \mathbb{P}[f(Y) \le f(q_Y(t))] \ge \mathbb{P}[Y \le q_Y(t)] \\ &= F_Y(q_Y(t)) \ge t \ge F_Y(q_Y(t)-) \\ &= \mathbb{P}[Y < q_Y(t)] \ge \mathbb{P}[f(Y) < f(q_Y(t))] \\ &= F_X(f(q_Y(t)-)). \end{aligned}$$

Hence, $f(q_Y(t))$ is a quantile function for $X = f(Y)$. If f is decreasing, then $-f$ is increasing and the preceding argument yields that $-f(q_Y(t))$ is a quantile function for $-X$. Thus, the assertion will follow once we have shown that $-q_Y(1-t)$ is a quantile function for $-Y$. Indeed, we have

$$F_{-Y}(-q_Y(1-t)) = \mathbb{P}[-Y \le -q_Y(1-t)] = 1 - F_Y(q_Y(1-t)-) \ge t,$$

and a similar argument yields $F_{-Y}(-q_Y(1-t)) \le t$. □

Lemma A.22. *Let X be a real-valued random variable with CDF F_X and quantile function q_X, and suppose $\mathbb{E}[|X|] < \infty$. Define the convex function*

$$\Psi(x) \;:=\; \int_{-\infty}^{x} F_X(z)\,dz \;=\; \mathbb{E}[(x-X)^+], \qquad x \in \mathbb{R}.$$

Then its Fenchel–Legendre transform is given by

$$\Psi^*(y) \;:=\; \sup_{x\in\mathbb{R}}\bigl(xy - \Psi(x)\bigr) \;=\; \begin{cases} \displaystyle\int_0^y q_X(t)\,dt & \text{if } 0 \le y \le 1, \\ +\infty & \text{otherwise.} \end{cases}$$

Moreover, for $0 < y < 1$, the supremum is attained at x if and only if x is a y-quantile of X.

The function Ψ^* is the Fenchel–Legendre transform of Subsection 2.5.1, applied to the convex function $\Psi(x) = \mathbb{E}\big[(x - X)^+\big]$; the identity above makes the abstract conjugate of Definition 2.28 fully explicit in terms of the quantile function of X, and the characterization of the maximiser is the case of equality in the Fenchel–Young inequality (Proposition 2.29).

Proposition A.23. *For any probability space $(\Omega, \mathcal{F}, \mathbb{P})$, the following conditions are equivalent:*

(a) $(\Omega, \mathcal{F}, \mathbb{P})$ *is atomless.*

(b) There exists an i.i.d. sequence $X_1, X_2, \ldots$ of random variables with symmetric Bernoulli distribution, $\mathbb{P}(X_1 = 0) = \mathbb{P}(X_1 = 1) = \frac{1}{2}$.

(c) For any $\mu \in \mathcal{M}_1(\mathbb{R})$, there exist i.i.d. random variables $Y_1, Y_2, \ldots$ with common distribution μ.

(d) $(\Omega, \mathcal{F}, \mathbb{P})$ *supports a random variable with a continuous distribution.*

A.2 Expectation and Integrable Random Variables

Throughout this section, let $(\Omega, \mathcal{F}, \mathbb{P})$ be a fixed probability space. We now briefly recall basic notions from probability theory that are used throughout the thesis. In particular, we introduce the expectation, integrability, and the spaces $L^p(\Omega, \mathcal{F}, \mathbb{P})$, with special emphasis on L^∞, which will serve as the underlying domain for risk measures.

Definition A.24. [16] Let X be a real-valued random variable on $(\Omega, \mathcal{F}, \mathbb{P})$.

- *Simple case.* If X takes finitely many distinct values $a_1, \ldots, a_k$ with $\mathbb{P}(\{a_1, \ldots, a_k\}) = 1$, the *expectation* (or *mean*) of X is

$$\mathbb{E}[X] \;:=\; \sum_{j=1}^{k} a_j \, \mathbb{P}(X = a_j).$$

- *Non-negative case.* If $X \geq 0$ $\mathbb{P}$-a.s., let $\mathcal{E}_X$ denote the set of simple random variables that are $\mathbb{P}$-a.s. bounded above by X. Then

$$\mathbb{E}[X] \;:=\; \sup_{Y \in \mathcal{E}_X} \mathbb{E}[Y] \;\in\; [0, +\infty].$$

- *General case.* Decompose $X = X^+ - X^-$, where $X^+ := \max(X, 0)$ and $X^- := \max(-X, 0)$ are both non-negative. If $\mathbb{E}[X^+] < \infty$ and $\mathbb{E}[X^-] < \infty$, then X is called *integrable* and one sets

$$\mathbb{E}[X] \;:=\; \mathbb{E}[X^+] - \mathbb{E}[X^-].$$

Remark A.25. [16] The abstract definition of the expectation unifies the classical formulas used in concrete situations. In particular, if X is a discrete random variable, then

$$\mathbb{E}[X] = \sum_i x_i \, \mathbb{P}(X = x_i),$$

whenever the series converges absolutely. If X admits a density f with respect to the Lebesgue measure, then

$$\mathbb{E}[X] = \int_{\mathbb{R}} x f(x) \, dx,$$

provided the integral is finite. Both expressions are special cases of the Lebesgue integral

$$\mathbb{E}[X] = \int_{\Omega} X \, d\mathbb{P}.$$

Proposition A.26. *[9, 16] Let X, Y be integrable random variables on $(\Omega, \mathcal{F}, \mathbb{P})$ and let $a, b \in \mathbb{R}$.*

- *Linearity. $aX + bY$ is integrable and $\mathbb{E}[aX + bY] = a\,\mathbb{E}[X] + b\,\mathbb{E}[Y]$.*
- *Monotonicity. If $X \le Y$ $\mathbb{P}$-a.s., then $\mathbb{E}[X] \le \mathbb{E}[Y]$.*
- *Positivity. If $X \ge 0$ $\mathbb{P}$-a.s., then $\mathbb{E}[X] \ge 0$.*
- *Absolute value bound. $|\mathbb{E}[X]| \le \mathbb{E}[|X|]$.*
- *Constants. For any $c \in \mathbb{R}$, $\mathbb{E}[c] = c$.*

Before introducing function spaces, we first recall the notion of integrability, which ensures that the expectation is well-defined.

Definition A.27. A real-valued random variable X on $(\Omega, \mathcal{F}, \mathbb{P})$ is *integrable* if

$$\mathbb{E}[|X|] \;=\; \mathbb{E}[X^+] + \mathbb{E}[X^-] \;<\; \infty.$$

The collection of all (equivalence classes of $\mathbb{P}$-a.s. equal) integrable random variables is denoted

$$L^1(\Omega, \mathcal{F}, \mathbb{P}) \quad \text{or simply } L^1(\mathbb{P}).$$

Equipped with the norm $\|X\|_1 := \mathbb{E}[|X|]$, the space $L^1(\mathbb{P})$ is a *Banach space*.

Definition A.28. For $p \in [1, \infty)$, the space $L^p(\Omega, \mathcal{F}, \mathbb{P})$ consists of all (equivalence classes of $\mathbb{P}$-a.s. equal) real-valued random variables X satisfying

$$\mathbb{E}\Big[|X|^p\Big] \;<\; \infty.$$

It is a Banach space under the norm

$$\|X\|_p \;:=\; \Big(\mathbb{E}[|X|^p]\Big)^{1/p}.$$

Proposition A.29. *For $1 \le p < q \le \infty$ one has the continuous embedding*

$$L^q(\mathbb{P}) \;\subseteq\; L^p(\mathbb{P}).$$

In particular, every $X \in L^q(\mathbb{P})$ is integrable.

Remark A.30. The most important special case is $L^2(\mathbb{P})$: every $X \in L^2(\mathbb{P})$ has a well-defined *variance*

$$\operatorname{Var}(X) \;:=\; \mathbb{E}\Big[(X - \mathbb{E}[X])^2\Big] \;=\; \mathbb{E}[X^2] - \Big(\mathbb{E}[X]\Big)^2 \;\ge\; 0.$$

Of specific importance for this thesis is the space L^∞, which we introduce next.

Definition A.31. A random variable X is *essentially bounded* if there exists a constant $c > 0$ such that $|X| \le c$ $\mathbb{P}$-a.s. The *essential supremum of X* is the smallest such constant:

$$\operatorname{ess\,sup} X \;:=\; \inf\Big\{c \in \mathbb{R} : \mathbb{P}(X > c) = 0\Big\} \;\in\; (-\infty, +\infty].$$

The space of all (equivalence classes of) essentially bounded random variables is

$$L^\infty(\Omega, \mathcal{F}, \mathbb{P}) \quad \text{or simply } L^\infty(\mathbb{P}),$$

with norm $\|X\|_\infty := \operatorname{ess\,sup} |X|$. It is a Banach space, and $L^\infty(\mathbb{P}) \subseteq L^p(\mathbb{P})$ for all $p \in [1, \infty)$.

Remark A.32. The spaces introduced above, in particular $L^\infty(\Omega, \mathcal{F}, \mathbb{P})$, will serve as the underlying domain for the risk measures studied in this thesis. Moreover, expectation functionals of the form $\mathbb{E}_Q[X]$ will play a central role in the dual representation of risk measures.

The notion of essential supremum extends naturally from a single random variable to an arbitrary *family* of random variables.

Theorem A.33. *Let Φ be any set of random variables on $(\Omega, \mathcal{F}, \mathbb{P})$.*

(a) There exists a $[-\infty, +\infty]$-valued random variable φ^* with the following two properties:

- *$\varphi^* \ge \varphi$ $\mathbb{P}$-a.s. for all $\varphi \in \Phi$.*
- *If ψ is any random variable satisfying $\psi \ge \varphi$ $\mathbb{P}$-a.s. for all $\varphi \in \Phi$, then $\varphi^* \le \psi$ $\mathbb{P}$-a.s.*

(b) If, in addition, Φ is directed upward*, i.e., for every $\varphi, \tilde{\varphi} \in \Phi$ there exists $\psi \in \Phi$ with $\psi \ge \varphi \vee \tilde{\varphi}$ $\mathbb{P}$-a.s. then there exists an increasing sequence $\varphi_1 \le \varphi_2 \le \cdots$ in Φ such that $\varphi^* = \lim_n \varphi_n$ $\mathbb{P}$-a.s.*

Definition A.34. The random variable φ^* in Theorem A.33 is called the *essential supremum* of Φ with respect to $\mathbb{P}$, written

$$\operatorname{ess\,sup}_{\varphi\in\Phi}\varphi \;:=\; \varphi^*.$$

The *essential infimum* of Φ is defined by

$$\operatorname{ess\,inf}_{\varphi\in\Phi}\varphi \;:=\; -\operatorname{ess\,sup}_{\varphi\in\Phi}(-\varphi).$$

Remark A.35. For a given random variable X, let Φ be the set of all constants c such that $\mathbb{P}[X > c] > 0$. Then

$$\operatorname{ess\,sup} X := \sup\Phi$$

is the smallest constant $c \le +\infty$ such that $X \le c$ $\mathbb{P}$-a.s., and is called the *essential supremum* of X with respect to $\mathbb{P}$. The *essential infimum* of X is defined as

$$\operatorname{ess\,inf} X := \operatorname{ess\,sup}(-X).$$

A.3 Important Integration Results

Throughout this section, $(\Omega, \mathcal{A}, \mu)$ denotes a measure space and all functions are assumed to be measurable.

We collect a number of fundamental results from integration theory that will be used repeatedly throughout the thesis. In particular, the following theorems provide the key tools for interchanging limits and integrals, which is essential in the analysis of risk measures and their representations.

We begin with the classical Monotone Convergence Theorem.

Theorem A.36 (Monotone Convergence Theorem)**.** *[44] Let $\{f_n\}_{n\in\mathbb{N}}$ be a sequence of measurable functions $f_n : \Omega \to [0,\infty]$ such that*

- *$f_n(x) \le f_{n+1}(x)$ for all $n \in \mathbb{N}$ and μ-almost all $x \in \Omega$, and*
- *$f_n(x) \to f(x)$ for μ-almost all $x \in \Omega$.*

Then

$$\lim_{n\to\infty}\int f_n\,d\mu \;=\; \int f\,d\mu.$$

Remark A.37. [44] The functions f_n do not need to be integrable; the integrals are well-defined (possibly $+\infty$) since the f_n are non-negative. An equivalent formulation (sometimes called the *Beppo Levi theorem*) states that for any sequence of measurable functions $f_k : \Omega \to [0,\infty]$,

$$\int \sum_{k=1}^{\infty} f_k\,d\mu \;=\; \sum_{k=1}^{\infty}\int f_k\,d\mu.$$

Lemma A.38 (Fatou). *[44] Let $\{f_n\}_{n\in\mathbb{N}}$ be a sequence of measurable functions $f_n : \Omega \to [0,\infty]$. Then*

$$\int \liminf_{n\to\infty} f_n \, d\mu \;\le\; \liminf_{n\to\infty} \int f_n \, d\mu.$$

Remark A.39. [44] The inequality can be strict. Fatou's Lemma requires only nonnegativity and imposes no domination condition. It is the key tool in proving the Dominated Convergence Theorem, which is indispensable in many applications.

Theorem A.40 (Dominated Convergence Theorem). *[44] Let $f_n : \Omega \to [-\infty,\infty]$ be a sequence of measurable functions and let $g : \Omega \to [0,\infty]$ be integrable. Suppose*

- $\lim_{n\to\infty} f_n(x) = f(x)$ *for μ-almost all $x \in \Omega$, and*
- $|f_n(x)| \le g(x)$ *for μ-almost all $x \in \Omega$.*

Then

$$\lim_{n\to\infty} \int f_n \, d\mu \;=\; \int f \, d\mu.$$

We next consider integration on product spaces. The following result, known as Tonelli's theorem, ensures that iterated integrals are well-defined for nonnegative functions.

Theorem A.41 (Tonelli). *[44] Let $(X,\mathcal{A},\mu)$ and $(Y,\mathcal{B},\nu)$ be σ-finite measure spaces, and let $f : X \times Y \to [0,\infty]$ be $\mathcal{A}\times\mathcal{B}$-measurable. Then:*

(a) The section functions $x \mapsto \int_Y f_x \, d\nu$ and $y \mapsto \int_X f^y \, d\mu$ are measurable (with respect to $\mathcal{A}$ and $\mathcal{B}$, respectively).

(b) The iterated integrals equal the product integral:

$$\int_{X\times Y} f \, d(\mu\times\nu) \;=\; \int_X\!\int_Y f(x,y)\, d\nu(y)\, d\mu(x) \;=\; \int_Y\!\int_X f(x,y)\, d\mu(x)\, d\nu(y).$$

Tonelli's theorem guarantees that iterated integrals are well-defined for nonnegative functions. For integrable functions, this result extends to the more general Fubini theorem, which allows for the interchange of the order of integration.

Theorem A.42 (Fubini). *[44] Let $(X,\mathcal{A},\mu)$ and $(Y,\mathcal{B},\nu)$ be σ-finite measure spaces, and let $f : X \times Y \to [-\infty,\infty]$ be $\mathcal{A}\times\mathcal{B}$-measurable and integrable with respect to $\mu\times\nu$. Then:*

(a) $f_x(\cdot) = f(x,\cdot)$ is ν-integrable for μ-almost all $x \in X$, and $f^y(\cdot) = f(\cdot,y)$ is μ-integrable for ν-almost all $y \in Y$.

(b) Setting

$$I_f(x) = \begin{cases} \displaystyle\int_Y f(x,y)\, d\nu(y) & \text{if } f_x \text{ is } \nu\text{-integrable,} \\ 0 & \text{otherwise,} \end{cases}$$

$$J_f(x) = \begin{cases} \int_X f(x,y)\,d\mu(x) & \text{if } f^y \text{ is } \mu\text{-integrable,} \\ 0 & \text{otherwise,} \end{cases}$$

we have $I_f \in L^1(X, \mathcal{A}, \mu)$ *and* $I_f \in L^1(Y, \mathcal{B}, \nu)$ *and*

$$\int_{X\times Y} f\,d(\mu\times\nu) \;=\; \int_X I_f\,d\mu \;=\; \int_Y J_f\,d\nu.$$

The following inequality will be particularly useful when dealing with convexity properties of risk measures.

Theorem A.43 (Jensen's Inequality)**.** *[9] Let* $(\Omega, \mathcal{A}, \mu)$ *be a probability space (i.e.* $\mu(\Omega) = 1$*), let* $f \in L^1(\Omega, \mathcal{A}, \mu)$ *with values in an interval* $I \subseteq \mathbb{R}$*, and let* $\varphi : I \to \mathbb{R}$ *be a convex function. Then*

$$\varphi\Big(\int f\,d\mu\Big) \;\le\; \int(\varphi\circ f)\,d\mu.$$

Finally, we recall Markov's inequality, which provides a simple but important bound for tail probabilities in terms of expectations.

Proposition A.44 (Markov's inequality)**.** *[9] Let* $X \in L^1(\Omega, \mathcal{F}, \mathbb{P})$*. Then for every* $a > 0$*,*

$$\mathbb{P}(|X| \ge a) \le \frac{1}{a}\,\mathbb{E}[|X|].$$

The results collected in this section form the technical backbone for many of the arguments used throughout the thesis. In particular, they will be applied repeatedly in the study of convergence properties and representation results of risk measures.

A.4 Change of Probability Measure

Throughout this section, $(\Omega, \mathcal{F}, \mathbb{P})$ denotes a probability space and when a filtration is involved we have $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\ge0}, \mathbb{P})$. The aim of this section is to recall the Radon–Nikodym formalism, which lies at the heart of many arguments throughout the thesis. In particular, changes of probability measure are essential in the study of risk measures and their representations.

Given $Z \in L^1(\Omega, \mathcal{F}, \mathbb{P})$ with $Z \ge 0$ and $\mathbb{E}_{\mathbb{P}}[Z] = 1$, the prescription

$$Q(A) \;:=\; \int_A Z\,d\mathbb{P} \;=\; \mathbb{E}_{\mathbb{P}}[Z\mathbf{1}_A], \qquad A \in \mathcal{F}, \tag{A.6}$$

defines a probability measure Q on $(\Omega, \mathcal{F})$. The random variable Z is called the *density* (or *Radon–Nikodym derivative*) of Q with respect to $\mathbb{P}$, and we write

$$Q = Z\cdot\mathbb{P}, \qquad dQ = Z\,d\mathbb{P}, \qquad Z = \frac{dQ}{d\mathbb{P}}.$$

[17]

Definition A.45. [17] Let $\mathbb{P}$ and Q be two probability measures on a measurable space $(\Omega, \mathcal{F})$. We say that Q is *absolutely continuous with respect to* $\mathbb{P}$ on $\mathcal{F}$, and write $Q \ll \mathbb{P}$, if

$$\mathbb{P}(A) = 0 \implies Q(A) = 0 \quad \text{for all } A \in \mathcal{F}.$$

If both $Q \ll \mathbb{P}$ and $\mathbb{P} \ll Q$ hold, then $\mathbb{P}$ and Q are called *equivalent*, and we write $Q \sim \mathbb{P}$.

The next theorem is the key characterization linking Definition A.45 to formula (A.6).

Theorem A.46 (Radon–Nikodym)**.** *[17]*

(a) $Q \ll \mathbb{P}$ on $\mathcal{F}$ if and only if there exists a density of Q with respect to $\mathbb{P}$.

(b) $Q \sim \mathbb{P}$ if and only if there exists a density of Q with respect to $\mathbb{P}$ that is strictly positive $\mathbb{P}$-a.s.

The density acts as a multiplicative weight when integrating under the new measure:

Proposition A.47. *[17] Let $Z = dQ/d\mathbb{P}$. Then for every $Y \in L^1(\Omega, \mathcal{F}, Q)$,*

$$\mathbb{E}_Q[Y] \;=\; \mathbb{E}_{\mathbb{P}}[ZY].$$

We now extend the previous results to filtered probability spaces, which allows us to describe changes of measure in a dynamic framework. Suppose that $Q \ll \mathbb{P}$ on $\mathcal{F}$. Since $\mathcal{F}_t \subseteq \mathcal{F}$ for every $t \geq 0$, absolute continuity is inherited on each $\mathcal{F}_t$, and the Radon–Nikodym theorem yields a family of densities

$$D_t \;:=\; \left.\frac{dQ}{d\mathbb{P}}\right|_{\mathcal{F}_t}, \qquad t \geq 0.$$

The process $D = (D_t)_{t \geq 0}$ is called the *density process* of Q with respect to $\mathbb{P}$ and will be useful when working with conditional expectations under different probability measures [17].

The following result describes how Radon–Nikodym derivatives behave under successive changes of measure.

Proposition A.48 (Chain rule for Radon–Nikodym derivatives)**.** *Let $\mathbb{P}$, Q, and R be probability measures on $(\Omega, \mathcal{F})$ with $Q \ll \mathbb{P}$.*

(a) If $R \ll Q$, then $R \ll \mathbb{P}$ and

$$\frac{dR}{d\mathbb{P}} \;=\; \frac{dR}{dQ} \cdot \frac{dQ}{d\mathbb{P}} \qquad \mathbb{P}\text{-a.s.}$$

(b) If $R \ll \mathbb{P}$, then $R \ll Q$ if and only if $dQ/d\mathbb{P} > 0$ holds $\mathbb{R}$-a.s., and in this case

$$\frac{dR}{dQ} \;=\; \frac{dR}{d\mathbb{P}} \cdot \left(\frac{dQ}{d\mathbb{P}}\right)^{-1} \qquad Q\text{-a.s.}$$

In particular, choosing $R = \mathbb{P}$ in (b) yields

$$Q \sim \mathbb{P} \iff \frac{dQ}{d\mathbb{P}} > 0 \quad \mathbb{P}\text{-}a.s., \qquad with \qquad \frac{d\mathbb{P}}{dQ} = \left(\frac{dQ}{d\mathbb{P}}\right)^{-1}.$$

The results collected in this section provide the fundamental tools for handling changes of probability measure. In particular, they will be used repeatedly in the analysis of expectations under different measures and in the derivation of representation results for risk measures.

Appendix B

Additional Material on Risk Measures

B.1 Supplementary Results

Proposition B.1. *A monetary risk measure is quasi-convex if and only if it is convex.*

Proof. Let ρ be a monetary risk measure. If ρ is convex, then for all $X, Y \in \mathcal{X}$ and $\lambda \in [0,1]$,

$$\rho(\lambda X + (1-\lambda)Y) \leq \lambda\rho(X) + (1-\lambda)\rho(Y) \leq \max\{\rho(X), \rho(Y)\},$$

which shows that ρ is quasi-convex.

Conversely, suppose that ρ is quasi-convex. Let $X, Y \in \mathcal{X}$ and set $a := \rho(X)$ and $b := \rho(Y)$. By cash invariance, we have $\rho(X + a) = 0$ and $\rho(Y + b) = 0$. Applying quasi-convexity yields

$$\rho\Big(\lambda(X+a) + (1-\lambda)(Y+b)\Big) \leq 0.$$

Rewriting the argument gives

$$\lambda(X+a) + (1-\lambda)(Y+b) = \lambda X + (1-\lambda)Y + \lambda a + (1-\lambda)b.$$

Using cash invariance once more, we obtain

$$\rho(\lambda X + (1-\lambda)Y) \leq \lambda a + (1-\lambda)b.$$

Substituting back $a = \rho(X)$ and $b = \rho(Y)$ shows that ρ is convex. □

Remark B.2. The equivalence between quasi-convexity and convexity for monetary risk measures is a direct consequence of cash invariance. While quasi-convexity only controls the risk of convex combinations up to the worst-case level, cash invariance

allows for a normalization of positions to a common benchmark level. This effectively restores linear control and upgrades quasi-convexity to full convexity. Without cash invariance, quasi-convexity would in general be strictly weaker than convexity.

Proposition B.3. *Let ρ be a convex and normalized risk measure. Then for all $X \in \mathcal{X}$ and all $\lambda \geq 0$,*

- *if $0 \leq \lambda \leq 1$, then $\rho(\lambda X) \leq \lambda \rho(X)$,*
- *if $\lambda \geq 1$, then $\rho(\lambda X) \geq \lambda \rho(X)$.*

Proof. Since ρ is normalized, $\rho(0) = 0$. For $0 \leq \lambda \leq 1$, convexity applied to X and 0 yields $\rho(\lambda X) = \rho(\lambda X + (1-\lambda)0) \leq \lambda \rho(X) + (1-\lambda)\rho(0) = \lambda \rho(X)$. If $\lambda \geq 1$, then $\frac{1}{\lambda} \in [0,1]$, and applying the previous inequality to λX gives

$$\rho(X) = \rho\left(\frac{1}{\lambda}(\lambda X)\right) \leq \frac{1}{\lambda}\rho(\lambda X),$$

hence $\rho(\lambda X) \geq \lambda \rho(X)$. □

Proposition B.4. *Let ρ be a convex risk measure on $\mathcal{X}$. Then the following two conditions are equivalent:*

(a) ρ is continuous from below.

(b) ρ satisfies the Lebesgue property, i.e., for every bounded sequence (X_n) in $\mathcal{X}$ that converges pointwise to X, we have

$$\rho(X_n) \longrightarrow \rho(X).$$

Proof. (b)⇒(a): Suppose that ρ satisfies the Lebesgue property. Let (X_n) be a sequence such that $X_n \nearrow X$ pointwise. Since (X_n) is increasing and bounded from above by X, it is bounded in $\mathcal{X}$. Hence, by the Lebesgue property,

$$\rho(X_n) \longrightarrow \rho(X),$$

which shows continuity from below.

(a)⇒(b): Assume now that ρ is continuous from below, and let (X_n) be a bounded sequence in $\mathcal{X}$ converging pointwise to X. Define

$$Y_n := \inf_{k \geq n} X_k, \qquad Z_n := \sup_{k \geq n} X_k.$$

Then

$$Y_n \leq X_n \leq Z_n \qquad \text{for all } n,$$

and, since $X_n \to X$ pointwise,

$$Y_n \nearrow X \qquad \text{and} \qquad Z_n \searrow X$$

pointwise.

By monotonicity of ρ, we obtain

$$\rho(Y_n) \geq \rho(X_n) \geq \rho(Z_n) \qquad \text{for all } n.$$

Since ρ is continuous from below, it follows that

$$\rho(Y_n) \searrow \rho(X).$$

Moreover, by Remark 3.31, continuity from below implies continuity from above for convex risk measures. Hence,

$$\rho(Z_n) \nearrow \rho(X).$$

Therefore, $\rho(X_n)$ is squeezed between two sequences converging to $\rho(X)$, and thus

$$\rho(X_n) \longrightarrow \rho(X).$$

This proves the Lebesgue property. □

Proposition B.5. *Let ρ be a convex risk measure on $\mathcal{X}$. Then the following two conditions are equivalent:*

(a) ρ is continuous from below.

(b) For every $c > -\rho(0)$, the coherent risk measure

$$\rho_c(X) := \sup_{Q \in \Lambda_c} \mathbb{E}_Q[-X]$$

is continuous from below, where $\Lambda_c := \{Q \in \mathcal{M}_{1,f} \mid \alpha^{\min}(Q) \leq c\}$.

Proof. Throughout the proof we use that, by the robust representation theorem, ρ admits the representation

$$\rho(X) = \sup_{Q \in \mathcal{M}_{1,f}} \Big(\mathbb{E}_Q[-X] - \alpha^{\min}(Q) \Big), \qquad X \in \mathcal{X}.$$

(a) ⇒ (b): Assume that ρ is continuous from below. By Theorem 3.28, the minimal penalty function $\alpha^{\min}$ is concentrated on $\mathcal{M}_1$, i.e.,

$$\alpha^{\min}(Q) < \infty \implies Q \text{ is } \sigma\text{-additive.}$$

In particular, for every $c > -\rho(0)$ we have $\Lambda_c \subseteq \mathcal{M}_1$.

Fix $c > -\rho(0)$ and let (X_n) be a sequence in $\mathcal{X}$ with $X_n \nearrow X$ pointwise. By cash invariance, we may assume without loss of generality that $X_n \geq 0$ for all n, so (X_n) is bounded in $\mathcal{X}$.

We argue exactly as in the proof of the implication (b) $\Rightarrow$ (a) of Theorem 3.28. Since each $Q \in \Lambda_c \subseteq \mathcal{M}_1$ is σ-additive and (X_n) is bounded, monotone convergence yields

$$\mathbb{E}_Q[-X_n] \searrow \mathbb{E}_Q[-X] \quad \text{for every } Q \in \Lambda_c.$$

Thus the functions $\ell_n(Q) := \mathbb{E}_Q[-X_n]$ form a decreasing sequence of weak* continuous functions on Λ_c that converge pointwise to the weak* continuous function $\ell(Q) := \mathbb{E}_Q[-X]$.

Recall that Λ_c is weak* compact in ba: indeed, $\mathcal{M}_{1,f}$ is weak* compact by the Banach–Alaoglu theorem, and $\alpha^{\min}$ is weak* lower semicontinuous as the supremum of the weak* continuous maps $Q \mapsto \mathbb{E}_Q[-Y]$ for $Y \in \mathcal{A}_\rho$, so Λ_c is weak* closed and hence compact.

By Dini's lemma, the convergence $\ell_n \to \ell$ is uniform on Λ_c, that is,

$$\mathbb{E}_Q[-X_n] \longrightarrow \mathbb{E}_Q[-X] \quad \text{uniformly in } Q \in \Lambda_c.$$

Consequently,

$$\rho_c(X_n) = \sup_{Q \in \Lambda_c} \mathbb{E}_Q[-X_n] \longrightarrow \sup_{Q \in \Lambda_c} \mathbb{E}Q[-X] = \rho_c(X).$$

Combined with monotonicity of ρ_c, this gives $\rho_c(X_n) \searrow \rho_c(X)$, so ρ_c is continuous from below.

(b) $\Rightarrow$ (a): Conversely, assume that ρ_c is continuous from below for every $c > -\rho(0)$. Let (X_n) be a sequence in $\mathcal{X}$ with $X_n \nearrow X$ pointwise. By cash invariance we may again assume that $X_n \geq 0$ for all n.

Choose $c > 1 - \rho(X)$. As established in the proof of Theorem 3.28, only those Q with $\alpha^{\min}(Q) \leq c$ contribute to the supremum representing $\rho(X_n)$, that is,

$$\rho(X_n) = \sup_{Q \in \Lambda_c} \Big(\mathbb{E}_Q[-X_n] - \alpha^{\min}(Q) \Big),$$

and analogously for $\rho(X)$.

By assumption, $\rho_c(X_n) \searrow \rho_c(X)$, which by the same Dini-type argument as above (now read in the reverse direction) means that

$$\mathbb{E}_Q[-X_n] \longrightarrow \mathbb{E}_Q[-X] \quad \text{uniformly in } Q \in \Lambda_c.$$

We therefore obtain

$$\rho(X_n) = \sup_{Q\in\Lambda_c}\Big(\mathbb{E}_Q[-X_n] - \alpha^{\min}(Q)\Big) \longrightarrow \sup_{Q\in\Lambda_c}\Big(\mathbb{E}_Q[-X] - \alpha^{\min}(Q)\Big) = \rho(X).$$

Together with monotonicity of ρ this gives $\rho(X_n) \searrow \rho(X)$, so ρ is continuous from below. □

Proposition B.6. *Let $0 < \lambda < 1$. Then for any $X \in L^\infty$, the Average Value at Risk satisfies*

$$\mathrm{AV@R}_\lambda(X) = \frac{1}{\lambda}\int_{1-\lambda}^1 q_{-X}(t)\,dt = \frac{1}{\lambda}\int_0^\lambda \mathrm{V@R}_\gamma(X)\,d\gamma.$$

Proof. Recall that

$$\mathrm{AV@R}_\lambda(X) = \sup\left\{\mathbb{E}_Q[-X] \;\middle|\; Q \in \mathcal{M}_1(\mathbb{P}),\ \frac{dQ}{d\mathbb{P}} \le \frac{1}{\lambda}\ \mathbb{P}\text{-a.s.}\right\}.$$

By the representation of law-invariant convex risk measures (Theorem 3.52), we obtain

$$\mathrm{AV@R}_\lambda(X) = \sup_{Q\in\mathcal{Q}_\lambda}\int_0^1 q_{-X}(t)\,q_{\varphi_Q}(t)\,dt,$$

where $\mathcal{Q}_\lambda := \left\{Q \in \mathcal{M}_1(\mathbb{P}) \;\middle|\; \varphi_Q := \frac{dQ}{d\mathbb{P}} \le \frac{1}{\lambda}\ \mathbb{P}\text{-a.s.}\right\}$. For $Q \in \mathcal{Q}_\lambda$, the density φ_Q satisfies

$$0 \le q_{\varphi_Q}(t) \le \frac{1}{\lambda}, \qquad \int_0^1 q_{\varphi_Q}(t)\,dt = 1.$$

Let $A := [1-\lambda, 1]$. Since q_{-X} is increasing, we obtain

$$\int_0^1 q_{-X}(t)\,q_{\varphi_Q}(t)\,dt \le \frac{1}{\lambda}\int_{1-\lambda}^1 q_{-X}(t)\,dt.$$

Hence,

$$\mathrm{AV@R}_\lambda(X) \le \frac{1}{\lambda}\int_{1-\lambda}^1 q_{-X}(t)\,dt.$$

To prove equality, let U be uniformly distributed on $(0,1)$ such that

$$-X = q_{-X}(U) \quad \mathbb{P}\text{-a.s.}$$

Define

$$\varphi^* := \frac{1}{\lambda}\,\mathbf{1}_{\{U>1-\lambda\}}.$$

Then φ^* is the density of some probability measure $Q^* \in \mathcal{Q}_\lambda$, and

$$\mathbb{E}_{Q^*}[-X] = \mathbb{E}[-X\,\varphi^*] = \frac{1}{\lambda}\int_{1-\lambda}^1 q_{-X}(t)\,dt.$$

Thus,

$$\mathrm{AV@R}_\lambda(X) = \frac{1}{\lambda}\int_{1-\lambda}^{1} q_{-X}(t)\,dt.$$

Finally, using the identity $\mathrm{V@R}_\gamma(X) = q_{-X}(1-\gamma)$, and the substitution $t = 1-\gamma$, we obtain

$$\mathrm{AV@R}_\lambda(X) = \frac{1}{\lambda}\int_0^{\lambda} \mathrm{V@R}_\gamma(X)\,d\gamma.$$

□

Proposition B.7. *[20] Let $c : \mathcal{F} \to [0,1]$ be a normalized and monotone set function, and define its dual set function $\tilde{c}$ by*

$$\tilde{c}(A) := 1 - c(\Omega \setminus A), \quad A \in \mathcal{F}.$$

Then, for every bounded measurable function X, the corresponding Choquet integrals satisfy

$$\int (-X)\,dc \;=\; -\int X\,d\tilde{c}.$$

Proof. Recall the definition of the Choquet integral for a bounded measurable function X:

$$\int X\,dc = \int_0^\infty c(\{X > t\})\,dt + \int_{-\infty}^0 [c(\{X > t\}) - 1]\,dt.$$

Step 1: Compute the left-hand side. Substituting $-X$ for X in the Choquet integral definition:

$$\int (-X)\,dc = \int_0^\infty c(\{-X > t\})\,dt + \int_{-\infty}^0 [c(\{-X > t\}) - 1]\,dt.$$

Since $\{-X > t\} = \{X < -t\}$, we have:

$$\int (-X)\,dc = \int_0^\infty c(\{X < -t\})\,dt + \int_{-\infty}^0 [c(\{X < -t\}) - 1]\,dt.$$

Step 2: Change of variables. In the first integral, substitute $s = -t$ (so $dt = -ds$). When t goes from 0 to ∞, s goes from 0 to $-\infty$:

$$\int_0^\infty c(\{X < -t\})\,dt = \int_0^{-\infty} c(\{X < s\})(-ds) = \int_{-\infty}^0 c(\{X < s\})\,ds.$$

Similarly, for the second integral with $s = -t$:

$$\int_{-\infty}^0 [c(\{X < -t\}) - 1]\,dt = \int_0^\infty [c(\{X < s\}) - 1]\,ds.$$

Therefore:

$$\int (-X)\,dc = \int_{-\infty}^0 c(\{X < s\})\,ds + \int_0^\infty [c(\{X < s\}) - 1]\,ds.$$

Step 3: Use the dual relationship. Since $\{X \le s\} = \Omega \setminus \{X > s\}$, by definition of the dual:

$$c(\{X < s\}) = c(\Omega \setminus \{X > s\}) = 1 - \tilde{c}(\{X \ge s\}).$$

Substituting this into our expression:

$$\begin{aligned}\int (-X)\, dc &= \int_{-\infty}^{0} [1 - \tilde{c}(\{X \ge s\})]\, ds + \int_{0}^{\infty} [1 - \tilde{c}(\{X \ge s\}) - 1]\, ds \\ &= \int_{-\infty}^{0} [1 - \tilde{c}(\{X \ge s\})]\, ds - \int_{0}^{\infty} \tilde{c}(\{X \ge s\})\, ds.\end{aligned}$$

Step 4: Recognize the negative of the Choquet integral with respect to $\tilde{c}$. Rearranging the terms:

$$\begin{aligned}\int (-X)\, dc &= -\int_{0}^{\infty} \tilde{c}(\{X \ge s\})\, ds + \int_{-\infty}^{0} ds - \int_{-\infty}^{0} \tilde{c}(\{X \ge s\})\, ds \\ &= -\int_{0}^{\infty} \tilde{c}(\{X > s\})\, ds - \int_{-\infty}^{0} [\tilde{c}(\{X > s\}) - 1]\, ds \\ &= -\left[\int_{0}^{\infty} \tilde{c}(\{X > s\})\, ds + \int_{-\infty}^{0} [\tilde{c}(\{X > s\}) - 1]\, ds\right] \\ &= -\int X\, d\tilde{c}.\end{aligned}$$

This completes the proof. □

Proposition B.8. *Let ψ be a concave distortion function on $[0,1]$, and define its convex conjugate $\varphi : [0,\infty) \to \mathbb{R}$ by*

$$\varphi(x) := \sup_{y \in [0,1]} \Big(\psi(y) - xy\Big), \quad x \ge 0.$$

Here φ is once more the Fenchel–Legendre transform of Subsection 2.5.1, applied to the concave distortion ψ (equivalently, the conjugate of the convex function $-\psi$ in the sense of Definition 2.28). The representation of $\mathcal{Q}_\mu$ below thus expresses the dual set of Theorem 4.13 through the same conjugation that underlies the robust representation of Chapter 3.

Then the set $\mathcal{Q}_\mu$ in Theorem 4.13 admits the representation

$$\mathcal{Q}_\mu = \left\{ Q \in \mathcal{M}_1(\mathbb{P}) \;\middle|\; Z := \frac{dQ}{d\mathbb{P}} \text{ satisfies } \mathbb{E}\Big[(Z - x)^+\Big] \le \varphi(x) \text{ for all } x \ge 0 \right\}.$$

Proof. We show that the two characterizations of $\mathcal{Q}_\mu$ are equivalent, i.e.,

$$\int_t^1 q_Z(s)\, \mathrm{d}s \;\le\; \psi(1-t) \quad \forall\, t \in (0,1) \qquad \Longleftrightarrow \qquad \mathbb{E}\Big[(Z - x)^+\Big] \;\le\; \varphi(x) \quad \forall\, x \ge 0,$$

where $\varphi(x) := \sup_{y \in [0,1]} \Big(\psi(y) - xy\Big)$.

Step 1: Quantile representation of $\mathbb{E}[(Z-x)^+]$. For $Z \ge 0$ with non-decreasing

quantile function q_Z, we have

$$\mathbb{E}\big[(Z-x)^+\big] \;=\; \int_0^1 (q_Z(s)-x)^+ \,\mathrm{d}s.$$

Since q_Z is non-decreasing, there exists $t_x \in [0,1]$ such that $q_Z(s) > x \Leftrightarrow s > t_x$, where $t_x = F_Z(x) = \mathbb{P}(Z \le x)$. Therefore,

$$\mathbb{E}\big[(Z-x)^+\big] \;=\; \int_{t_x}^1 q_Z(s)\,\mathrm{d}s \;-\; x(1-t_x).$$

Step 2: Evaluation of $\varphi(x)$. For fixed $x \ge 0$, the function $g(y) = \psi(y) - xy$ is concave in y (as the sum of a concave and a linear function). Since ψ is concave and continuous on the compact set $[0,1]$, the supremum is attained at $y^* \in [0,1]$. At an interior maximum, $\psi'(y^*) = x$, which corresponds to $y^* = 1 - t_x$. Hence,

$$\varphi(x) \;=\; \psi(1-t_x) \;-\; x(1-t_x).$$

Step 3: Equivalence. The condition $\mathbb{E}[(Z-x)^+] \le \varphi(x)$ becomes, after substituting Steps 1 and 2 and cancelling the term $-x(1-t_x)$ from both sides,

$$\int_{t_x}^1 q_Z(s)\,\mathrm{d}s \;\le\; \psi(1-t_x).$$

As x ranges over $[0,\infty)$, the value $t_x = F_Z(x)$ ranges over $(0,1)$. This is precisely the condition in Theorem 4.13, namely

$$\int_t^1 q_Z(s)\,\mathrm{d}s \;\le\; \psi(1-t) \quad \text{for all } t \in (0,1).$$

Step 4: Law-invariance. The condition depends only on the quantile function q_Z, and hence only on the distribution of $Z = \mathrm{d}Q/\mathrm{d}\mathbb{P}$ under $\mathbb{P}$, i.e., on the cumulative distribution function F_X. Therefore, $\mathcal{Q}_\mu$ is law-invariant and can equivalently be represented through F_X. □

B.2 Technical Proofs

Corollary B.9 (Restatement of Corollary 3.34)**.** *For a convex risk measure ρ on L^∞, the following conditions are equivalent:*

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

(b) ρ satisfies the Lebesgue property: $\rho(X_n) \to \rho(X)$ whenever (X_n) is a bounded sequence in L^∞ that converges $\mathbb{P}$-a.s. to X.

(c) The minimal penalty function $\alpha^{\min}$ is concentrated on $\mathcal{M}_1(\mathbb{P})$, that is,

$$\alpha^{\min}(Q) < \infty \;\Rightarrow\; Q \in \mathcal{M}_1(\mathbb{P}).$$

(d) For each $c \in \mathbb{R}$, the level set $\Lambda_c := \{Q \mid \alpha^{\min}(Q) \le c\}$ is contained in $\mathcal{M}_1(\mathbb{P})$, and the corresponding set of densities,

$$\left\{ \frac{dQ}{d\mathbb{P}} \;\middle|\; Q \in \Lambda_c \right\}$$

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

Particularly, we have

$$\rho(X) = \max_{Q \in \mathcal{M}_1(\mathbb{P})} \Big(\mathbb{E}_Q[-X] - \alpha^{\min}(Q) \Big), \quad X \in L^\infty, \tag{B.1}$$

whenever one of these equivalent conditions is satisfied.

Proof of Corollary 3.34. As already shown in the main text, conditions (a), (b), (c) are equivalent. It thus suffices to show (a) $\iff$ (d). Throughout, fix $c > -\rho(0) = \inf_{Q \in \mathcal{M}_{1,f}} \alpha^{\min}(Q)$ and write

$$\Lambda_c \;=\; \Big\{ Q \in \mathcal{M}_{1,f} \;\Big|\; \alpha^{\min}(Q) \le c \Big\}, \qquad \mathcal{D}_c \;:=\; \left\{ \tfrac{dQ}{d\mathbb{P}} \;\middle|\; Q \in \Lambda_c \right\}.$$

The set $\mathcal{D}_c$ is well defined as a subset of $L^1(\Omega, \mathcal{F}, \mathbb{P})$ only once we know $\Lambda_c \subseteq \mathcal{M}_1(\mathbb{P})$; otherwise it should be read as the empty collection of densities of the absolutely continuous elements of Λ_c. The argument below makes this distinction explicit.

Step 1: (d) $\Longrightarrow$ (a). Assume (d). In particular, $\Lambda_c \subseteq \mathcal{M}_1(\mathbb{P})$ for every $c > -\rho(0)$, so $\alpha^{\min}(Q) < \infty$ implies $Q \in \mathcal{M}_1(\mathbb{P})$, which is condition (c). By Corollary 3.34 from the main text, this gives (a).

(Alternatively, one can argue directly from Lemma 3.29: take a sequence $X_n \nearrow X$ in L^∞ and show $\inf_{Q \in \Lambda_c} \mathbb{E}_Q[X_n - X] \to 0$. Since $\mathcal{D}_c$ is uniformly integrable by Dunford–Pettis, the expectations $\mathbb{E}_Q[X_n - X] = \mathbb{E}_{\mathbb{P}}\big[\frac{dQ}{d\mathbb{P}}(X_n - X)\big]$ converge to 0 uniformly in $Q \in \Lambda_c$, and Lemma 3.29 yields $\rho(\lambda X_n) \to \rho(\lambda X)$ for all $\lambda \ge 1$, hence (a) by cash invariance.)

Step 2: (a) $\Longrightarrow$ (d). Assume (a). By Corollary 3.34, condition (c) holds, i.e., $\Lambda_c \subseteq \mathcal{M}_1(\mathbb{P})$ for every $c > -\rho(0)$. In particular, $\mathcal{D}_c \subseteq L^1(\Omega, \mathcal{F}, \mathbb{P})$ is well defined. It remains to prove that $\mathcal{D}_c$ is weakly compact in L^1. By the Dunford–Pettis theorem, this is equivalent to showing that $\mathcal{D}_c$ is uniformly integrable.

Step 2a: Reduction to a uniform-tail estimate. A family $\mathcal{D} \subseteq L^1$ is uniformly integrable if and only if

$$\lim_{K \to \infty} \sup_{\varphi \in \mathcal{D}} \mathbb{E}_{\mathbb{P}}\Big[\varphi \, \mathbf{1}_{\{\varphi > K\}} \Big] \;=\; 0. \tag{B.2}$$

Equivalently, since each $\varphi = dQ/d\mathbb{P}$ satisfies $\mathbb{E}_{\mathbb{P}}[\varphi] = 1$, condition (B.2) is the same as

$$\lim_{K\to\infty} \inf_{Q\in\Lambda_c} \mathbb{E}_Q\Big[\mathbf{1}_{\{\varphi_Q \le K\}}\Big] \;=\; 1, \tag{B.3}$$

where $\varphi_Q := dQ/d\mathbb{P}$. Indeed, $\mathbb{E}_Q[\mathbf{1}_{\{\varphi_Q\le K\}}] = \mathbb{E}_P[\varphi_Q \mathbf{1}_{\{\varphi_Q\le K\}}] = 1 - \mathbb{E}_{\mathbb{P}}[\varphi_Q \mathbf{1}_{\{\varphi_Q > K\}}]$.

Step 2b: Constructing a suitable sequence in L^∞. For each $n \in \mathbb{N}$, set

$$A_n \;:=\; \bigcap_{Q\in\Lambda_c} \{\varphi_Q \le n\}^c \;=\; \emptyset \qquad \text{(in general not constructive)},$$

so we proceed differently and exploit the fact that we only need (B.3) to hold along *some* sequence $K_n \uparrow \infty$. Suppose, for contradiction, that $\mathcal{D}_c$ is *not* uniformly integrable. Then by (B.3) there exist $\varepsilon > 0$, a sequence $K_n \uparrow \infty$, and measures $Q_n \in \Lambda_c$ such that

$$\mathbb{E}_{Q_n}\Big[\mathbf{1}_{\{\varphi_{Q_n}\le K_n\}}\Big] \;\le\; 1-\varepsilon \qquad \text{for all } n. \tag{B.4}$$

Define $X_n := \mathbf{1}_{B_n} \in L^\infty$ where $B_n := \bigcup_{k\ge n}\{\varphi_{Q_k} > K_k\}$. Then X_n is decreasing in n. Moreover, since each $\varphi_{Q_k} \in L^1(\mathbb{P})$, Markov's inequality gives $\mathbb{P}[\varphi_{Q_k} > K_k] \le 1/K_k \to 0$, so by the Borel–Cantelli lemma applied to a fast-growing subsequence K_{k_j} with $\sum_j 1/K_{k_j} < \infty$ (which we may pass to without loss of generality), $\mathbb{P}[\limsup_n\{\varphi_{Q_n} > K_n\}] = 0$. Hence $X_n \searrow 0$ $\mathbb{P}$-a.s., and equivalently $1 - X_n \nearrow 1$ pointwise $\mathbb{P}$-a.s.

Step 2c: Applying Lemma 3.29 to derive a contradiction. Set $Y_n := 1 - X_n \in [0,1]$, so $Y_n \nearrow 1$ $\mathbb{P}$-a.s. By condition (a), ρ is continuous from below, hence (using cash invariance) $\rho(\lambda Y_n) \to \rho(\lambda)$ for every $\lambda \ge 1$. By the implication (a) $\Longrightarrow$ (b) of Lemma 3.29,

$$\inf_{Q\in\Lambda_c} \mathbb{E}_Q[Y_n] \;\longrightarrow\; 1 \qquad \text{as } n\to\infty. \tag{B.5}$$

On the other hand, for the specific sequence $(Q_n) \subseteq \Lambda_c$ constructed above,

$$\mathbb{E}_{Q_n}[Y_n] \;=\; 1 - \mathbb{E}_{Q_n}[X_n] \;\le\; 1 - \mathbb{E}_{Q_n}\Big[\mathbf{1}_{\{\varphi_{Q_n}>K_n\}}\Big] \;=\; \mathbb{E}_{Q_n}\Big[\mathbf{1}_{\{\varphi_{Q_n}\le K_n\}}\Big] \;\le\; 1-\varepsilon$$

by (B.4), where the first inequality uses $\{\varphi_{Q_n} > K_n\} \subseteq B_n$ and hence $\mathbf{1}_{\{\varphi_{Q_n}>K_n\}} \le X_n$. Therefore

$$\inf_{Q\in\Lambda_c} \mathbb{E}_Q[Y_n] \;\le\; \mathbb{E}_{Q_n}[Y_n] \;\le\; 1-\varepsilon \qquad \text{for all } n,$$

contradicting (B.5). Hence $\mathcal{D}_c$ is uniformly integrable, and by the Dunford–Pettis theorem $\mathcal{D}_c$ is weakly relatively compact in L^1.

Step 2d: Weak closedness. It remains to upgrade weak relative compactness to weak compactness, i.e., to show that $\mathcal{D}_c$ is weakly closed. The minimal penalty function $\alpha^{\min}$ is convex and lower semicontinuous on $\mathcal{M}_{1,f}$ with respect to the $\sigma(\mathcal{M}_{1,f}, L^\infty)$-topology. When $\Lambda_c \subseteq \mathcal{M}_1(\mathbb{P})$, identifying Q with $\varphi_Q = dQ/d\mathbb{P}$ and using that the duality (L^1, L^∞) induces precisely the weak topology on L^1, the level set $\mathcal{D}_c = \{\varphi_Q : \alpha^{\min}(Q) \le c\}$ is the level set of a convex, $\sigma(L^1, L^\infty)$-lower semicontinuous functional and hence weakly

closed in L^1. Combined with Step 2c, $\mathcal{D}_c$ is weakly compact in L^1, proving (d).

This completes the proof of (a) $\Longleftrightarrow$ (d) and thereby the equivalence of (a)–(d). □

Lemma B.10 (Restatement of Lemma 3.46)**.** *For $\lambda \in (0,1)$, the risk measure* $\mathrm{AV@R}_\lambda$ *is subadditive, i.e.,*

$$\mathrm{AV@R}_\lambda(X+Y) \le \mathrm{AV@R}_\lambda(X) + \mathrm{AV@R}_\lambda(Y) \quad \textit{for all } X, Y \in \mathcal{X}.$$

Proof of Lemma 3.46. By Proposition 3.45, we have

$$\mathrm{AV@R}_\lambda(Z) = -\frac{1}{\lambda} \inf_{r\in\mathbb{R}} \Big(\mathbb{E}[(r-Z)^+] - \lambda r \Big) \qquad \text{for every } Z \in \mathcal{X}.$$

Applying this to $Z = X + Y$, we obtain

$$\mathrm{AV@R}_\lambda(X+Y) = -\frac{1}{\lambda} \inf_{r\in\mathbb{R}} \Big(\mathbb{E}[(r-X-Y)^+] - \lambda r \Big).$$

Now let $r_1, r_2 \in \mathbb{R}$ and set $r = r_1 + r_2$. Then $r - (X+Y) = (r_1 - X) + (r_2 - Y)$. Using the elementary inequality $(a+b)^+ \le a^+ + b^+$ for all $a, b \in \mathbb{R}$, it follows that $(r - X - Y)^+ = \Big((r_1 - X) + (r_2 - Y) \Big)^+ \le (r_1 - X)^+ + (r_2 - Y)^+$. Taking expectations yields

$$\mathbb{E}[(r-X-Y)^+] \le \mathbb{E}[(r_1 - X)^+] + \mathbb{E}[(r_2 - Y)^+].$$

Hence,

$$\mathbb{E}[(r-X-Y)^+] - \lambda r \le \Big(\mathbb{E}[(r_1 - X)^+] - \lambda r_1 \Big) + \Big(\mathbb{E}[(r_2 - Y)^+] - \lambda r_2 \Big).$$

Since this holds for all $r_1, r_2 \in \mathbb{R}$, we get

$$\inf_{r\in\mathbb{R}} \Big(\mathbb{E}[(r-X-Y)^+] - \lambda r \Big) \le \inf_{r_1\in\mathbb{R}} \Big(\mathbb{E}[(r_1 - X)^+] - \lambda r_1 \Big) + \inf_{r_2\in\mathbb{R}} \Big(\mathbb{E}[(r_2 - Y)^+] - \lambda r_2 \Big).$$

Multiplying by $-\frac{1}{\lambda}$ reverses the inequality, and therefore

$$\mathrm{AV@R}_\lambda(X+Y) \le -\frac{1}{\lambda} \inf_{r_1\in\mathbb{R}} \Big(\mathbb{E}[(r_1 - X)^+] - \lambda r_1 \Big) - \frac{1}{\lambda} \inf_{r_2\in\mathbb{R}} \Big(\mathbb{E}[(r_2 - Y)^+] - \lambda r_2 \Big).$$

Using Proposition 3.45 once more, this is exactly

$$\mathrm{AV@R}_\lambda(X+Y) \le \mathrm{AV@R}_\lambda(X) + \mathrm{AV@R}_\lambda(Y).$$

Thus, $\mathrm{AV@R}_\lambda$ is subadditive. □

Theorem B.11 (Restatement of Theorem 3.47)**.** *For $\lambda \in (0,1]$, $\mathrm{AV@R}_\lambda$ is a coherent risk measure that is continuous from below. Moreover, it admits the representation*

$$\mathrm{AV@R}_\lambda(X) = \max_{Q\in\mathcal{Q}_\lambda} \mathbb{E}_Q[-X], \quad X \in \mathcal{X},$$

where $\mathcal{Q}_\lambda$ is the set of all probability measures $Q \ll \mathbb{P}$ whose density $dQ/d\mathbb{P}$ is $\mathbb{P}$-a.s. bounded by $1/\lambda$. Furthermore, $\mathcal{Q}_\lambda$ coincides with the maximal set $\mathcal{Q}_{\max}$ of Corollary 3.36.

Proof of Theorem 3.47. Since $\mathcal{Q}_1 = \{\mathbb{P}\}$, the assertion is obvious for $\lambda = 1$. For $0 < \lambda < 1$, consider the coherent risk measure $\rho_\lambda(X) := \sup_{Q\in\mathcal{Q}_\lambda} \mathbb{E}_Q[-X]$. First, we assume that we are given some $X < 0$. We define a measure $\tilde{\mathbb{P}} \approx \mathbb{P}$ by $d\tilde{\mathbb{P}}/d\mathbb{P} = X/\mathbb{E}[X]$. Then

$$\rho_\lambda(X) = \frac{\mathbb{E}[-X]}{\lambda} \sup\Big\{\tilde{\mathbb{E}}[\varphi] \mid 0 \le \varphi \le 1,\ \mathbb{E}[\varphi] = \lambda\Big\}.$$

Clearly, the condition $\mathbb{E}[\varphi] = \lambda$ on the right can be replaced by $\mathbb{E}[\varphi] \le \lambda$. Thus, we can apply the Neyman–Pearson lemma and conclude that the supremum is attained by

$$\varphi_0 = \mathbf{1}_{\{X<q\}} + \kappa \mathbf{1}_{\{X=q\}},$$

for a λ-quantile q of X and a constant $\kappa \in [0,1]$ that is determined by the condition $\mathbb{E}[\varphi_0] = \lambda$. Hence,

$$\rho_\lambda(X) = \frac{\mathbb{E}[-X]}{\lambda} \cdot \tilde{\mathbb{E}}[\varphi_0] = \frac{1}{\lambda}\mathbb{E}[-X\varphi_0].$$

Since $dQ_0 = \lambda^{-1}\varphi_0\, d\mathbb{P}$ defines a probability measure in $\mathcal{Q}_\lambda$, we conclude that

$$\begin{aligned}
\rho_\lambda(X) &= \max_{Q\in\mathcal{Q}_\lambda} \mathbb{E}_Q[-X] = \mathbb{E}_{Q_0}[-X] \\
&= \frac{1}{\lambda}\Big(\mathbb{E}[-X; X < q] - q\lambda + q\mathbb{P}[X < q]\Big) \\
&= \frac{1}{\lambda}\mathbb{E}[(q - X)^+] - q \\
&= \mathrm{AV@R}_\lambda(X).
\end{aligned}$$

This proves the representation for $X < 0$.

For arbitrary $X \in L^\infty$, we use the cash invariance of both ρ_λ and $\mathrm{AV@R}_\lambda$. Notably, the representation in the theorem shows that $\mathrm{AV@R}_\lambda$ is a coherent risk measure.

It remains to prove that $\mathcal{Q}_\lambda$ is the maximal set of Corollary 3.36. To this end, we show that

$$\sup_{X\in\mathcal{X}}\Big(\mathbb{E}_Q[-X] - \mathrm{AV@R}_\lambda(X)\Big) = +\infty$$

for $Q \notin \mathcal{Q}_\lambda$. We denote by φ the density $dQ/d\mathbb{P}$. There exist $\lambda' \in (0,\lambda)$ and $k > 1/\lambda'$ such that $\mathbb{P}[\varphi \wedge k \ge 1/\lambda'] > 0$. For $c > 0$, define $X^{(c)} \in \mathcal{X}$ by

$$X^{(c)} := -c(\varphi \wedge k)\, \mathbf{1}_{\{\varphi \ge 1/\lambda'\}}.$$

Since, by Markov's inequality,

$$\mathbb{P}[X^{(c)} < 0] = \mathbb{P}\Big[\varphi \ge \frac{1}{\lambda'}\Big] \le \frac{\mathbb{E}[\varphi]}{1/\lambda'} = \lambda' < \lambda,$$

we have $\mathrm{V@R}_\lambda(X^{(c)}) = 0$, and we get that

$$\mathrm{AV@R}_\lambda(X^{(c)}) = \frac{1}{\lambda}\mathbb{E}[-X^{(c)}] = \frac{c}{\lambda}\mathbb{E}\Big[\varphi \wedge k;\ \varphi \geq \frac{1}{\lambda'}\Big] > 0.$$

On the other hand,

$$\mathbb{E}_Q[-X^{(c)}] = c\,\mathbb{E}\Big[\varphi \cdot (\varphi \wedge k);\ \varphi \geq \frac{1}{\lambda'}\Big] \geq \frac{c}{\lambda'}\mathbb{E}\Big[\varphi \wedge k;\ \varphi \geq \frac{1}{\lambda'}\Big].$$

Thus, the difference between $\mathbb{E}_Q[-X^{(c)}]$ and $\mathrm{AV@R}_\lambda(X^{(c)})$ becomes arbitrarily large as $c \uparrow \infty$.

Finally, to show continuity from below, let q be a λ-quantile of $X \in \mathcal{X}$. For any sequence (X_n) in $\mathcal{X}$ that increases to X, dominated convergence implies that

$$\mathrm{AV@R}_\lambda(X) = \frac{1}{\lambda}\mathbb{E}[(q - X)^+] - q = \lim_{n\to\infty}\left(\frac{1}{\lambda}\mathbb{E}[(q - X_n)^+] - q\right) \geq \limsup_{n\to\infty} \mathrm{AV@R}_\lambda(X_n),$$

using Proposition 3.45. Since $\mathrm{AV@R}_\lambda(X) \leq \mathrm{AV@R}_\lambda(X_n)$ by monotonicity, we have shown continuity from below. □

Lemma B.12 (Restatement of Lemma 5.5)**.** *For two measurable functions X and Y on $(\Omega, \mathcal{F})$, the following conditions are equivalent:*

(a) X and Y are comonotone.

(b) There exists a measurable function Z on $(\Omega, \mathcal{F})$ and increasing functions f and g on $\mathbb{R}$ such that

$$X = f(Z) \quad \text{and} \quad Y = g(Z).$$

(c) There exist increasing and Lipschitz continuous functions f and g on $\mathbb{R}$ such that

$$X = f(X + Y) \quad \text{and} \quad Y = g(X + Y).$$

Proof of Lemma 5.5. The implication $(c) \Rightarrow (b)$ is immediate by choosing $Z := X + Y$. The implication $(b) \Rightarrow (a)$ follows from the fact that increasing transformations of the same random variable are comonotone.

To prove $(a) \Rightarrow (c)$, suppose that X and Y are comonotone and define

$$Z := X + Y.$$

We show that $z := Z(\omega)$ admits a unique decomposition of the form $z = x + y$, where $(x, y) = (X(\omega'), Y(\omega'))$ for some $\omega' \in \Omega$. The existence of such a decomposition is immediate by taking $x := X(\omega)$ and $y := Y(\omega)$. It remains to show uniqueness. To this end, suppose that $X(\omega) + Y(\omega) = z = X(\omega') + Y(\omega')$ for some $\omega' \in \Omega$. Then

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

and comonotonicity implies that this expression vanishes. Hence, $X(\omega) = X(\omega')$ and $Y(\omega) = Y(\omega')$, which shows uniqueness. We can therefore define functions f and g on $Z(\Omega)$ by

$$f(z) := x, \qquad g(z) := y.$$

Next, we check that both f and g are increasing and Lipschitz continuous on $Z(\Omega)$. Let

$$z_1 = X(\omega_1) + Y(\omega_1) \le z_2 = X(\omega_2) + Y(\omega_2). \tag{B.6}$$

Then

$$X(\omega_2) - X(\omega_1) \ge -\big(Y(\omega_2) - Y(\omega_1)\big), \tag{B.7}$$

and comonotonicity yields $X(\omega_2) - X(\omega_1) \ge 0$ and $Y(\omega_2) - Y(\omega_1) \ge 0$, whence $f(z_1) \le f(z_2)$ and $g(z_1) \le g(z_2)$, so f and g are increasing. Moreover, (B.6) and (B.7) imply that

$$0 \le f(z_2) - f(z_1) = X(\omega_2) - X(\omega_1) \le X(\omega_2) - X(\omega_1) + Y(\omega_2) - Y(\omega_1) = z_2 - z_1,$$

which shows $|f(z_2) - f(z_1)| \le |z_2 - z_1|$.

Thus, f is Lipschitz continuous, and the same argument applies to g.

It remains to extend f and g to increasing Lipschitz continuous functions on all of $\mathbb{R}$. Since Lipschitz continuity implies uniform continuity, both functions can be extended to continuous functions on the closure $A := \overline{Z(\Omega)}$. These extensions are still increasing and Lipschitz continuous. Next, let us fix $x \in \mathbb{R} \setminus A$. If $x < z$ for all $z \in A$, then $z_1 := \min A$ belongs to the closed set A, and we can define $f(x) := f(z_1)$. Similarly, if $x > z$ for all $z \in A$, we let $f(x) := f(\max A)$. If neither of the preceding two cases holds, then we let $z_1 := \max\{z \in A \mid z < x\}$ and $z_2 := \min\{z \in A \mid z > x\}$, and note that both z_1 and z_2 belong to the closed set A. We may thus define

$$f(x) := \frac{x - z_1}{z_2 - z_1} f(z_2) + \frac{z_2 - x}{z_2 - z_1} f(z_1).$$

It is easy to see that this construction yields the desired extension of f to an increasing and Lipschitz continuous function on $\mathbb{R}$. □

Theorem B.13 (Restatement of Theorem 5.17)**.** *For the Choquet integral with respect to a normalized monotone set function c, the following conditions are equivalent:*

(a) $\rho(X) := \int (-X)\, dc$ *is a convex risk measure on* $\mathcal{X}$.

(b) $\rho(X) := \int (-X)\, dc$ *is a coherent risk measure on* $\mathcal{X}$.

(c) For $\mathcal{Q}_c := \{Q \in \mathcal{M}_{1,f} \mid Q[A] \le c(A) \text{ for all } A \in \mathcal{F}\}$,

$$\int X\, dc = \max_{Q \in \mathcal{Q}_c} \mathbb{E}_Q[X] \qquad \text{for } X \in \mathcal{X}.$$

(d) The set function c is submodular: $c(A \cap B) + c(A \cup B) \leq c(A) + c(B)$.

In this case, $\mathcal{Q}_c$ is equal to the maximal representing set $\mathcal{Q}_{\max}$ for ρ.

In preparation for the proof of the theorem, we first establish the following two auxiliary lemmas.

Lemma B.14. *For $X \in \mathcal{X}$ and $Q \in \mathcal{M}_{1,f}$, the integral $\mathbb{E}_Q[X]$ is equal to the Choquet integral $\int X\, dQ$.*

Proof. It is enough to prove the result for $X \geq 0$. Suppose first that $X = \sum_{i=1}^n x_i \mathbf{1}_{A_i}$ as in Remark 4.6. Then

$$\int X\, dQ = \sum_{i=1}^n (x_i - x_{i+1}) Q\left[\bigcup_{k=1}^i A_k\right] = \sum_{i=1}^n x_i Q[A_i] = \mathbb{E}_Q[X].$$

The result for general $X \in \mathcal{X}$ follows by approximating X uniformly with X_n that take only finitely many values, and by using the Lipschitz continuity of both $\mathbb{E}_Q[\cdot]$ and $\int \cdot\, dQ$ with respect to the supremum norm. □

Lemma B.15. *Let $A_1, \ldots, A_n$ be a partition of Ω into disjoint measurable sets, and suppose that the normalized monotone set function c is submodular. Let Q be the probability measure on $\mathcal{F}_0 := \sigma(A_1, \ldots, A_n)$ with weights*

$$Q[A_k] := c(B_k) - c(B_{k-1}) \quad \text{for } B_0 := \emptyset \text{ and } B_k := \bigcup_{j=1}^k A_j,\ k \geq 1. \tag{B.8}$$

Then $\int X\, dc \geq \mathbb{E}_Q[X]$ for all $\mathcal{F}_0$-measurable $X = \sum_{i=1}^n x_i \mathbf{1}_{A_i}$, and equality holds if the values of X are arranged in decreasing order: $x_1 \geq \cdots \geq x_n$.

Proof. Clearly, it suffices to consider only the case $X \geq 0$. Then Remark 4.6 implies $\int X\, dc = \mathbb{E}_Q[X]$ as soon as the values of X are arranged in decreasing order.

Now we prove $\int X\, dc \geq \mathbb{E}_Q[X]$ for arbitrary $\mathcal{F}_0$-measurable $X \geq 0$. To this end, note that any permutation σ of $\{1, \ldots, n\}$ induces a probability measure Q_σ on $\mathcal{F}_0$ by applying the definition of Q to the relabeled partition $A_{\sigma(1)}, \ldots, A_{\sigma(n)}$. If σ is a permutation such that $x_{\sigma(1)} \geq \cdots \geq x_{\sigma(n)}$, then we have $\int X\, dc = \mathbb{E}_{Q_\sigma}[X]$, and so the assertion will follow if we can prove that $\mathbb{E}_{Q_\sigma}[X] \geq \mathbb{E}_Q[X]$. To this end, it is enough to show that $\mathbb{E}_{Q_\tau}[X] \geq \mathbb{E}_Q[X]$ if τ is the transposition of two indices i and $i+1$ such that $x_i < x_{i+1}$, because σ can be represented as a finite composition of such transpositions.

Note next that

$$\mathbb{E}_{Q_\tau}[X] - \mathbb{E}_Q[X] = x_i\Big(Q_\tau[A_i] - Q[A_i]\Big) + x_{i+1}\Big(Q_\tau[A_{i+1}] - Q[A_{i+1}]\Big). \tag{B.9}$$

To compute the probabilities $Q_\tau[A_k]$, let us introduce

$$B_0^\tau := \emptyset, \quad B_k^\tau := \bigcup_{j=1}^k A_{\tau(j)}, \quad k = 1, \ldots, n.$$

Then $B_k^\tau = B_k$ for $k \neq i$. Hence,

$$\begin{aligned} Q_\tau[A_i] + Q_\tau[A_{i+1}] &= Q_\tau[A_{\tau(i)}] + Q_\tau[A_{\tau(i+1)}] = c(B_{i+1}^\tau) - c(B_{i-1}^\tau) \\ &= c(B_{i+1}) - c(B_{i-1}) = Q[A_i] + Q[A_{i+1}]. \end{aligned} \tag{B.10}$$

Moreover, $B_i^\tau \cap B_i = B_{i-1}$, $\quad B_i^\tau \cup B_i = B_{i+1}$, and hence $c(B_{i-1}) + c(B_{i+1}) \leq c(B_i^\tau) + c(B_i)$, due to the submodularity of c. Thus,

$$Q[A_{i+1}] = c(B_{i+1}) - c(B_i) \leq c(B_i^\tau) - c(B_{i-1}^\tau) = Q_\tau[A_{\tau(i)}] = Q_\tau[A_{i+1}].$$

Using (B.9), (B.10), and our assumption $x_i < x_{i+1}$ thus yields $\mathbb{E}_{Q_\tau}[X] \geq \mathbb{E}_Q[X]$. □

Proof of Theorem 5.17. $(a) \Leftrightarrow (b)$*:* According to Proposition 5.6, the property of positive homogeneity is shared by all Choquet integrals, and so $(a) \Rightarrow (b)$ follows. The converse implication is obvious.

$(b) \Rightarrow (c)$*:* By Corollary 3.20, $\rho(-X) = \max_{Q \in \mathcal{Q}_{\max}} \mathbb{E}_Q[X]$, where $Q \in \mathcal{M}_{1,f}$ belongs to $\mathcal{Q}_{\max}$ if and only if

$$\mathbb{E}_Q[X] \leq \rho(-X) = \int X \, dc \qquad \text{for all } X \in \mathcal{X}. \tag{B.11}$$

We will now show that this set $\mathcal{Q}_{\max}$ coincides with the set $\mathcal{Q}_c$. If $Q \in \mathcal{Q}_{\max}$, then in particular, $Q[A] \leq \int \mathbf{1}_A \, dc = c(A)$ for all $A \in \mathcal{F}$. Hence, $Q \in \mathcal{Q}_c$. Conversely, suppose $Q \in \mathcal{Q}_c$. If $X \geq 0$, then

$$\int X \, dc = \int_0^\infty c(X > x) \, dx \geq \int_0^\infty Q[X > x] \, dx = \mathbb{E}_Q[X],$$

where we have used Lemma B.14. Cash invariance yields (B.11)

$(c) \Rightarrow (b)$*:* This is obvious.

$(b) \Rightarrow (d)$*:* This follows precisely as in (5.1).

$(d) \Rightarrow (c)$*:* To prove this implication, it is enough to show that the Choquet integral is subadditive. By Lemma 3.6, it is enough to prove this for random variables that take only finitely many values. Thus, let $A_1, \ldots, A_n$ be a partition of Ω into finitely many disjoint measurable sets. Let us write $X = \sum_i x_i \mathbf{1}_{A_i}, Y = \sum_i y_i \mathbf{1}_{A_i}$, and let the indices $i = 1, \ldots, n$ be arranged in such a way that $x_1 + y_1 \geq \cdots \geq x_n + y_n$. Then the probability measure Q constructed in Lemma B.15 is such that

$$\int (X + Y) \, dc = \mathbb{E}_Q[X + Y] = \mathbb{E}_Q[X] + \mathbb{E}_Q[Y] \leq \int X \, dc + \int Y \, dc.$$

But this is the required subadditivity of the Choquet integral. □

Appendix C

Technical Details of the Application

C.1 Auxiliary Results from Optimization

The analytical solution of Problem 6.2 in Section 6.3 rests on two tools from optimization theory: a minimax theorem, which permits the interchange of the infimum and supremum, and Slater's condition, which guarantees strong duality for the resulting convex maximization problem. For completeness, we collect both here and indicate how their hypotheses are met in our setting.

C.1.1 The Minimax Theorem

The reversal of the order of optimization in Problem 6.2, which is the starting point of the solution in Section 6.3, rests on a minimax theorem. We state the version used by Boonen and Jiang [11], due to Fan [22], and explain how its hypotheses are met.

The theorem is formulated on a Hausdorff topological vector space, that is, a topological vector space in which any two distinct points can be separated by disjoint open sets.

Theorem C.1 (Minimax theorem)**.** *[22] Let Ξ_1 be a non-empty compact convex Hausdorff topological vector space and let Ξ_2 be a convex set. Let $\mathcal{H}$ be a real-valued function on $\Xi_1 \times \Xi_2$ such that*

- *$\xi_1 \mapsto \mathcal{H}(\xi_1, \xi_2)$ is convex and lower semi-continuous on Ξ_1 for each $\xi_2 \in \Xi_2$;*
- *$\xi_2 \mapsto \mathcal{H}(\xi_1, \xi_2)$ is concave on Ξ_2 for each $\xi_1 \in \Xi_1$.*

Then

$$\inf_{\xi_1 \in \Xi_1} \sup_{\xi_2 \in \Xi_2} \mathcal{H}(\xi_1, \xi_2) = \sup_{\xi_2 \in \Xi_2} \inf_{\xi_1 \in \Xi_1} \mathcal{H}(\xi_1, \xi_2).$$

The two conditions of Theorem C.1 are the heart of the matter. Convex (and lower semi-continuous) behavior in the first variable together with concave behavior in the second guarantees that minimizing and then maximizing yields the same value

as maximizing and then minimizing. Compactness and convexity of the first domain ensure that the relevant extrema are attained.

In the application to Problem 6.2, the first variable is the indemnity function and the second is the loss distribution. The set of admissible indemnities $\mathcal{I}$ in (6.3) plays the role of Ξ_1. It is convex and, under the metric $d(I_1, I_2) = \max_{t\in[0,M]} |I_1(t) - I_2(t)|$, it is compact by the Arzelà–Ascoli theorem, since its elements are uniformly bounded and 1-Lipschitz. The ambiguity set $\mathcal{P}_\epsilon$ in (6.8) plays the role of Ξ_2 and is convex. The objective

$$\mathcal{H}(I, P) = \rho_g^P\Big(X - I(X) + \pi(I)\Big)$$

is linear, and in particular convex and continuous, in I, because ρ_g^P is translation invariant and comonotone additive, and it is concave in F_P because the distortion function g is concave. Both conditions of Theorem C.1 are thus satisfied, and the inf–sup formulation of Problem 6.2 coincides with its sup–inf counterpart (6.10). This interchange is what allows the inner problem to be solved for a fixed distribution, and it is therefore the structural pivot on which the entire solution rests.

C.1.2 Slater's Condition and Strong Duality

Once the optimal indemnity has been substituted, the outer problem becomes a constrained maximization over survival functions; see (6.12). To solve it through its Lagrangian dual, one needs strong duality, which is ensured by Slater's constraint qualification.

Theorem C.2 (Slater's condition)**.** *[12] Consider a convex optimization problem, that is, the maximization of a concave objective subject to convex inequality constraints. If there exists a strictly feasible point, which is a point satisfying every inequality constraint with strict inequality, then strong duality holds. The optimal values of the primal and the dual problem coincide, and an optimal Lagrange multiplier exists.*

In our setting the decision variable is the survival function S_P. The objective of (6.12) is concave in S_P, since the distortion function g is concave, and the Wasserstein constraint, once expressed in terms of survival functions, is convex in S_P. The problem is therefore convex in the sense of Theorem C.2. A strictly feasible point is readily available: because $\epsilon > 0$, the benchmark survival function $S_{\mathbb{B}}$ itself satisfies the Wasserstein constraint strictly, as it lies at the centre of the ball. Slater's condition thus holds, strong duality applies, and the problem can be solved through its dual, in which a multiplier $\beta \geq 0$ enforces the Wasserstein constraint. The two cases $\beta = 0$ and $\beta > 0$ of Section 6.3 correspond, respectively, to a slack and a binding constraint, the multiplier being determined in the latter case by the requirement that the constraint hold with equality.

C.2 Implementation Code

The numerical study in Chapter 6 was implemented in Python. The code reproduces all figures and results presented there, including the optimal stop-loss contracts and worst-case distributions for the GlueV@R risk measure under the Wasserstein and L^2 ambiguity sets. The complete, documented source code is available at https://github.com/mariazeppa/Distributionally-Robust-Insurance.git.

References

[1] Carlo Acerbi. 'Spectral Measures of Risk: A Coherent Representation of Subjective Risk Aversion'. In: *Journal of Banking & Finance* 26.7 (2002), pp. 1505–1518.

[2] Leonidas Alaoglu. 'Weak Topologies of Normed Linear Spaces'. In: *Annals of Mathematics (2)* 41.1 (1940), pp. 252–267.

[3] Philippe Artzner, Freddy Delbaen, Jean-Marc Eber and David Heath. 'Coherent Measures of Risk'. In: *Mathematical Finance* 9.3 (1999), pp. 203–228.

[4] Hirbod Assa. 'On optimal reinsurance policy with distortion risk measures and premiums'. In: *Insurance: Mathematics and Economics* 61 (2015), pp. 70–75.

[5] Stefan Banach. 'Sur les fonctionnelles linéaires I, II'. In: *Studia Mathematica* 1 (1929), pp. 211–216, 223–239.

[6] Nicole Bäuerle and Alfred Müller. 'Stochastic Orders and Risk Measures: Consistency and Bounds'. In: *Insurance: Mathematics and Economics* 38.1 (2006), pp. 132–148.

[7] Jaume Belles-Sampera, Montserrat Guillén and Miguel Santolino. 'Beyond Value-at-Risk: GlueVaR Distortion Risk Measures'. In: *Risk Analysis* 34.1 (2014), pp. 121–134.

[8] Carole Bernard, Silvana M. Pesenti and Steven Vanduffel. 'Robust distortion risk measures'. In: *Mathematical Finance* 34.3 (2024), pp. 774–818.

[9] Jean Bertoin. *Probabilités: Cours de licence de mathématiques appliquées (LM 390)*. Lecture notes. Université Pierre et Marie Curie (Paris 6), 2007.

[10] Tim J. Boonen and Wenjun Jiang. 'Robust Insurance Design with Distortion Risk Measures'. In: *European Journal of Operational Research* 316.2 (2024), pp. 694–706.

[11] Tim J. Boonen and Wenjun Jiang. 'Distributionally robust insurance under the Wasserstein distance'. In: *Insurance: Mathematics and Economics* 120 (2025), pp. 61–78.

[12] Stephen P. Boyd and Lieven Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004.

[13] Haim Brezis. *Functional Analysis, Sobolev Spaces and Partial Differential Equations*. Universitext. New York: Springer, 2011.

[14] Patrick Cheridito. *Convex Analysis*. Lecture notes. Princeton University, 2013.

[15] Ka Chun Cheung, Wing Fung Chong and Sheung Chi Phillip Yam. 'Convex Ordering for Insurance Preferences'. In: *Insurance: Mathematics and Economics* 64 (2015), pp. 409–416.

[16] Delia Coculescu. *FIM 102 Mathematical Finance I, Lecture Notes Spring Semester 2025*. Lecture notes. University of Zurich, 2025.

[17] Delia Coculescu. *FIM 202 Mathematical Finance II, Lecture Notes Fall Semester 2025*. Lecture notes. University of Zurich, 2025.

[18] Freddy Delbaen. *Draft: Coherent Risk Measures*. Lecture notes. Scuola Normale di Pisa, 2000.

[19] Freddy Delbaen. 'Coherent Risk Measures on General Probability Spaces'. In: *Advances in Finance and Stochastics: Essays in Honour of Dieter Sondermann*. Ed. by Klaus Sandmann and Philipp J. Schönbucher. Berlin: Springer, 2002, pp. 1–37.

[20] Dieter Denneberg. *Non-Additive Measure and Integral*. Vol. 27. Theory and Decision Library B. Dordrecht: Kluwer Academic Publishers, 1994.

[21] Michel Denuit, Jan Dhaene, Marc Goovaerts and Rob Kaas. *Actuarial Theory for Dependent Risks: Measures, Orders and Models*. John Wiley & Sons, 2006.

[22] Ky Fan. 'Minimax theorems'. In: *Proceedings of the National Academy of Sciences* 39.1 (1953), pp. 42–47.

[23] Werner Fenchel. 'On Conjugate Convex Functions'. In: *Canadian Journal of Mathematics* 1.1 (1949), pp. 73–77.

[24] G. Fichtenholz and L. Kantorovich. 'Sur les opérations linéaires dans l'espace des fonctions bornées'. In: *Studia Mathematica* 5 (1934), pp. 69–98.

[25] Hans Föllmer and Alexander Schied. 'Convex Measures of Risk and Trading Constraints'. In: *Finance and Stochastics* 6.4 (2002), pp. 429–447.

[26] Hans Föllmer and Alexander Schied. *Stochastic Finance: An Introduction in Discrete Time*. 5th revised and extended edition. De Gruyter, 2025.

[27] Marco Frittelli and Emanuela Rosazza Gianin. 'Putting Order in Risk Measures'. In: *Journal of Banking & Finance* 26.7 (2002), pp. 1473–1486.

[28] Marco Frittelli and Emanuela Rosazza Gianin. 'Law Invariant Convex Risk Measures'. In: *Advances in Mathematical Economics* 7 (2005), pp. 33–46.

[29] Michael Grabisch. *Set Functions, Games and Capacities in Decision Making*. Springer, 2016.

[30] Hans Hahn. 'Über lineare Gleichungssysteme in linearen Räumen'. In: *Journal für die reine und angewandte Mathematik* 157 (1927), pp. 214–229.

[31] T. H. Hildebrandt. 'On Bounded Linear Functional Operations'. In: *Transactions of the American Mathematical Society* 36.4 (1934), pp. 868–875.

[32] Gur Huberman, David Mayers and Clifford W. Smith Jr. 'Optimal insurance policy indemnity schedules'. In: *The Bell Journal of Economics* 14.2 (1983), pp. 415–426.

[33] Elyes Jouini, Walter Schachermayer and Nizar Touzi. 'Law Invariant Risk Measures Have the Fatou Property'. In: *Advances in Mathematical Economics* 9 (2006), pp. 49–71.

[34] Mark Krein and Vitold Šmulian. 'On Regularly Convex Sets in the Space Conjugate to a Banach Space'. In: *Annals of Mathematics (2)* 41.3 (1940), pp. 556–583.

[35] Shigeo Kusuoka. 'On Law Invariant Coherent Risk Measures'. In: *Advances in Mathematical Economics* 3 (2001), pp. 83–95.

[36] Stanisław Mazur. 'Über konvexe Mengen in linearen normierten Räumen'. In: *Studia Mathematica* 4 (1933), pp. 70–84.

[37] Hermann Minkowski. 'Theorie der konvexen Körper, insbesondere Begründung ihres Oberflächenbegriffs'. In: *Gesammelte Abhandlungen*. Ed. by David Hilbert. Vol. 2. Leipzig and Berlin: B. G. Teubner, 1911, pp. 131–229.

[38] Jean-Jacques Moreau. 'Fonctions convexes en dualité'. In: *Séminaires de Mathématiques, Faculté des Sciences de Montpellier* (1962). Multigraph.

[39] Victor M. Panaretos and Yoav Zemel. 'Statistical aspects of Wasserstein distances'. In: *Annual Review of Statistics and its Application* 6 (2019), pp. 405–431.

[40] R. Tyrrell Rockafellar. *Convex Analysis*. Princeton Mathematical Series 28. Princeton, NJ: Princeton University Press, 1970.

[41] R. Tyrrell Rockafellar and Stanislav Uryasev. 'Optimization of Conditional Value-at-Risk'. In: *Journal of Risk* 2.3 (2000), pp. 21–41.

[42] R. Tyrrell Rockafellar and Stanislav Uryasev. 'Conditional Value-at-Risk for General Loss Distributions'. In: *Journal of Banking & Finance* 26.7 (2002), pp. 1443–1471.

[43] Benjamin Schlein. *Skript zur Vorlesung Analysis 2, Frühling 2022*. Lecture notes. Universität Zürich, 2022.

[44] Benjamin Schlein. *Skript zur Vorlesung Analysis 3, Herbstsemester 2022*. Lecture notes. Universität Zürich, 2022.

[45] David Schmeidler. 'Integral Representation Without Additivity'. In: *Proceedings of the American Mathematical Society* 97.2 (1986), pp. 255–261.

[46] David Schmeidler. 'Subjective Probability and Expected Utility Without Additivity'. In: *Econometrica* 57.3 (1989), pp. 571–587.

[47] Alexander Shapiro. 'On Kusuoka Representation of Law Invariant Risk Measures'. In: *Mathematics of Operations Research* 38.1 (2013), pp. 142–152.

[48] Gregor Svindland. 'Continuity Properties of Law-Invariant (Quasi-)Convex Risk Functions on L^∞'. In: *Mathematics and Financial Economics* 3.1 (2010), pp. 39–43.

[49] Hugo Touchette. *Legendre-Fenchel transforms in a nutshell*. Lecture notes. School of Mathematical Sciences, Queen Mary, University of London, London E1 4NS, UK, 2005.

[50] Cédric Villani. *Optimal Transport: Old and New*. Vol. 338. Grundlehren der mathematischen Wissenschaften. Springer, 2009.

[51] Shaun S. Wang, Virginia R. Young and Harry H. Panjer. 'Axiomatic characterization of insurance prices'. In: *Insurance: Mathematics and Economics* 21.2 (1997), pp. 173–183.

[52] Sheng Chao Zhuang, Chengguo Weng, Ken Seng Tan and Hirbod Assa. 'Marginal indemnification function formulation for optimal reinsurance'. In: *Insurance: Mathematics and Economics* 67 (2016), pp. 65–76.

Declarations

Use of Generative AI

During the preparation of this thesis, generative artificial intelligence tools were used to support language editing, improve readability, refine the structure of the text, assist in digitizing figures, and support the implementation of the replication. All mathematical results, proofs, interpretations, and the final scientific content were independently verified by the author, who takes full responsibility for the accuracy and originality of the thesis.

Declaration of Authenticity

I hereby declare that I have written this thesis independently and have not used any sources or aids other than those indicated. All passages taken verbatim or in substance from published or unpublished sources are identified as such. Any use of generative AI is fully disclosed above. I am aware that violations of this declaration may have legal and disciplinary consequences.

Zürich, 7th September 2026 Maria Elena Zeppa