Optimization of Risk Measures

Andrzej Ruszczyński[1] **and Alexander Shapiro**[2]

[1] Rutgers University, Piscataway, NJ 08854, USA,
e-mail: `rusz@rutcor.rutgers.edu`
[2] Georgia Institute of Technology, Atlanta, Georgia 30332-0205, USA,
e-mail: `ashapiro@isye.gatech.edu`

1 Introduction

Consider a stochastic system whose output variable Z is a real valued random variable. If it depends on some decision vector $x \in \mathbb{R}^n$, we can write the relation:

$$Z(\omega) = f(x,\omega), \quad \omega \in \Omega.$$

Here $f : \mathbb{R}^n \times \Omega \to \mathbb{R}$, and $(\Omega, \mathcal{F})$ is a measurable space. To focus attention, we shall be interested in the case when smaller values of Z are "better", for example, Z may represent random cost or losses. It will be obvious how our considerations can be adapted to the case of reverse preferences.

In order to find the "best" values of the decision vector x we can formulate the stochastic optimization problem:

$$\operatorname*{Min}_{x \in S} \big\{ \phi(x) := \mathbb{E}_P[f(x,\omega)] \big\}, \tag{1}$$

where $S \subset \mathbb{R}^n$ is a set of feasible decision vectors, and P is a probability measure (distribution) on the sample space $(\Omega, \mathcal{F})$. The theory of such stochastic optimization problems and numerical methods for their solution are well developed (see [22]).

There are two basic difficulties associated with the above formulation. First, it is assumed that the probability distribution P is known. In real life applications the probability distribution is never known exactly. In some cases it can be estimated from historical data by statistical techniques. However, in many cases the probability distribution neither can be estimated accurately nor remains constant. Even worse, quite often one subjectively assigns certain weights (probabilities) to a finite number of possible realizations (called *scenarios*) of the uncertain data. Such a simplified model can hardly be considered an accurate description of the reality.

The second basic question is why we want to optimize the *expected* value of the random outcome Z. In some situations the same decisions under similar conditions are made repeatedly over a certain period of time. In such cases one can justify optimization of the expected value by arguing that, by the Law of Large Numbers, it gives an optimal decision on *average*. However, because of the variability of the data, the average of the first few results may be very bad. For example, one may lose all his investments, and it does not help that the decisions were optimal on average.

For these reasons, quantitative models of *risk* and *risk aversion* are needed. There exist several approaches to model decision making under risk. The classical approach is based on the *expected utility theory* of von Neumann and Morgenstern [25]. One specifies a *disutility function*[3] $g : \mathbb{R} \to \mathbb{R}$ and formulate the problem:

$$\operatorname*{Min}_{x \in S} \Big\{ \phi_g(x) := \mathbb{E}_P\big[g(f(x,\omega))\big] \Big\}. \tag{2}$$

Unfortunately, it is extremely difficult to elicit the disutility function of a decision maker.

The second approach is to specify *constraints* on risk. The most common is the *Value at Risk* constraint, which involves the critical value $z_{\max}$ allowed for risk exposure, and the probability $p_{\max}$ of excessive outcomes:

$$P\big[Z \geq z_{\max}\big] \leq p_{\max}.$$

In the stochastic optimization literature such constraints are called *probabilistic* or *chance constraints* [15]. Variations of this concept are known as integrated chance constraints [6], Conditional Value at Risk [20], or expected shortfall [1].

A direct way to deal with the issue of uncertain probability distribution, is to identify a plausible family $\mathcal{A}$ of probability distributions and, consequently, to consider the min-max problem

$$\operatorname*{Min}_{x \in S} \left\{ \phi(x) := \sup_{P \in \mathcal{A}} \mathbb{E}_P[f(x,\omega)] \right\}. \tag{3}$$

The idea of the worst-case (min-max) formulation is not new of course. It goes back to von Neumann's game theory and was already discussed, for example in the context of stochastic programming, in Žáčková [26] almost 40 years ago.

The attempts to overcome the drawbacks of the expected value optimization have also a long history. One can try to reach a compromise between the optimization on average and the minimization of a certain measure of the involved risk. This leads to the formulation

$$\operatorname*{Min}_{x \in S} \big\{ \phi(x) := \rho[F(x)] \big\}, \tag{4}$$

[3] We consider here minimization problems, and that is why we speak about disutility. Any disutility function g corresponds to a utility function $u : \mathbb{R} \to \mathbb{R}$ defined by $u(-z) = -g(z)$. Note that the function u is concave and increasing (nondecreasing) iff the function g is convex and increasing (nondecreasing).

where $\rho(Z)$ is a mean-risk measure, defined on a space of random variables $Z : \Omega \to \mathbb{R}$, and $[F(x)](\omega) = f(x, \omega)$. The classical mean-variance risk measure $\rho(Z) := \mathbb{E}[Z] + c\mathbb{V}\mathrm{ar}[Z]$, where c is a nonnegative constant, is going back to Markowitz [8].

There are several problems with the mean-variance risk measure. First, the expectation and variance are measured in different units. Secondly, the mean-variance model is not consistent with the classical relation of stochastic dominance, which formalizes risk-averse preferences [11].

In recent years risk analysis came under intensive investigation, in particular from the point of view of the optimization theory. In this chapter we discuss a general theory of optimization of risk measures. We show, in particular, that the above approaches of min-max formulation (3) and risk measure formulation (4), in a sense, are equivalent to each other.

We also introduce and analyze new models of dynamic optimization problems involving risk functions. We introduce the concept of conditional risk mappings, and we derive dynamic programming relations for the corresponding optimization models. In this way we provide an alternative approach to the recent works [3, 4, 14, 16], where various dynamic risk models are considered.

2 Risk Functions

In this section we give a formal definition of risk functions and we discuss their basic properties. Let $(\Omega, \mathcal{F})$ be a sample space, equipped with sigma algebra $\mathcal{F}$, on which considered uncertain outcomes (random functions $Z = Z(\omega)$) are defined. By a *risk function* we understand a function $\rho(Z)$ which maps Z into the extended real line $\overline{\mathbb{R}} = \mathbb{R} \cup \{+\infty\} \cup \{-\infty\}$. In order to make this concept precise we need to define a space $\mathcal{Z}$ of allowable random functions $Z(\omega)$ for which $\rho(Z)$ is defined. It seems that a natural choice of $\mathcal{Z}$ will be the space of all $\mathcal{F}$-measurable functions $Z : \Omega \to \mathbb{R}$. However, typically, this space is too large for development of a meaningful theory. In almost all interesting examples considered in this chapter we deal with the space[4] $\mathcal{Z} := \mathcal{L}_p(\Omega, \mathcal{F}, P)$. We will discuss an appropriate choice of the space $\mathcal{Z}$ later.

We assume throughout this chapter that $\mathcal{Z}$ is a linear space of $\mathcal{F}$-measurable functions and considered risk functions $\rho : \mathcal{Z} \to \overline{\mathbb{R}}$ are *proper*. That is, $\rho(Z) > -\infty$ for all $Z \in \mathcal{Z}$ and the domain

$$\mathrm{dom}(\rho) := \{Z \in \mathcal{Z} : \rho(Z) < +\infty\}$$

[4] Recall that $\mathcal{L}_p(\Omega, \mathcal{F}, P, \mathbb{R}^n)$ denotes the linear space of all $\mathcal{F}$-measurable functions $\psi : \Omega \to \mathbb{R}^n$ such that $\int_\Omega \|\psi(\omega)\|^p \, dP(\omega) < +\infty$. More precisely, an element of $\mathcal{L}_p(\Omega, \mathcal{F}, P, \mathbb{R}^n)$ is a class of such functions $\psi(\omega)$ which may differ from each other on sets of P-measure zero. For $n = 1$ we denote this space by $\mathcal{L}_p(\Omega, \mathcal{F}, P)$. Unless stated otherwise, while dealing with these spaces we assume that $p \in [1, +\infty)$, P is a probability measure on $(\Omega, \mathcal{F})$ and expectations are taken with respect to P. For $\psi \in \mathcal{L}_p(\Omega, \mathcal{F}, P)$, its norm $\|\psi\|_p := \left(\int_\Omega |\psi(\omega)|^p \, dP(\omega)\right)^{1/p}$.

is nonempty. We consider the following axioms associated with a risk function ρ. For $Z_1, Z_2 \in \mathcal{Z}$ we denote by $Z_2 \succeq Z_1$ the pointwise partial order meaning $Z_2(\omega) \geq Z_1(\omega)$ for all $\omega \in \Omega$.

(A1) *Convexity:*

$$\rho(\alpha Z_1 + (1-\alpha)Z_2) \leq \alpha\rho(Z_1) + (1-\alpha)\rho(Z_2)$$

for all $Z_1, Z_2 \in \mathcal{Z}$ and all $\alpha \in [0,1]$.

(A2) *Monotonicity:* If $Z_1, Z_2 \in \mathcal{Z}$ and $Z_2 \succeq Z_1$, then $\rho(Z_2) \geq \rho(Z_1)$.

(A3) *Translation Equivariance:* If $a \in \mathbb{R}$ and $Z \in \mathcal{Z}$, then $\rho(Z+a) = \rho(Z)+a$.

(A4) *Positive Homogeneity:* If $\alpha > 0$ and $Z \in \mathcal{Z}$, then $\rho(\alpha Z) = \alpha\rho(Z)$.

These axioms were introduced, and risk functions satisfying (A1)–(A4) were called *coherent risk measures*, in Artzner, Delbaen, Eber and Heath [2].

In order to proceed with the analysis we need to associate with the space $\mathcal{Z}$ a dual space $\mathcal{Z}^*$ of measures such that the scalar product

$$\langle \mu, Z \rangle := \int_\Omega Z(\omega)\, d\mu(\omega) \tag{5}$$

is well defined for all $Z \in \mathcal{Z}$ and $\mu \in \mathcal{Z}^*$. That is, we assume that $\mathcal{Z}^*$ is a linear space of finite signed measures[5] μ on $(\Omega, \mathcal{F})$ such that $\int_\Omega |Z|\, d|\mu| < +\infty$ for all $Z \in \mathcal{Z}$. We assume that $\mathcal{Z}$ and $\mathcal{Z}^*$ are *paired* (locally convex topological vector) spaces. That is, $\mathcal{Z}$ and $\mathcal{Z}^*$ are equipped with respective topologies which make them locally convex topological vector spaces and these topologies are compatible with the scalar product (5), i.e., every linear continuous functional on $\mathcal{Z}$ can be represented in the form $\langle \mu, \cdot \rangle$ for some $\mu \in \mathcal{Z}^*$, and every linear continuous functional on $\mathcal{Z}^*$ can be represented in the form $\langle \cdot, Z \rangle$ for some $Z \in \mathcal{Z}$. In particular, we can equip each space $\mathcal{Z}$ and $\mathcal{Z}^*$ with its weak topology induced by its paired space. This will make $\mathcal{Z}$ and $\mathcal{Z}^*$ paired locally convex topological vector spaces provided that for any $Z \in \mathcal{Z} \setminus \{0\}$ there exists $\mu \in \mathcal{Z}^*$ such that $\langle \mu, Z \rangle \neq 0$, and for any $\mu \in \mathcal{Z}^* \setminus \{0\}$ there exists $Z \in \mathcal{Z}$ such that $\langle \mu, Z \rangle \neq 0$.

If $\mathcal{Z} := \mathcal{L}_p(\Omega, \mathcal{F}, P)$, we can consider its dual space $\mathcal{Z}^* := \mathcal{L}_q(\Omega, \mathcal{F}, P)$, where $q \in (1, +\infty]$ is such that $1/p + 1/q = 1$. Here $\mathcal{Z}$, equipped with the respective norm, is a Banach space and $\mathcal{Z}^*$ is its dual Banach space. In order to make these spaces paired spaces we can equip $\mathcal{Z}$ with its strong (norm) topology and $\mathcal{Z}^*$ with its weak* topology. Moreover, if $p \in (1, +\infty)$, then $\mathcal{Z}$ and $\mathcal{Z}^*$ are reflexive Banach spaces. In that case, they are paired spaces when equipped with their strong topologies. Note also that in this case every measure $\mu \in \mathcal{Z}^*$ has a density $\zeta \in \mathcal{L}_q(\Omega, \mathcal{F}, P)$, i.e., $d\mu = \zeta dP$. When dealing

[5] Recall that a finite signed measure μ can be represented in the form $\mu = \mu^+ - \mu^-$, where μ^+ and μ^- are nonnegative finite measures on $(\Omega, \mathcal{F})$. This representation is called the Jordan decomposition of μ. The measure $|\mu| = \mu^+ + \mu^-$ is called the total variation of μ.

with these spaces we identify the corresponding measure with its density and for $Z \in \mathcal{L}_p(\Omega, \mathcal{F}, P)$ and $\zeta \in \mathcal{L}_q(\Omega, \mathcal{F}, P)$ we use the scalar product

$$\langle \zeta, Z \rangle := \int_\Omega \zeta(\omega) Z(\omega)\, dP(\omega). \tag{6}$$

Unless stated otherwise we always assume the following.

(C) For every $A \in \mathcal{F}$ the space $\mathcal{Z}$ contains the indicator[6] function $\mathbb{1}_A$.

Since the space $\mathcal{Z}$ is linear, this implies that $\mathcal{Z}$ contains all step functions of the form $\sum_{i=1}^m \alpha_i \mathbb{1}_{A_i}$, where $a_i \in \mathbb{R}$ and $A_i \in \mathcal{F}$, $i = 1, \dots, m$. This holds true, in particular, for every space $\mathcal{Z} := \mathcal{L}_p(\Omega, \mathcal{F}, P)$.

The partial order in the space $\mathcal{Z}$, appearing in condition (A2), is defined by the cone

$$\mathcal{Z}_+ := \{Z \in \mathcal{Z} : Z(\omega) \geq 0,\ \forall\, \omega \in \Omega\},$$

i.e., $Z_2 \succeq Z_1$ iff $Z_2 - Z_1 \in \mathcal{Z}_+$. Consider the cone $\mathcal{Z}^*_+$ of all nonnegative measures in the space $\mathcal{Z}^*$. For any $Z \in \mathcal{Z}_+$ and any $\mu \in \mathcal{Z}^*_+$, we have that $\langle \mu, Z \rangle \geq 0$. Moreover, because of assumption (C) above, we have that $\mathcal{Z}^*_+$ coincides with the dual cone of the cone $\mathcal{Z}_+$, which is defined as the set of $\mu \in \mathcal{Z}^*$ such that $\langle \mu, Z \rangle \geq 0$ for all $Z \in \mathcal{Z}_+$.

We can now formulate the basic (conjugate) duality result. Recall that the conjugate function $\rho^* : \mathcal{Z}^* \to \overline{\mathbb{R}}$ of a risk function ρ is defined as

$$\rho^*(\mu) := \sup_{Z \in \mathcal{Z}} \big\{ \langle \mu, Z \rangle - \rho(Z) \big\}, \tag{7}$$

and the conjugate of ρ^* (the biconjugate function) as

$$\rho^{**}(Z) := \sup_{\mu \in \mathcal{Z}^*} \big\{ \langle \mu, Z \rangle - \rho^*(\mu) \big\}. \tag{8}$$

By $\operatorname{lsc}(\rho)$ we denote the lower semicontinuous hull of ρ taken with respect to the considered topology of $\mathcal{Z}$. The following is the basic duality result of convex analysis (see, e.g., [17, Theorem 5] for a proof).

Theorem 1 (Fenchel-Moreau). *Suppose that function $\rho : \mathcal{Z} \to \overline{\mathbb{R}}$ is convex and proper. Then $\rho^{**} = \operatorname{lsc}(\rho)$.*

It follows that if ρ is convex and proper, then the representation

$$\rho(Z) = \sup_{\mu \in \mathcal{Z}^*} \big\{ \langle \mu, Z \rangle - \rho^*(\mu) \big\} \tag{9}$$

holds true if ρ is lower semicontinuous. Conversely, if (9) is satisfied for some function $\rho^*(\cdot)$, then ρ is lower semicontinuous and convex. Note also that if ρ is proper, lower semicontinuous and convex, then its conjugate function ρ^*

[6] Recall that the indicator function $\mathbb{1}_A$ is defined as $\mathbb{1}_A(\omega) = 1$ for $\omega \in A$ and $\mathbb{1}_A(\omega) = 0$ for $\omega \notin A$.

is proper. Let us also remark that if $\mathcal{Z}$ is a Banach space and $\mathcal{Z}^*$ is its dual (e.g., $\mathcal{Z} = \mathcal{L}_p(\Omega, \mathcal{F}, P)$ and $\mathcal{Z}^* = \mathcal{L}_q(\Omega, \mathcal{F}, P)$) and ρ is convex, then ρ is lower semicontinuous in the weak topology iff it is lower semicontinuous in the strong (norm) topology.

If the set Ω is finite, say $\Omega = \{\omega_1, \dots, \omega_K\}$, then the technical level of the analysis simplifies considerably. Every function $Z \in \mathcal{Z}$ can be identified with the vector $(Z(\omega_1), \dots, Z(\omega_K))$. Thus the space $\mathcal{Z}$ is finite dimensional, $\mathcal{Z} = \mathbb{R}^K$, and can be paired with itself. Moreover, in the finite dimensional case, if ρ is proper and convex, then it is continuous (and hence lower semicontinuous) at every point in the interior of its domain. In particular, it is continuous at every point if it is real valued. In order to avoid technical details one can be tempted to restrict the discussion to finite sample spaces. However, apart from restricting the generality, this would result in losing some important essentials of the analysis. It turns out that some important properties enjoyed by risk functions for continuous distributions do not extend to the discrete case of finite Ω (see the examples in the next section).

As it was discussed above, in order for the representation (9) to hold we only need the convexity (condition (A1)) and lower semicontinuity properties to be satisfied. Let us observe that (9) is equivalent to

$$\rho(Z) = \sup_{\mu \in \mathcal{A}} \big\{ \langle \mu, Z \rangle - \rho^*(\mu) \big\}, \tag{10}$$

where

$$\mathcal{A} := \{\mu \in \mathcal{Z}^* : \rho^*(\mu) < +\infty\} \tag{11}$$

is the domain of ρ^*. It is not difficult to show that if representation (9) (or, equivalently, representation (10)) holds true, then condition (A2) is satisfied iff the set $\mathcal{A}$ contains only nonnegative measures, and condition (A3) is satisfied iff $\mu(\Omega) = 1$ for every $\mu \in \mathcal{A}$ (cf., [23]). We obtain that if conditions (A1)–(A3) are satisfied and ρ is lower semicontinuous, then the representation (10) holds true with $\mathcal{A} \subset \mathcal{P}$, where $\mathcal{P}$ denotes the set of all probability measures in the space $\mathcal{Z}^*$.

Moreover, if $\mathcal{Z}$ is a Banach lattice[7] and ρ satisfies conditions (A1) and (A2), then ρ is continuous at every point[8] $Z \in \operatorname{int}(\operatorname{dom}(\rho))$ ([23]). Note that every space $\mathcal{L}_p(\Omega, \mathcal{F}, P)$ is a Banach lattice. Also if ρ is positively homogeneous, then $\rho^*(\mu) = 0$ for $\mu \in \mathcal{A}$ and $\rho^*(\mu) = +\infty$ otherwise. Therefore we have the following. Recall that

$$\mathcal{P} := \big\{ \zeta \in \mathcal{L}_q(\Omega, \mathcal{F}, P) : \textstyle\int_\Omega \zeta(\omega)\, dP(\omega) = 1,\ \zeta \succeq 0 \big\} \tag{12}$$

[7] It is said that a Banach space $\mathcal{Z}$ is a Banach lattice, with respect to the considered partial order defined by the cone $\mathcal{Z}_+$, if $\mathcal{Z}$ is a lattice, i.e., for any $Z_1, Z_2 \in \mathcal{Z}$ the element $\max\{Z_1(\cdot), \mathcal{Z}_2(\cdot)\}$ also belongs to $\mathcal{Z}$, and moreover if $|Z_1(\cdot)| \le |Z_2(\cdot)|$, then $\|Z_1\| \le \|Z_2\|$.

[8] We denote by $\operatorname{int}(\operatorname{dom}(\rho))$ the interior of the domain of ρ. That is, $Z \in \operatorname{int}(\operatorname{dom}(\rho))$ if there is a neighborhood $\mathcal{N}$ of Z such that $\rho(Z')$ is finite for all $Z' \in \mathcal{N}$.

denotes the set of probability measures in the dual space $\mathcal{L}_q(\Omega, \mathcal{F}, P)$.

Theorem 2. *Suppose that $\mathcal{Z} := \mathcal{L}_p(\Omega, \mathcal{F}, P)$, risk function $\rho : \mathcal{Z} \to \overline{\mathbb{R}}$ is proper and conditions* (A1)–(A3) *are satisfied. Then for all $Z \in \operatorname{int}(\operatorname{dom}(\rho))$ it holds that*

$$\rho(Z) = \sup_{\zeta \in \mathcal{P}} \left\{ \langle \zeta, Z \rangle - \rho^*(\zeta) \right\}. \tag{13}$$

If, moreover, ρ is positively homogeneous, then there exists a nonempty convex closed set $\mathcal{A} \subset \mathcal{P}$ such that for all $Z \in \operatorname{int}(\operatorname{dom}(\rho))$ it holds that

$$\rho(Z) = \sup_{\zeta \in \mathcal{A}} \langle \zeta, Z \rangle. \tag{14}$$

In this way we have established the equivalent representation of convex risk functions, which corresponds to the min-max model (3).

In various forms of generality the above dual representations of convex risk functions were derived in [2, 5, 21, 23]. If the set Ω is finite, say $\Omega = \{\omega_1, \dots, \omega_K\}$ with respective (positive) probabilities $p_1, \dots, p_K$, then the corresponding set

$$\mathcal{P} = \left\{ \zeta \in \mathbb{R}^K : \textstyle\sum_{k=1}^K p_k \zeta_k = 1,\ \zeta \geq 0 \right\}$$

is bounded, and hence the set $\mathcal{A}$ is also bounded. It follows that if Ω is finite and ρ is proper and conditions (A1)–(A4) are satisfied, then $\rho(\cdot)$ is real valued and representation (14) holds.

3 The Utility Model

It is also possible to relate the theory of convex risk functions with the utility model (2). Let $\mathcal{Z} := \mathcal{L}_p(\Omega, \mathcal{F}, P)$ and $\mathcal{Z}^* := \mathcal{L}_q(\Omega, \mathcal{F}, P)$, and let $g : \mathbb{R} \to \overline{\mathbb{R}}$ be a proper convex lower semicontinuous function such that the expectation $\mathbb{E}[g(Z)]$ is well defined[9] for all $Z \in \mathcal{Z}$. We can view the function g as a disutility function. Consider the risk function

$$\rho(Z) := \mathbb{E}[g(Z)] \tag{15}$$

and assume that ρ is proper. Since g is lower semicontinuous and convex, we have that

$$g(z) = \sup_{\alpha \in \mathbb{R}} \left\{ \alpha z - g^*(\alpha) \right\},$$

where g^* is the conjugate of g. As g is proper, the conjugate function g^* is also proper. It follows that

[9] It is allowed here for $\mathbb{E}[g(Z)]$ to take value $+\infty$, but not $-\infty$ since the corresponding risk function is required to be proper.

$$\rho(Z) = \mathbb{E}\left[\sup_{\alpha\in\mathbb{R}} \{\alpha Z - g^*(\alpha)\}\right]. \tag{16}$$

We use the following interchangeability principle (e.g., Rockafellar and Wets [19, Theorem 14.60]). It is said that a linear space $\mathcal{M}$ of $\mathcal{F}$-measurable functions $\psi : \Omega \to \mathbb{R}^m$ is *decomposable* if for every $\psi \in \mathcal{M}$ and $B \in \mathcal{F}$, and every bounded and $\mathcal{F}$-measurable function $W : \Omega \to \mathbb{R}^m$, the space $\mathcal{M}$ also contains the function $V(\cdot) := \mathbb{1}_{\Omega\setminus B}(\cdot)\psi(\cdot) + \mathbb{1}_B(\cdot)W(\cdot)$. In the subsequent analysis we work with spaces $\mathcal{M} := \mathcal{L}_p(\Omega, \mathcal{F}, P, \mathbb{R}^m)$ which are decomposable. Now let $\mathcal{M}$ be a decomposable space and $h : \mathbb{R}^m \times \Omega \to \overline{\mathbb{R}}$ be a random lower semicontinuous function[10]. Then

$$\mathbb{E}\left[\inf_{y\in\mathbb{R}^m} h(y,\omega)\right] = \inf_{Y\in\mathcal{M}} \mathbb{E}\big[H_Y\big], \tag{17}$$

where $H_Y(\omega) := h(Y(\omega), \omega)$, provided that the right hand side of (17) is less than $+\infty$. Moreover, if the common value of both sides in (17) is not $-\infty$, then

$$\bar{Y} \in \operatorname*{argmin}_{Y\in\mathcal{M}} \mathbb{E}[H_Y] \;\text{ iff }\; \bar{Y}(\omega) \in \operatorname*{argmin}_{y\in\mathbb{R}^m} h(y,\omega) \;\text{ for a.e. } \omega \in \Omega.$$

Clearly the above interchangeability principle can be applied to a maximization, rather than minimization, procedure simply by replacing function $h(y,\omega)$ with $-h(y,\omega)$.

Let us return to the dual formulation (16) of the risk function (15). By using the interchangeability formula (17) with $h(\alpha,\omega) := -[\alpha Z(\omega) - g^*(\alpha)]$ we obtain

$$\rho(Z) = \sup_{\zeta\in\mathcal{Z}^*} \big\{\langle\zeta, Z\rangle - \mathbb{E}[g^*(\zeta)]\big\}. \tag{18}$$

It follows that ρ is convex and lower semicontinuous, and representation (9) holds with

$$\rho^*(\zeta) = \mathbb{E}[g^*(\zeta)].$$

Moreover, if the function g is nondecreasing, then ρ satisfies the monotonicity condition (A2). However, the risk function ρ does not satisfy condition (A3) unless $g(z) \equiv z$, and ρ is not positively homogeneous unless g is positively homogeneous.

4 Examples of Risk Functions

In this section we discuss several examples of risk functions which are commonly used in applications. In the following, P is a (reference) probability measure on $(\Omega, \mathcal{F})$ and, unless stated otherwise, all expectations and probabilistic statements are made with respect to P.

---

[10] A function $h : \mathbb{R}^m \times \Omega \to \overline{\mathbb{R}}$ is said to be *random lower semicontinuous* if its epigraphical mapping is closed valued and measurable. Random lower semicontinuous functions are also called *normal integrands* (cf., [19, Definition 14.27]).

Example 1 (Mean-variance risk function). Consider

$$\rho(Z) := \mathbb{E}[Z] + c\mathbb{V}\mathrm{ar}[Z], \tag{19}$$

where $c \geq 0$ is a given constant. It is natural to use here the space $\mathcal{Z} := \mathcal{L}_2(\Omega, \mathcal{F}, P)$ since for any $Z \in \mathcal{L}_2(\Omega, \mathcal{F}, P)$ the expectation $\mathbb{E}[Z]$ and variance $\mathbb{V}\mathrm{ar}[Z]$ are well defined and finite.

By direct calculation we can verify that

$$\mathbb{V}\mathrm{ar}[Z] = \big\|Z - \mathbb{E}[Z]\big\|^2 = \sup_{\zeta \in \mathcal{Z}} \big\{ \langle \zeta, Z - \mathbb{E}[Z] \rangle - \tfrac{1}{4}\|\zeta\|^2 \big\},$$

where the scalar products and the norms are in the sense of the (Hilbert) space $\mathcal{L}_2(\Omega, \mathcal{F}, P)$. Since $\langle \zeta, Z - \mathbb{E}[Z] \rangle = \langle \zeta - \mathbb{E}[\zeta], Z \rangle$ we can rewrite the last expression as follows:

$$\begin{aligned} \mathbb{V}\mathrm{ar}[Z] &= \sup_{\zeta \in \mathcal{Z}} \big\{ \langle \zeta - \mathbb{E}[\zeta], Z \rangle - \tfrac{1}{4}\|\zeta\|^2 \big\} \\ &= \sup_{\zeta \in \mathcal{Z}} \big\{ \langle \zeta - \mathbb{E}[\zeta], Z \rangle - \tfrac{1}{4}\mathbb{V}\mathrm{ar}[\zeta] - \tfrac{1}{4}\big(\mathbb{E}[\zeta]\big)^2 \big\}. \end{aligned}$$

Consequently, the above maximization can be restricted to such $\zeta \in \mathcal{Z}$ that $\mathbb{E}[\zeta] = 0$, and hence

$$\mathbb{V}\mathrm{ar}[Z] = \sup_{\substack{\zeta \in \mathcal{Z} \\ \mathbb{E}[\zeta]=0}} \big\{ \langle \zeta, Z \rangle - \tfrac{1}{4}\mathbb{V}\mathrm{ar}[\zeta] \big\}.$$

Therefore the risk function ρ, defined in (19), can be equivalently expressed for $c > 0$ as follows:

$$\begin{aligned} \rho(Z) &= \mathbb{E}[Z] + c \sup_{\substack{\zeta \in \mathcal{Z} \\ \mathbb{E}[\zeta]=0}} \Big\{ \langle \zeta, Z \rangle - \tfrac{1}{4}\mathbb{V}\mathrm{ar}\ [\zeta] \Big\} \\ &= \sup_{\substack{\zeta \in \mathcal{Z} \\ \mathbb{E}[\zeta]=1}} \Big\{ \langle \zeta, Z \rangle - \frac{1}{4c}\mathbb{V}\mathrm{ar}[\zeta] \Big\}. \end{aligned} \tag{20}$$

It follows that for any $c \geq 0$ the function ρ is convex and lower semicontinuous. Furthermore

$$\rho^*(\zeta) = \begin{cases} \frac{1}{4c}\mathbb{V}\mathrm{ar}[\zeta], & \text{if } \mathbb{E}[\zeta] = 1, \\ +\infty, & \text{otherwise.} \end{cases}$$

The function ρ satisfies the translation equivariance condition (A3), because the domain of its conjugate contains only ζ such that $\mathbb{E}[\zeta] = 1$. However, for any $c \geq 0$ the function ρ is not positively homogeneous and it does not satisfy the monotonicity condition (A2), because the domain of ρ^* contains density functions which are not nonnegative.

Example 2 (Mean-deviation risk function of order p). For $\mathcal{Z} := \mathcal{L}_p(\Omega, \mathcal{F}, P)$, $\mathcal{Z}^* := \mathcal{L}_q(\Omega, \mathcal{F}, P)$ and $c \geq 0$ consider

$$\rho(Z) := \mathbb{E}[Z] + c\left(\mathbb{E}\big[|Z - \mathbb{E}[Z]|^p\big]\right)^{1/p}. \tag{21}$$

Note that $\left(\mathbb{E}\big[|Z|^p\big]\right)^{1/p} = \|Z\|_p$, where $\|\cdot\|_p$ denotes the norm of the space $\mathcal{L}_p(\Omega, \mathcal{F}, P)$. We have that

$$\|Z\|_p = \sup_{\|\zeta\|_q \leq 1} \langle \zeta, Z \rangle,$$

and hence

$$\left(\mathbb{E}\big[|Z - \mathbb{E}[Z]|^p\big]\right)^{1/p} = \sup_{\|\zeta\|_q \leq 1} \langle \zeta, Z - \mathbb{E}[Z] \rangle = \sup_{\|\zeta\|_q \leq 1} \langle \zeta - \mathbb{E}[\zeta], Z \rangle.$$

It follows that representation (14) holds with the set $\mathcal{A}$ given by

$$\mathcal{A} = \left\{\zeta' \in \mathcal{Z}^* : \zeta' = 1 + \zeta - \mathbb{E}[\zeta],\ \|\zeta\|_q \leq c\right\}. \tag{22}$$

We obtain here that ρ satisfies conditions (A1), (A3) and (A4).

The monotonicity condition (A2) is more involved. Suppose that $p = 1$. Then $q = +\infty$ and hence for any $\zeta' \in \mathcal{A}$ and P-almost every $\omega \in \Omega$ we have

$$\zeta'(\omega) = 1 + \zeta(\omega) - \mathbb{E}[\zeta] \geq 1 - |\zeta(\omega)| - \mathbb{E}[\zeta] \geq 1 - 2c.$$

It follows that if $c \in [0, 1/2]$, then $\zeta'(\omega) \geq 0$ for P-almost every $\omega \in \Omega$, and hence condition (A2) follows. Conversely, take $\zeta := c(-\mathbb{1}_A + \mathbb{1}_{\Omega \setminus A})$, for some $A \in \mathcal{F}$, and $\zeta' = 1 + \zeta - \mathbb{E}[\zeta]$. We have that $\|\zeta\|_\infty = c$ and $\zeta'(\omega) = 1 - 2c + 2cP(A)$ for all $\omega \in A$ It follows that if $c > 1/2$, then $\zeta'(\omega) < 0$ for all $\omega \in A$, provided that $P(A)$ is small enough. We obtain that for $c > 1/2$ the monotonicity property (A2) does not hold if the following condition is satisfied:

$$\text{For any } \varepsilon > 0 \text{ there exists } A \in \mathcal{F} \text{ such that } \varepsilon > P(A) > 0. \tag{23}$$

That is, for $p = 1$ the mean-deviation function ρ satisfies (A2) if, and provided that condition (23) holds, only if $c \in [0, 1/2]$.

Suppose now that $p > 1$. For a set $A \in \mathcal{F}$ and $\alpha > 0$ let us take $\zeta := -\alpha \mathbb{1}_A$ and $\zeta' = 1 + \zeta - \mathbb{E}[\zeta]$. Then $\|\zeta\|_q = \alpha P(A)^{1/q}$ and $\zeta'(\omega) = 1 - \alpha + \alpha P(A)$ for all $\omega \in A$. It follows that if $p > 1$, then for any $c > 0$ the mean-deviation function ρ does not satisfy (A2) provided that condition (23) holds.

Example 3 (Mean-upper-semideviation risk function of order p). Let $\mathcal{Z} := \mathcal{L}_p(\Omega, \mathcal{F}, P)$ and for $c \geq 0$ consider[11]

$$\rho(Z) := \mathbb{E}[Z] + c\left(\mathbb{E}\left[\big[Z - \mathbb{E}[Z]\big]_+^p\right]\right)^{1/p}. \tag{24}$$

[11] We denote $[a]_+^p := (\max\{0, a\})^p$.

For any $c \geq 0$ this function satisfies conditions (A1), (A3) and (A4), and similarly to the derivations of Example 2 it can be shown that representation (14) holds with the set $\mathcal{A}$ given by

$$\mathcal{A} = \big\{\zeta' \in \mathcal{Z}^* : \zeta' = 1 + \zeta - \mathbb{E}[\zeta],\ \|\zeta\|_q \leq c,\ \zeta \succeq 0\big\}. \tag{25}$$

Since $|\mathbb{E}[\zeta]| \leq \mathbb{E}|\zeta| \leq \|\zeta\|_q$ for any $\zeta \in \mathcal{L}_q(\Omega, \mathcal{F}, P)$, we have that every element of the above set $\mathcal{A}$ is nonnegative and has its expected value equal to 1. This means that the monotonicity condition (A2) holds true, if and, provided that condition (23) holds, only if $c \in [0,1]$ (see [23]). That is, ρ is a coherent risk function if $c \in [0,1]$.

Example 4 (Mean-upper-semivariance from a target). Let $\mathcal{Z} := \mathcal{L}_2(\Omega, \mathcal{F}, P)$ and for weight $c \geq 0$ and target $\tau \in \mathbb{R}$ consider

$$\rho(Z) := \mathbb{E}[Z] + c\,\mathbb{E}\left[\big[Z-\tau\big]_+^2\right]. \tag{26}$$

We can now use (18) with $g(z) = z + c(z-\tau)_+^2$. Since

$$g^*(\alpha) = \begin{cases} (\alpha-1)^2/4c + \tau(\alpha-1), & \text{if } \alpha \geq 1, \\ +\infty, & \text{otherwise,} \end{cases}$$

we obtain that

$$\rho(Z) = \sup_{\zeta \in \mathcal{Z},\, \zeta(\cdot) \geq 1} \left\{ \mathbb{E}[\zeta Z] - \tau \mathbb{E}[\zeta - 1] - \frac{1}{4c}\mathbb{E}[(\zeta-1)^2] \right\}. \tag{27}$$

Consequently, representation (10) holds with[12] $\mathcal{A} = \{\zeta \in \mathcal{Z} : \zeta - 1 \succeq 0\}$ and

$$\rho^*(\zeta) = \tau \mathbb{E}[\zeta - 1] + \frac{1}{4c}\mathbb{E}[(\zeta-1)^2], \quad \zeta \in \mathcal{A}.$$

If $c > 0$, none of the conditions (A3) and (A4) is satisfied by this risk function.

Example 5 (Mean-upper-semideviation of order p from a target). Let $\mathcal{Z} := \mathcal{L}_p(\Omega, \mathcal{F}, P)$ and for $c \geq 0$ and $\tau \in \mathbb{R}$ consider

$$\rho(Z) := \mathbb{E}[Z] + c\left(\mathbb{E}\left[\big[Z-\tau\big]_+^p\right]\right)^{1/p}. \tag{28}$$

For any $c \geq 0$ and τ this risk function satisfies conditions (A1) and (A2), but not (A3) and (A4), if $c > 0$. We have

$$\begin{aligned} \left(\mathbb{E}\left[\big[Z-\tau\big]_+^p\right]\right)^{1/p} &= \sup_{\|\zeta\|_q \leq 1} \mathbb{E}\big(\zeta[Z-\tau]_+\big) \\ &= \sup_{\|\zeta\|_q \leq 1,\, \zeta(\cdot) \geq 0} \mathbb{E}\big(\zeta[Z-\tau]_+\big) \\ &= \sup_{\|\zeta\|_q \leq 1,\, \zeta(\cdot) \geq 0} \mathbb{E}\big(\zeta[Z-\tau]\big) \\ &= \sup_{\|\zeta\|_q \leq 1,\, \zeta(\cdot) \geq 0} \mathbb{E}\big[\zeta Z - \tau\zeta\big]. \end{aligned}$$

---

[12] Recall that $\mathcal{A} := \mathrm{dom}(\rho^*)$.

We obtain that representation (10) holds with $\mathcal{A} = \{\zeta \in \mathcal{Z}^* : \|\zeta\|_q \le c,\ \zeta \succeq 0\}$ and $\rho^*(\zeta) = \tau\mathbb{E}[\zeta]$ for $\zeta \in \mathcal{A}$.

Example 6. Let $v : \mathbb{R} \to \overline{\mathbb{R}}$ be a proper lower semicontinuous convex function. For $\mathcal{Z} := \mathcal{L}_p(\Omega, \mathcal{F}, P)$ consider the function

$$\rho(Z) := \inf_{\alpha \in \mathbb{R}} \mathbb{E}\big[Z + v(Z - \alpha)\big]. \tag{29}$$

Assume that functions $\psi_\alpha(z) := z + v(z - \alpha)$, $\alpha \in \mathbb{R}$, are bounded from below by a P-integrable function, and hence $\rho(Z) > -\infty$ for all $Z \in \mathcal{Z}$. Since the function $(Z, \alpha) \mapsto \mathbb{E}\big[Z + v(Z - \alpha)\big]$ is convex, it follows that $\rho(\cdot)$ is convex. Also $\rho(Z + a) = \rho(Z) + a$ for any $a \in \mathbb{R}$ and $Z \in \mathcal{Z}$. This can be shown by making the change of variables $z \mapsto z + a$ in the calculation of $\rho(Z + a)$. That is, ρ satisfies conditions (A1) and (A3).

Let us calculate the conjugate of ρ:

$$\begin{aligned}
\rho^*(\zeta) &= \sup_{Z \in \mathcal{Z}} \big\{\mathbb{E}[\zeta Z] - \rho(Z)\big\} \\
&= \sup_{Z \in \mathcal{Z},\, \alpha \in \mathbb{R}} \mathbb{E}\big[\zeta Z - Z - v(Z - \alpha)\big] \\
&= \sup_{Z \in \mathcal{Z},\, \alpha \in \mathbb{R}} \mathbb{E}\big[(Z + \alpha)\zeta - Z - \alpha - v(Z)\big] \\
&= \sup_{Z \in \mathcal{Z}} \big\{\mathbb{E}[\zeta Z - Z - v(Z)]\big\} + \sup_{\alpha \in \mathbb{R}} \big\{\alpha(\mathbb{E}[\zeta] - 1)\big\}.
\end{aligned} \tag{30}$$

By the interchangeability formula (17), the first term in (30) can be expressed as follows:

$$\sup_{Z \in \mathcal{Z}} \mathbb{E}\big[\zeta Z - Z - v(Z)\big] = \mathbb{E}\left[\sup_{z \in \mathbb{R}} \big\{z(\zeta - 1) - v(z)\big\}\right] = \mathbb{E}\left[v^*(\zeta - 1)\right],$$

where $v^*(\cdot)$ is the conjugate function of $v(\cdot)$. The supremum with respect to α in (30) is $+\infty$, unless $\mathbb{E}[\zeta] = 1$. We conclude that

$$\rho^*(\zeta) = \begin{cases} \mathbb{E}\left[v^*(\zeta - 1)\right], & \text{if } \mathbb{E}[\zeta] = 1, \\ +\infty, & \text{otherwise.} \end{cases} \tag{31}$$

The function ρ satisfies the monotonicity condition (A2) iff its domain contains only probability density functions. This is equivalent to the condition that $\mathbb{E}[v^*(\zeta - 1)] = +\infty$ for any such $\zeta \in \mathcal{Z}^*$ that the event "$\zeta(\omega) < 0$" happens with positive probability. In particular, ρ satisfies (A2) if $v^*(t) = +\infty$ for $t < -1$. This is the same as requiring that the function $\phi(z) := z + v(z)$ is monotonically nondecreasing on $\mathbb{R}$.

Example 7 (Conditional value at risk). For $\mathcal{Z} := \mathcal{L}_1(\Omega, \mathcal{F}, P)$, $\mathcal{Z}^* := \mathcal{L}_\infty(\Omega, \mathcal{F}, P)$ and constants $\varepsilon_1 \ge 0$ and $\varepsilon_2 \ge 0$ consider

$$\rho(Z) := \mathbb{E}[Z] + \inf_{\alpha \in \mathbb{R}} \mathbb{E}\big(\varepsilon_1[\alpha - Z]_+ + \varepsilon_2[Z - \alpha]_+\big). \tag{32}$$

Note that the above function ρ is of the form (29) with

$$v(z) := \varepsilon_1[-z]_+ + \varepsilon_2[z]_+. \tag{33}$$

We have here that the function $z+v(z)$ is positively homogeneous, and monotonically nondecreasing iff $\varepsilon_1 \le 1$. It follows that for any $\varepsilon_1 \in [0,1]$ and $\varepsilon_2 \ge 0$, the above function ρ is a coherent risk function satisfying conditions (A1)–(A4). Moreover,

$$v^*(t) = \begin{cases} 0, & \text{if } t \in [-\varepsilon_1, \varepsilon_2], \\ +\infty, & \text{otherwise.} \end{cases}$$

Consequently we have that, for any $\varepsilon_1 \ge 0$ and $\varepsilon_2 \ge 0$, representation (14) holds with

$$\mathcal{A} = \big\{\zeta \in \mathcal{Z}^* : 1 - \varepsilon_1 \le \zeta(\omega) \le 1 + \varepsilon_2, \text{ a.e. } \omega \in \Omega, \ \mathbb{E}[\zeta] = 1\big\}. \tag{34}$$

For $\varepsilon_1 > 0$ and $\varepsilon_2 > 0$ we can write ρ in the form

$$\rho(Z) = (1 - \varepsilon_1)\mathbb{E}[Z] + \varepsilon_1 \mathrm{CV@R}_\kappa[Z], \tag{35}$$

where $\kappa := \varepsilon_2/(\varepsilon_1 + \varepsilon_2)$ and

$$\mathrm{CV@R}_\kappa[Z] := \inf_{a\in\mathbb{R}} \left\{ a + \frac{1}{1-\kappa}\mathbb{E}\big([Z-a]_+\big) \right\} \tag{36}$$

is the so called Conditional Value at Risk function, [20]. By the above analysis we have that $\mathrm{CV@R}_\kappa[Z]$ is a coherent risk function for any $\kappa \in (0,1)$ and the corresponding set $\mathcal{A}$ is given by

$$\mathcal{A} = \big\{\zeta \in \mathcal{Z}^* : 0 \le \zeta(\omega) \le (1-\kappa)^{-1}, \text{ a.e. } \omega \in \Omega, \ \mathbb{E}[\zeta] = 1\big\}. \tag{37}$$

5 Stochastic Dominance Conditions

In all examples considered in Section 4, the space $\mathcal{Z}$ was given by $\mathcal{L}_p(\Omega, \mathcal{F}, P)$ with $\mathcal{Z}^* := \mathcal{L}_q(\Omega, \mathcal{F}, P)$ and, moreover, the risk functions $\rho(Z)$ discussed there were dependent only on the distribution of Z. That is, each risk function $\rho(Z)$, considered in Section 4, could be formulated in terms of the cumulative distribution function (cdf) $F_Z(z) := P(Z \le z)$ associated with $Z \in \mathcal{Z}$. In other words these risk functions satisfied the following condition:

(D) If $Z_1, Z_2 \in \mathcal{Z}$ are such that $P(Z_1 \le z) = P(Z_2 \le z)$ for all $z \in \mathbb{R}$, then $\rho(Z_1) = \rho(Z_2)$.

We say that risk function $\rho : \mathcal{Z} \to \overline{\mathbb{R}}$ is *distribution invariant* if it satisfies the above condition (D). For distribution invariant risk functions it makes sense to discuss their monotonicity properties with respect to various stochastic orders defined for (real valued) random variables.

Many stochastic orders can be characterized by a class $\mathcal{G}$ of functions $g : \mathbb{R} \to \mathbb{R}$ as follows. For (real valued) random variables Z_1 and Z_2 it is said that Z_2 dominates Z_1, denoted $Z_2 \succeq_{\mathcal{G}} Z_1$, if $\mathbb{E}[g(Z_2)] \geq \mathbb{E}[g(Z_1)]$ for all $g \in \mathcal{G}$ for which the corresponding expectations do exist. This stochastic order is called the *integral stochastic order* with *generator* $\mathcal{G}$. We refer to [10, Chapter 2] for a thorough discussion of this concept. For example, the *usual stochastic order*, written $Z_2 \succeq_{\rm st} Z_1$, corresponds to the generator $\mathcal{G}$ formed by all non-decreasing functions $g : \mathbb{R} \to \mathbb{R}$. It is possible to show that $Z_2 \succeq_{\rm st} Z_1$ iff $F_{Z_2}(z) \leq F_{Z_1}(z)$ for all $z \in \mathbb{R}$ (e.g., [10, Theorem 1.2.8]). We say that the integral stochastic order is *increasing* if all functions in the set $\mathcal{G}$ are non-decreasing. The usual stochastic order is an example of increasing integral stochastic order.

We say that (distribution invariant) risk function ρ is *consistent* with the integral stochastic order if $Z_2 \succeq_{\mathcal{G}} Z_1$ implies $\rho(Z_2) \geq \rho(Z_1)$ for all $Z_1, Z_2 \in \mathcal{Z}$, i.e., ρ is monotone with respect to $\succeq_{\mathcal{G}}$. For an increasing integral stochastic order we have that if $Z_2(\omega) \geq Z_1(\omega)$ for a.e. $\omega \in \Omega$, then $g(Z_2(\omega)) \geq g(Z_1(\omega))$ for any $g \in \mathcal{G}$ and a.e. $\omega \in \Omega$, and hence $\mathbb{E}[g(Z_2(\omega))] \geq \mathbb{E}[g(Z_1(\omega))]$. That is, if $Z_2 \succeq Z_1$ in the almost sure sense, then $Z_2 \succeq_{\mathcal{G}} Z_1$. It follows that if ρ is distribution invariant and consistent with respect to an increasing integral stochastic order, then it satisfies the monotonicity condition (A2). In other words if ρ does not satisfy condition (A2), then it cannot be consistent with any increasing integral stochastic order. In particular, for $c > 0$ the mean-variance risk function, defined in (19), is not consistent with any increasing integral stochastic order, and for $p > 1$ the mean-deviation risk function, defined in (21), is not consistent with any increasing integral stochastic order provided that condition (23) holds.

Consider now the usual stochastic order. By Strassen's localization theorem, we have that $Z_2 \succeq_{\rm st} Z_1$ iff there exists a probability space $(\Omega, \mathcal{F}, P)$ and random variables $\hat{Z}_1$ and $\hat{Z}_2$ on it such that[13] $\hat{Z}_1 \overset{D}{\sim} Z_1$ and $\hat{Z}_2 \overset{D}{\sim} Z_2$, and $\hat{Z}_2(\omega) \geq \hat{Z}_1(\omega)$ for all $\omega \in \Omega$ (e.g., [10, Theorem 1.2.4]). In our context, this relation between the usual stochastic order and the almost sure order cannot be used directly, because, we are not allowed to freely change the probability space $(\Omega, \mathcal{F}, P)$, which is an integral part of our definition of a risk function.

However, if our space $(\Omega, \mathcal{F}, P)$ is sufficiently rich, so that a *uniform*[14] random variable $U(\omega)$ exists on this space, we can easily link the monotonicity assumption (A2) with the consistency with the usual stochastic order. Suppose that the risk function ρ is distribution invariant and satisfies the monotonicity condition (A2). Recall that $Z_2 \succeq_{\rm st} Z_1$ iff $F_{Z_2}(z) \leq F_{Z_1}(z)$ for all $z \in \mathbb{R}$. Consider random variables $\hat{Z}_1 := F_{Z_1}^{-1}(U)$ and $\hat{Z}_2 := F_{Z_2}^{-1}(U)$, where the inverse distribution function is defined as

[13] The notation $X \overset{D}{\sim} Y$ means that random variables X and Y, which can be defined on different probability spaces, have the same cumulative distribution function.

[14] Random variable $U : \Omega \to [0, 1]$ is said to be uniform if $P(U \leq z) = z$ for every $z \in [0, 1]$.

$$F_Z^{-1}(t) := \inf\{z : F_Z(z) \geq t\}.$$

We obtain that $\hat{Z}_2(\omega) \geq \hat{Z}_1(\omega)$ for all $\omega \in \Omega$, and by virtue of (A2), $\rho(\hat{Z}_2) \geq \rho(\hat{Z}_1)$. By construction, $\hat{Z}_1$ has the same distribution as Z_1, and $\hat{Z}_2$ has the same distribution as Z_2. Since the risk function is distribution invariant, we conclude that $\rho(Z_2) \geq \rho(Z_1)$. Consequently, the risk function ρ is consistent with the usual stochastic order. It follows that in a sufficiently rich probability space the monotonicity condition (A2) and the consistency with the usual stochastic order are equivalent (for distribution invariant risk functions).

It is said that Z_2 is bigger than Z_1 in *increasing convex order*, written $Z_2 \succeq_{\text{icx}} Z_1$, if $\mathbb{E}[g(Z_2)] \geq \mathbb{E}[g(Z_1)]$ for all increasing convex functions $g : \mathbb{R} \to \mathbb{R}$ such that the expectations exist. Clearly this is an integral stochastic order with the corresponding generator given by the set of increasing convex functions. It is the counterpart of the classical stochastic dominance relation, which is the increasing concave order (recall that we are dealing here with minimization rather than maximization procedures). Consider the setting of Example 6 with risk function ρ defined in (29). Suppose that the function $\phi(z) := z + v(z)$ is monotonically nondecreasing on $\mathbb{R}$. Note that $\phi(\cdot)$ is convex, since $v(\cdot)$ is convex. We obtain that if $Z_2 \geq_{\text{icx}} Z_1$, then $\mathbb{E}[\phi(Z_2 - \alpha)] \geq \mathbb{E}[\phi(Z_2 - \alpha)]$ for any fixed $\alpha \in \mathbb{R}$, and hence (by taking minimum over $\alpha \in \mathbb{R}$) that $\rho(Z_2) \geq \rho(Z_1)$. That is, the risk function defined in (29) is consistent with the increasing convex order. We have in this way re-established the stochastic dominance consistency result of [13].

The mean-upper-semideviation risk function of order $p \geq 1$ (Example 3) is also consistent with the increasing convex order, provided that $c \in [0, 1]$. We can prove this for $p = 1$ as follows (see [11]).

Suppose that $Z_2 \succeq_{\text{icx}} Z_1$. First, using $g(z) := z$ we see that

$$\mathbb{E}[Z_1] \leq \mathbb{E}[Z_2]. \tag{38}$$

Secondly, setting $g(z) := \big(z - \mathbb{E}[Z_1]\big)_+$ we obtain that

$$\mathbb{E}\Big[\big(Z_1 - \mathbb{E}[Z_1]\big)_+\Big] \leq \mathbb{E}\Big[\big(Z_2 - \mathbb{E}[Z_1]\big)_+\Big].$$

Using (38) we can continue this estimate as follows

$$\begin{aligned}\mathbb{E}\Big[\big(Z_1 - \mathbb{E}[Z_1]\big)_+\Big] &\leq \mathbb{E}\Big[\big(Z_2 - \mathbb{E}[Z_2] + \mathbb{E}[Z_2] - \mathbb{E}[Z_1]\big)_+\Big] \\ &\leq \mathbb{E}\Big[\big(Z_2 - \mathbb{E}[Z_2]\big)_+\Big] + \mathbb{E}[Z_2] - \mathbb{E}[Z_1].\end{aligned}$$

This can be rewritten as

$$\mathbb{E}[Z_1] + \mathbb{E}\Big[\big(Z_1 - \mathbb{E}[Z_1]\big)_+\Big] \leq \mathbb{E}[Z_2] + \mathbb{E}\Big[\big(Z_2 - \mathbb{E}[Z_2]\big)_+\Big], \tag{39}$$

which is the required relation $\rho(Z_1) \leq \rho(Z_2)$ for $c = 1$. Combining inequalities (38) and (39) with coefficients $1 - c$ and c, we obtain the required result for any $c \in [0, 1]$. The proof for $p > 1$ can be found in [12] (in the stochastic dominance setting).

6 Differentiability of Risk Functions

In this section we discuss differentiability properties of risk functions. In the analysis of optimization of risk measures we also have to deal with composite functions of the form

$$\phi(x) := \rho(F(x)).$$

Here $F : \mathbb{R}^n \to \mathcal{Z}$ is a mapping, defined by $[F(x)](\cdot) := f(x, \cdot)$, associated with a function $f : \mathbb{R}^n \times \Omega \to \mathbb{R}$. Of course, in order for this mapping F to be well defined we have to assume that the random variable $Z(\omega) = f(x, \omega)$ belongs to $\mathcal{Z}$ for any $x \in \mathbb{R}^n$. We say that the mapping F is *convex* if the function $f_\omega(\cdot) := f(\cdot, \omega)$ is convex for every $\omega \in \Omega$. It is not difficult to verify and is well known that the composite function $\phi(x)$ is convex if F is convex and ρ is convex and satisfies the monotonicity condition (A2). Let us emphasize that in order to preserve convexity of the composite function ϕ we need convexity of F and ρ *and* the monotonicity property (A2).

Consider a point $\bar{Z} \in \mathcal{Z}$ such that $\rho(\bar{Z})$ is finite valued. Since it is assumed that ρ is proper, this means that $\bar{Z} \in \mathrm{dom}(\rho)$. The following limit (provided that it exists)

$$\rho'(\bar{Z}, Z) := \lim_{t \downarrow 0} \frac{\rho(\bar{Z} + tZ) - \rho(\bar{Z})}{t} \tag{40}$$

is called the *directional derivative* of ρ at $\bar{Z}$ in direction Z. If this limit exists for all $Z \in \mathcal{Z}$, it is said that ρ is directionally differentiable at $\bar{Z}$. It is said that ρ is Hadamard directionally differentiable at $\bar{Z}$, if ρ is directionally differentiable at $\bar{Z}$ and, moreover, the following limit holds

$$\rho'(\bar{Z}, Z) = \lim_{\substack{Z' \to Z \\ t \downarrow 0}} \frac{\rho(\bar{Z} + tZ') - \rho(\bar{Z})}{t}. \tag{41}$$

It can be observed that $\rho'(\bar{Z}, Z)$ is just the one sided derivative of the function $g(t) := \rho(\bar{Z} + tZ)$ at $t = 0$. If ρ is convex, then the function $g : \mathbb{R} \to \overline{\mathbb{R}}$ is also convex, and hence $\rho'(\bar{Z}, Z)$ exists, although it can take values $+\infty$ or $-\infty$.

It said that an element $\mu \in \mathcal{Z}^*$ is a *subgradient* of ρ at $\bar{Z}$ if

$$\rho(Z) \geq \rho(\bar{Z}) + \langle \mu, Z - \bar{Z} \rangle, \quad \forall\, Z \in \mathcal{Z}. \tag{42}$$

The set of all subgradients of ρ, at $\bar{Z}$, is called the *subdifferential* of ρ and denoted $\partial\rho(\bar{Z})$. It is said that ρ is subdifferentiable at $\bar{Z}$ if $\partial\rho(\bar{Z})$ is nonempty. By convex analysis we have that if ρ is convex and continuous at $\bar{Z}$, then it is subdifferentiable at $\bar{Z}$, and, moreover, if $\mathcal{Z}$ is a Banach space (equipped with its norm topology), then ρ is Hadamard directionally differentiable at $\bar{Z}$.

It is said that ρ is Gâteaux (Hadamard) differentiable at $\bar{Z}$ if it is (Hadamard) directionally differentiable at $\bar{Z}$ and there exists $\bar{\mu} \in \mathcal{Z}^*$ such that $\rho'(\bar{Z}, Z) = \langle \bar{\mu}, Z \rangle$ for all $Z \in \mathcal{Z}$. The functional $\bar{\mu}$ represents the derivative of ρ at $\bar{Z}$ and denoted $\nabla\rho(\bar{Z})$. If the space $\mathcal{Z}$ is finite dimensional, then the concept of Hadamard differentiability coincides with the usual concept of differentiability. By convex analysis we have the following.

Theorem 3. *Suppose that $\mathcal{Z}$ is a Banach space (e.g., $\mathcal{Z} := \mathcal{L}_p(\Omega, \mathcal{F}, P)$), and ρ is convex and finite valued and continuous at $\bar{Z}$. Then ρ is subdifferentiable and Hadamard directionally differentiable at $\bar{Z}$, and the following formulas hold*

$$\partial\rho(\bar{Z}) = \operatorname{argmax}_{\mu \in \mathcal{Z}^*} \left\{ \langle \mu, \bar{Z} \rangle - \rho^*(\mu) \right\}, \tag{43}$$

$$\rho'(\bar{Z}, Z) = \sup_{\mu \in \partial\rho(\bar{Z})} \langle \mu, Z \rangle. \tag{44}$$

Moreover, ρ is Hadamard differentiable at $\bar{Z}$ if and only if $\partial\rho(\bar{Z}) = \{\bar{\mu}\}$ is a singleton, in which case $\nabla\rho(\bar{Z}) = \bar{\mu}$.

As we mentioned earlier, if $\mathcal{Z} := \mathcal{L}_p(\Omega, \mathcal{F}, P)$ and ρ satisfies conditions (A1) and (A2), then ρ is continuous and subdifferentiable at every point of the interior of its domain ([23]). In particular, if ρ is real valued, then ρ is continuous and subdifferentiable at every point of $\mathcal{Z}$ and formulas (43) and (44) hold. Moreover, if ρ is a real valued coherent risk function, then representation (14) holds and

$$\partial\rho(\bar{Z}) = \underset{\zeta \in \mathcal{A}}{\operatorname{argmax}} \langle \zeta, \bar{Z} \rangle. \tag{45}$$

Consider now the composite function $\phi(x) := \rho(F(x))$. Since $f_\omega(\cdot)$ is real valued, we have that if $f_\omega(\cdot)$ is convex, then it is directionally differentiable at every point $\bar{x} \in \mathbb{R}^n$ and its directional derivative $f'_\omega(\bar{x}, x)$ is finite valued. By using the chain rule for directional derivatives and (44) we obtain the following differentiability properties of the composite function, at a point $\bar{x} \in \mathbb{R}^n$ (cf., [23]).

Proposition 1. *Suppose that $\mathcal{Z}$ is a Banach space, the mapping $F : \mathbb{R}^n \to \mathcal{Z}$ is convex, the function ρ is convex, finite valued and continuous at $\bar{Z} := F(\bar{x})$. Then the composite function $\phi(x) = \rho(F(x))$ is directionally differentiable at $\bar{x}$, its directional derivative $\phi'(\bar{x}, x)$ is finite valued for every $x \in \mathbb{R}^n$ and*

$$\phi'(\bar{x}, x) = \sup_{\mu \in \partial\rho(\bar{Z})} \int_\Omega f'_\omega(\bar{x}, x)\, d\mu(\omega). \tag{46}$$

Moreover, if $\partial\rho(\bar{Z}) = \{\bar{\mu}\}$ is a singleton, then the composite function ϕ is differentiable at $\bar{x}$ if and only if $f_\omega(\cdot)$ is differentiable at $\bar{x}$ for $\bar{\mu}$-almost every ω, in which case

$$\nabla\phi(\bar{x}) = \int_\Omega \nabla f_\omega(\bar{x})\, d\bar{\mu}(\omega). \tag{47}$$

It is also possible to write the above differentiability formulas in terms of subdifferentials. Suppose that F is convex. Then for any[15] measure $\mu \in \mathcal{Z}^*_+$ the integral function $\psi_\mu(x) := \int_\Omega f_\omega(x)\, d\mu(\omega)$ is also convex. Moreover,

[15] Recall that $\mathcal{Z}^*_+$ denotes the set of nonnegative measures $\mu \in \mathcal{Z}^*$.

if the integral function $\psi_\mu(\cdot)$ is finite valued (and hence continuous) in a neighborhood of a point $\bar{x} \in \mathbb{R}^n$, then

$$\psi'_\mu(\bar{x}, x) = \int_\Omega f'_\omega(\bar{x}, x)\, d\mu(\omega), \tag{48}$$

and by Strassen's disintegration theorem the following interchangeability formula holds

$$\partial\psi_\mu(\bar{x}) = \int_\Omega \partial f_\omega(\bar{x})\, d\mu(\omega). \tag{49}$$

The integral in the right hand side of (49) is understood as the set of all vectors of the form $\int_\Omega \delta(\omega)\, d\mu(\omega)$, where $\delta(\omega)$ is a μ-integrable selection[16] of $\partial f_\omega(\bar{x})$.

Suppose that the assumptions of Proposition 1 hold, and monotonicity condition (A2) is satisfied and hence ϕ is convex and $\partial\rho(\bar{Z}) \subset \mathcal{Z}^*_+$. Now formula (46) means that $\phi'(\bar{x}, \cdot)$ is equal to the supremum of $\psi'_\mu(\bar{x}, \cdot)$ over $\mu \in \partial\rho(\bar{Z})$. The functions $\psi'_\mu(\bar{x}, \cdot)$ are convex and positively homogeneous, and hence $\partial\phi(\bar{x})$ is equal to the topological closure of the union of the sets $\partial\psi'_\mu(\bar{x})$ over $\mu \in \partial\rho(\bar{Z})$. Consequently, we obtain that formula (46) can be written in the following equivalent form[17]

$$\partial\phi(\bar{x}) = \mathrm{cl}\left\{ \textstyle\bigcup_{\mu \in \partial\rho(\bar{Z})} \int_\Omega \partial f_\omega(\bar{x})\, d\mu(\omega) \right\}. \tag{50}$$

Note that since $\partial\rho(\bar{Z})$ is convex, it is straightforward to verify that the set inside the parentheses at the right hand side of (50) is convex. Moreover, if the maximum in the right hand side of (46) is attained for any $x \in \mathbb{R}^n$, then this set is closed. Now if $\mathcal{Z} := \mathcal{L}_p(\Omega, \mathcal{F}, P)$, with $p \in [1, +\infty)$, then $\partial\rho(\bar{Z})$ is weakly* compact, provided that ρ is continuous at $\bar{Z}$. In that case, indeed, the maximum in the right hand side of (46) is always attained. We obtain the following result.

Theorem 4. *Suppose that $\mathcal{Z} := \mathcal{L}_p(\Omega, \mathcal{F}, P)$, with $p \in [1, +\infty)$, the mapping $F : \mathbb{R}^n \to \mathcal{Z}$ is convex, the function ρ satisfies conditions* (A1) *and* (A2)*, finite valued and continuous at $\bar{Z} := F(\bar{x})$. Then*

$$\partial\phi(\bar{x}) = \bigcup_{\zeta \in \partial\rho(\bar{Z})} \int_\Omega \zeta(\omega) \partial f_\omega(\bar{x})\, dP(\omega). \tag{51}$$

Let us consider now some examples discussed in Section 4.

Example 8 (Mean-upper-semideviation risk function of order p). Consider the setting of Example 3. We have that the risk function ρ, defined in (24), is a convex real valued continuous function. It follows that for any $Z \in \mathcal{Z}$ the

[16] It is said that $\delta(\omega)$ is a selection of $\partial f_\omega(\bar{x})$ if $\delta(\omega) \in \partial f_\omega(\bar{x})$ for almost every ω.

[17] By $\mathrm{cl}(S)$ we denote the topological closure of the set $S \subset \mathbb{R}^n$.

subdifferential $\partial\rho(Z)$ is nonempty and formula (45) holds with the set $\mathcal{A}$ given in (25). That is,

$$\partial\rho(Z) = \big\{1 + \zeta - \mathbb{E}[\zeta] : \zeta \in \Delta_Z\big\}, \tag{52}$$

where

$$\Delta_Z := \operatorname*{argmax}_{\zeta\in\mathcal{Z}^*} \big\{\langle\zeta, Y\rangle : \|\zeta\|_q \le c,\ \zeta \succeq 0\big\} \ \text{ and } \ Y := Z - \mathbb{E}[Z]. \tag{53}$$

If $p \in (1, +\infty)$, then the set Δ_Z can be described as follows. If the function $Z(\cdot)$ is constant, then $Y(\cdot) \equiv 0$ and hence $\Delta_Z = \{\zeta : \|\zeta\|_q \le c,\ \zeta \succeq 0\}$. Suppose that $Z(\cdot)$ is not constant[18] and hence $Y(\cdot)$ is not identically zero. Note that the 'argmax' in (53) is not changed if Y is replaced $Y_+(\cdot) := [Y(\cdot)]_+$. With Y_+ is associated a unique point $\zeta^* \in \mathcal{Z}^*$ such that $\|\zeta^*\|_q = 1$ and $\langle\zeta^*, Y\rangle = \|Y\|_p$. Since $Y_+ \succeq 0$, it follows that $\zeta^* \succeq 0$ and $\Delta_Z = \{c\,\zeta^*\}$. That is, for $p > 1$ and nonconstant $Z \in \mathcal{Z}$, the subdifferential $\partial\rho(Z)$ is a singleton, and hence ρ is differentiable at Z.

Suppose now that $p = 1$ and hence $q = +\infty$. In that case

$$\Delta_Z = \left\{\zeta \in \mathcal{Z}^* : \begin{array}{l} \zeta(\omega) = c \text{ if } Y(\omega) > 0,\ \zeta(\omega) = 0 \text{ if } Y(\omega) < 0, \\ 0 \le \zeta(\omega) \le c \text{ if } Y(\omega) = 0 \end{array}\right\}. \tag{54}$$

It follows that Δ_Z is a singleton, and hence ρ is differentiable at Z, iff $Y(\omega) \ne 0$ for P-almost every $\omega \in \Omega$.

Example 9 (Mean-upper-semideviation of order p from a target). Consider the setting of Example 5. The risk function ρ, defined in (28), is real valued convex and continuous. We have that

$$\partial\rho(Z) = \operatorname*{argmax}_{\zeta\in\mathcal{Z}^*} \big\{\langle\zeta, Z - \tau\rangle : \|\zeta\|_q \le c,\ \zeta \succeq 0\big\}. \tag{55}$$

Similarly to the previous example, we have here that if $p > 1$, then ρ is differentiable at Z iff $P\{Z(\omega) \ne \tau\} > 0$. If $p = 1$, then ρ is differentiable at Z iff $P\{Z(\omega) \ne \tau\} = 1$.

Example 10. Consider the setting of Example 6 with the risk function ρ defined in (29). Because of (31) and by (43) we have

$$\partial\rho(Z) = \operatorname{argmax}_{\zeta\in\mathcal{Z}^*,\,\mathbb{E}[\zeta]=1} \mathbb{E}\big[\zeta Z - v^*(\zeta - 1)\big]. \tag{56}$$

Also the subdifferential of function $h(\zeta) := \mathbb{E}\big[\zeta Z - v^*(\zeta - 1)\big]$ is given by

$$\partial h(\zeta) = \big\{Z' \in \mathcal{Z} : Z'(\omega) \in Z(\omega) - \partial v^*(\zeta(\omega) - 1),\ \omega \in \Omega\big\}.$$

By the first order optimality conditions we have then that $\bar\zeta \in \mathcal{Z}^*$ is an optimal solution of the right hand side problem of (56) iff there exists $\bar\lambda \in \mathbb{R}$ such that

[18] Of course, this and similar statements here should be understood up to a set of P-measure zero.

$$Z(\omega) - \bar{\lambda} \in \partial v^*(\bar{\zeta}(\omega) - 1), \ a.e. \ \omega \in \Omega, \ \text{ and } \ \mathbb{E}[\bar{\zeta}] = 1. \tag{57}$$

Since the inclusion $a \in \partial v^*(z)$ is equivalent to $z \in \partial v(a)$ we obtain

$$\partial\rho(Z) = \big\{\zeta \in \mathcal{Z}^* : \zeta(\omega) \in 1 + \partial v(Z(\omega) - \bar{\lambda}), \ a.e. \ \omega \in \Omega, \ \mathbb{E}[\zeta] = 1\big\}. \tag{58}$$

Note that $\bar{\lambda}$ is an optimal solution of the dual problem

$$\operatorname*{Min}_{\lambda\in\mathbb{R}} \sup_{\zeta\in\mathcal{Z}^*} \mathbb{E}\big[\zeta Z - v^*(\zeta - 1) - \lambda(\zeta - 1)\big].$$

By interchanging the integral and max operators (see (17)), the above problem can be written in the following equivalent form

$$\mathrm{Min}_{\lambda\in\mathbb{R}}\, \mathbb{E}\left[\sup_{z\in\mathbb{R}} \big\{(Z - \lambda)z - v^*(z - 1) + \lambda\big\}\right].$$

Example 11 (Conditional value at risk). Consider the setting of Example 7 with ρ defined in (32). We can use results of the previous example with function $v(z)$ defined in (33). We have here that $\bar{\lambda}$ is an optimal solution of the problem

$$\mathrm{Min}_{\lambda\in\mathbb{R}}\, \mathbb{E}\big[- \varepsilon_1[\lambda - Z]_+ + \varepsilon_2[Z - \lambda]_+\big]. \tag{59}$$

For $\varepsilon_1 > 0$ and $\varepsilon_2 > 0$ an optimal solution $\bar{\lambda}$ of (59) is given by a κ-quantile of Z (recall that $\kappa = \varepsilon_2/(\varepsilon_1 + \varepsilon_2)$). That is, $\bar{\lambda} \in [a, b]$ where

$$a := \inf\{t : P(Z \le t) \ge \kappa\} \ \text{ and } \ b := \sup\{t : P(Z \le t) \ge \kappa\}.$$

By (58) we have

$$\partial\rho(Z) = \left\{ \zeta \in \mathcal{Z}^* : \begin{array}{ll} \zeta(\omega) = 1 - \varepsilon_1, & \text{if } Z(\omega) < \bar{\lambda} \\ \zeta(\omega) = 1 + \varepsilon_2, & \text{if } Z(\omega) > \bar{\lambda} \\ \zeta(\omega) \in [1 - \varepsilon_1, 1 + \varepsilon_2], & \text{if } Z(\omega) = \bar{\lambda} \\ \mathbb{E}[\zeta] = 1 & \end{array} \right\}. \tag{60}$$

Note that elements (functions) $\zeta \in \partial\rho(Z)$ are defined up to sets of P-measure zero and the above formula (60) holds for any κ-quantile $\bar{\lambda} \in [a, b]$. Also recall that for $\varepsilon_1 = 1$ the risk function $\rho(\cdot)$ coincides with $CV@R_\kappa[\,\cdot\,]$.

7 Optimization of Risk Functions

In this section we consider the optimization problem

$$\operatorname*{Min}_{x\in S} \big\{\phi(x) := \rho(F(x))\big\}. \tag{61}$$

Recall that with mapping $F : \mathbb{R}^n \to \mathcal{Z}$ is associated function $f(x, \omega) = [F(x)](\omega)$. We assume throughout this section, and the following sections 8 and 9, that:

(i) S is a nonempty closed convex subset of $\mathbb{R}^n$,
(ii) the mapping $F : \mathbb{R}^n \to \mathcal{Z}$ is convex,
(iii) the risk function $\rho : \mathcal{Z} \to \overline{\mathbb{R}}$ is proper, lower semicontinuous and satisfies conditions (A1) and (A2).

It follows that the composite function $\phi : \mathbb{R}^n \to \overline{\mathbb{R}}$ is convex, and hence optimization problem (61) is a convex problem. Because of the Fenchel-Moreau theorem, we can employ representation (10) of the risk function ρ to write problem (61) in the following min-max form

$$\operatorname*{Min}_{x\in S} \sup_{\mu\in\mathcal{A}} \Phi(x,\mu), \tag{62}$$

where $\mathcal{A} := \mathrm{dom}(\rho^*)$ and

$$\Phi(x,\mu) := \langle \mu, F(x)\rangle - \rho^*(\mu). \tag{63}$$

Note that because of the assumed monotonicity condition (A2), the set $\mathcal{A}$ contains only nonnegative measures, i.e., $\mathcal{A} \subset \mathcal{Z}^*_+$. If, moreover, assumption (A3) holds, then $\mathcal{A}$ is a subset of the set $\mathcal{P} \subset \mathcal{Z}^*$ of probability measures, and for $\mu \in \mathcal{P}$,

$$\langle \mu, F(x)\rangle = \mathbb{E}_\mu[F(x)] = \textstyle\int_\Omega f(x,\omega)\, d\mu(\omega).$$

If assumption (A4) also holds, then $\rho^*(\mu) = 0$ and hence $\Phi(x,\mu) = \mathbb{E}_\mu[F(x)]$ for any $\mu \in \mathcal{A}$. Therefore if ρ is a coherent risk function, then problem (61) can be written in the min-max form

$$\operatorname*{Min}_{x\in S} \sup_{\mu\in\mathcal{A}} \mathbb{E}_\mu[F(x)]. \tag{64}$$

We have here that the function $\Phi(x,\mu)$ is concave in μ and, since F is convex, is convex in x. Therefore, under various regularity conditions, the 'min' and 'max' operators in (62) can be interchanged to obtain the problem

$$\operatorname*{Max}_{\mu\in\mathcal{A}} \inf_{x\in S} \Phi(x,\mu). \tag{65}$$

For example, the following holds (cf., [23]).

Proposition 2. *Suppose that $\mathcal{Z}$ is a Banach space, the mapping F is convex, the function ρ is proper, lower semicontinuous and satisfies assumptions* (A1)–(A3). *Then the optimal values of problems* (62) *and* (65) *are equal to each other, and if their common optimal value is finite, then problem* (65) *has an optimal solution $\bar{\mu}$. Moreover, the optimal values of* (62) *and* (65) *are equal to the optimal value of the problem*

$$\operatorname*{Min}_{x\in S} \Phi(x,\bar{\mu}), \tag{66}$$

and if $\bar{x}$ is an optimal solution of (62)*, then $\bar{x}$ is also an optimal solution of* (66)*.*

We obtain that, under assumptions specified in the above proposition, there exists a probability measure $\bar{\mu} \in \mathcal{P}$ such that problem (61) is 'almost' equivalent to problem (66). That is, optimal values of problems (61) and (66) are equal to each other and the set of optimal solutions of problem (61) is contained in the set of optimal solutions of problem (66). Of course, the corresponding probability measure $\bar{\mu}$ is not known apriori and could be obtained by solving the dual problem (65).

We also have that if the optimal values of problems (62) and (65) are equal to each other, then $\bar{x}$ is an optimal solution of (62) and $\bar{\mu}$ is an optimal solution of (65) iff $(\bar{x}, \bar{\mu})$ is a *saddle point* of $\Phi(x, \mu)$, i.e.,

$$\bar{x} \in \operatorname*{argmin}_{x \in S} \Phi(x, \bar{\mu}) \ \text{ and } \ \bar{\mu} \in \operatorname*{argmax}_{\mu \in \mathcal{A}} \Phi(\bar{x}, \mu).$$

Conversely, if $\Phi(x, \mu)$ possesses a saddle point, then the optimal values of problems (62) and (65) are equal. Because of convexity and lower semicontinuity of ρ we have that $\rho^{**}(\cdot) = \rho(\cdot)$, and by (63) we obtain that

$$\operatorname*{argmax}_{\mu \in \mathcal{A}} \Phi(\bar{x}, \mu) = \partial \rho(\bar{Z}),$$

where $\bar{Z} := F(\bar{x})$. Moreover, if $\psi(\cdot) := \mathbb{E}_{\bar{\mu}}[F(\cdot)]$ is finite valued in a neighborhood of $\bar{x}$, then the first order optimality condition for $\bar{x}$ to be a minimizer of $\psi(x)$ over $x \in S$ is that[19] $0 \in N_S(\bar{x}) + \partial \psi(\bar{x})$. Together with Strassen's disintegration theorem (see (49)) this leads to the following optimality conditions.

Proposition 3. *Suppose that $\mathcal{Z}$ is a Banach space, the risk function ρ satisfies conditions* (A1)–(A3)*, the set S and the mapping F are convex, and $\bar{x} \in X$ and $\bar{\mu} \in \mathcal{P}$ are such that $\mathbb{E}_{\bar{\mu}}[F(\cdot)]$ is finite valued in a neighborhood of $\bar{x}$. Denote $\bar{Z} := F(\bar{x})$. Then $(\bar{x}, \bar{\mu})$ is a saddle point of $\Phi(x, \mu)$ if and only if:*

$$0 \in N_S(\bar{x}) + \mathbb{E}_{\bar{\mu}}[\partial f_\omega(\bar{x})] \ \text{ and } \ \bar{\mu} \in \partial \rho(\bar{Z}). \tag{67}$$

Under the assumptions of Proposition 3, conditions (67) can be viewed as optimality conditions for a point $\bar{x} \in S$ to be an optimal solution of problem (61). That is, if there exists a probability measure $\bar{\mu} \in \partial \rho(\bar{Z})$ such that the first condition of (67) holds, then $\bar{x}$ is an optimal solution of problem (61), i.e., (67) are sufficient conditions for optimality. Moreover, under the assumptions of Proposition 2, the existence of such a probability measure $\bar{\mu}$ is a necessary condition for optimality of $\bar{x}$.

8 Nonanticipativity Constraints

The optimization problem (61) can be written in the following equivalent form

[19] By $N_S(\bar{x}) := \left\{x \in \mathbb{R}^n : (x - \bar{x})^T y \le 0,\ \forall\, y \in S\right\}$ we denote the normal cone to S at $\bar{x} \in S$. By the definition $N_S(\bar{x}) = \emptyset$ if $\bar{x} \notin S$.

$$\operatorname*{Min}_{X \in \mathcal{M}_S,\, x \in \mathbb{R}^n} \rho(F_X) \;\text{ subject to } X(\omega) = x,\ \forall\, \omega \in \Omega, \tag{68}$$

where $\mathcal{M} := \mathcal{L}_p(\Omega, \mathcal{F}, P, \mathbb{R}^n)$ and $[F_X](\omega) := f(X(\omega), \omega)$, for $X \in \mathcal{M}$, and

$$\mathcal{M}_S := \{X \in \mathcal{M} : X(\omega) \in S,\ a.e.\ \omega \in \Omega\}.$$

Although the above problem involves optimization over the functional space $\mathcal{M}$, the constraints $X(\omega) = x$, $\omega \in \Omega$, ensure that this problem is equivalent to problem (61). These constraints are called the *nonanticipativity* constraints.

Ignoring the nonanticipativity constraints we can write the following relaxation of problem (68):

$$\operatorname*{Min}_{X \in \mathcal{M}_S} \rho(F_X). \tag{69}$$

Let us note now that the interchangeability principle, similar to (17), holds for risk functions as well.

Proposition 4. *Let $\mathcal{Z} := \mathcal{L}_p(\Omega, \mathcal{F}, P)$, $\rho : \mathcal{Z} \to \mathbb{R}$ be a real valued risk function satisfying conditions* (A1) *and* (A2), *$f : \mathbb{R}^n \times \Omega \to \mathbb{R}$ be a random lower semicontinuous function and $\mathcal{G} : \Omega \rightrightarrows \mathbb{R}^n$ be a closed valued measurable multifunction*[20]. *Let $F^{\mathcal{G}}(\omega) := \inf_{x \in \mathcal{G}(\omega)} f(x, \omega)$ and suppose that $F^{\mathcal{G}} \in \mathcal{Z}$. Then*

$$\rho(F^{\mathcal{G}}) = \inf_{X \in \mathcal{M}} \left\{ \rho\big(F_X\big) : X(\omega) \in \mathcal{G}(\omega)\ a.e.\ \omega \in \Omega \right\}. \tag{70}$$

The above interchangeability formula can be either derived from (17) by using the dual representation (10) or proved directly. Indeed, for any $X \in \mathcal{M}$ such that $X(\cdot) \in \mathcal{G}(\cdot)$ we have that $F^{\mathcal{G}}(\cdot) \le f(X(\cdot), \cdot)$, and hence it follows by assumption (A2) that $\rho(F^{\mathcal{G}}) \le \rho(F_X)$. This implies that $\rho(F^{\mathcal{G}})$ is less than or equal to the right hand side of (70). Conversely, suppose for the moment that the minimum of $f(x, \omega)$ over $x \in \mathcal{G}(\omega)$ is attained for a.e. $\omega \in \Omega$, and let $\bar{X}(\cdot) \in \arg\min_{x \in \mathcal{G}(\cdot)} f(x, \cdot)$ be a measurable selection such that $\bar{X} \in \mathcal{M}$. Then $\rho(F^{\mathcal{G}}) = \rho(F_{\bar{X}})$, and hence $\rho(F^{\mathcal{G}})$ is greater than or equal to the right hand side of (70). It also follows then that

$$\bar{X} \in \operatorname*{argmin}_{X \in \mathcal{M}} \left\{ \rho\big(F_X\big) : X(\omega) \in \mathcal{G}(\omega)\ a.e.\ \omega \in \Omega \right\}. \tag{71}$$

Such arguments can be also pushed through without assuming existence of optimal solutions by considering ε-optimal solutions with arbitrary $\varepsilon > 0$. Let us emphasize that the monotonicity assumption (A2) is the key condition for (70) to hold.

By employing (70) with $\mathcal{G}(\omega) \equiv S$ and denoting $F^S(\omega) := \inf_{x \in S} f(x, \omega)$, we obtain that the optimal value of problem (69) is equal to $\rho(F^S)$, provided

[20] A multifunction $\mathcal{G} : \Omega \rightrightarrows \mathbb{R}^n$ maps a point $\omega \in \Omega$ into a set $\mathcal{G}(\omega) \subset \mathbb{R}^n$. It is said that $\mathcal{G}$ is *closed valued* if $\mathcal{G}(\omega)$ is a closed subset of $\mathbb{R}^n$ for any $\omega \in \Omega$. It is said that $\mathcal{G}$ is *measurable* if for any closed set $A \subset \mathbb{R}^n$ the inverse image set $\mathcal{G}^{-1}(A) := \{\omega \in \Omega : \mathcal{G}(\omega) \in \mathcal{A}\}$ is $\mathcal{F}$-measurable.

that $F^S \in \mathcal{Z}$. The difference between the optimal values of problems (61) and (69), that is

$$\mathrm{RVPI}_\rho := \inf_{x \in S} \rho[F(x)] - \rho(F^S), \tag{72}$$

is called the *Risk Value of Perfect Information*. Since problem (69) is a relaxation of problem (61), we have that RVPI_ρ is nonnegative. It is also possible to show that if ρ is real valued and satisfies conditions (A1)–(A4), and hence representation (14) holds, then

$$\inf_{\mu \in \mathcal{A}} \mathrm{EVPI}_\mu \le \mathrm{RVPI}_\rho \le \sup_{\mu \in \mathcal{A}} \mathrm{EVPI}_\mu, \tag{73}$$

where

$$\mathrm{EVPI}_\mu := \inf_{x \in S} \mathbb{E}\big[f(x,\omega)\big] - \mathbb{E}_\mu\big[\inf_{x \in S} f(x,\omega)\big] \tag{74}$$

is the Expected Value of Perfect Information associated with the probability measure μ (cf., [23]).

9 Dualization of Nonanticipativity Constraints

In addition to the assumptions (i)–(iii) of section 7, we assume in this section that $\mathcal{Z} := \mathcal{L}_p(\Omega, \mathcal{F}, P)$ and $\mathcal{Z}^* := \mathcal{L}_q(\Omega, \mathcal{F}, P)$, and that $\mathcal{M}^* := \mathcal{L}_q(\Omega, \mathcal{F}, P, \mathbb{R}^n)$ is the dual of the space $\mathcal{M} := \mathcal{L}_p(\Omega, \mathcal{F}, P, \mathbb{R}^n)$. Consider the Lagrangian

$$L_0(X, x, \lambda) := \rho(F_X) + \mathbb{E}[\lambda^T (X - x)], \quad (X, x, \lambda) \in \mathcal{M} \times \mathbb{R}^n \times \mathcal{M}^*,$$

associated with the nonanticipativity constraints of problem (68). Note that problem (68) can be written in the following equivalent form

$$\operatorname*{Min}_{X \in \mathcal{M}_S,\, x \in \mathbb{R}^n} \left\{ \sup_{\lambda \in \mathcal{M}^*} L_0(X, x, \lambda) \right\}. \tag{75}$$

By interchanging the 'min' and 'max' operators in (75) we obtain the (Lagrangian) dual of problem (68). Observe that $\inf_{x \in \mathbb{R}^n} L_0(X, x, \lambda)$ is equal to $-\infty$ if $\mathbb{E}[\lambda] \neq 0$, and to $L(X, \lambda)$ if $\mathbb{E}[\lambda] = 0$, where

$$L(X, \lambda) := \rho(F_X) + \mathbb{E}[\lambda^T X].$$

Therefore the (Lagrangian) dual of problem (68) takes on the form

$$\operatorname*{Max}_{\lambda \in \mathcal{M}^*} \left\{ \inf_{X \in \mathcal{M}_S} L(X, \lambda) \right\} \text{ subject to } \mathbb{E}[\lambda] = 0. \tag{76}$$

By the standard theory of Lagrangian duality we have that the optimal value of the primal problem (68) is greater than or equal to the optimal value of the dual problem (76). Moreover, under appropriate regularity conditions, there is no duality gap between problems (68) and (76), i.e., their optimal values are equal to each other. In particular, if the Lagrangian $L_0(X, x, \lambda)$ possesses a saddle point $((\bar{X}, \bar{x}), \bar{\lambda})$, then $(\bar{X}, \bar{x})$ and $\bar{\lambda}$ are optimal solutions of problems (68) and (76), respectively, and there is no duality gap between problems (68) and (76). Noting that $L_0(X, x, \lambda)$ is linear in x and λ, we obtain that $((\bar{X}, \bar{x}), \bar{\lambda})$ is a saddle point iff the following conditions hold:

$$\bar{X}(\omega) = \bar{x}, \ a.e. \ \omega \in \Omega, \text{ and } \mathbb{E}[\bar{\lambda}] = 0, \tag{77}$$

$$\bar{X} \in \operatorname*{argmin}_{X \in \mathcal{M}_S} L(X, \bar{\lambda}). \tag{78}$$

Consider function $\Phi(X) := \rho(F_X) : \mathcal{M} \to \overline{\mathbb{R}}$. Because of convexity of F and assumptions (A1) and (A2), this function is convex. Its subdifferential $\partial\Phi(X) \subset \mathcal{M}^*$ is defined in the usual way. By convexity, assuming that ρ is continuous at $\bar{Z} := F(\bar{x})$, we can write the following optimality conditions for (78) to hold:

$$-\bar{\lambda} \in N_S(\bar{x}) + \partial\Phi(\bar{X}). \tag{79}$$

Therefore we obtain that if problem (61) possesses an optimal solution $\bar{x}$, then the Lagrangian $L_0(X, x, \lambda)$ has a saddle point iff there exists $\bar{\lambda} \in \mathcal{M}^*$ satisfying condition (79) and such that $\mathbb{E}[\bar{\lambda}] = 0$. We shall now show the existence of such $\bar{\lambda}$.

By formula (51) and the optimality condition

$$0 \in N_S(\bar{x}) + \partial\phi(\bar{x}),$$

we have that (under the assumptions of Theorem 4) there exists $\zeta \in \partial\rho(\bar{Z})$ such that

$$0 \in N_S(\bar{x}) + \int_\Omega \zeta(\omega)\partial f_\omega(\bar{x})\, dP(\omega).$$

This means that there exists a measurable selection $g(\omega) \in \partial f_\omega(\bar{x})$ such that

$$0 \in N_S(\bar{x}) + \int_\Omega \zeta(\omega)g(\omega)\, dP(\omega). \tag{80}$$

Let us now define

$$\bar{\lambda}(\omega) := \int_\Omega \zeta(\omega)g(\omega)\, dP(\omega) - \zeta(\omega)g(\omega), \quad \omega \in \Omega.$$

By construction, $\mathbb{E}[\bar{\lambda}] = 0$. Furthermore, since $g(\omega) \in \partial f_\omega(\bar{x})$, we have that

$$f_\omega(X(\omega)) \geq f_\omega(\bar{x}) + g(\omega)^T(X(\omega) - \bar{x}), \ \ \omega \in \Omega.$$

Because of $\zeta \in \partial\rho(\bar{Z})$, this implies that

$$\rho(F_X) \geq \rho(F(\bar{x})) + \int_\Omega \zeta(\omega)g(\omega)^T(X(\omega) - \bar{x})\, dP(\omega).$$

We obtain that $\zeta g \in \partial \Phi(\bar{X})$. This together with equation (80) imply that $\bar{\lambda}$ satisfies condition (79).

We obtain the following result.

Proposition 5. *Suppose that problem* (61) *possesses an optimal solution* $\bar{x}$ *and the assumptions of Theorem* 4 *hold. Then there exists* $\bar{\lambda}$ *such that* $((\bar{X},\bar{x}),\bar{\lambda})$*, where* $\bar{X}(\omega) \equiv \bar{x}$*, is a saddle point of the Lagrangian* $L_0(X,x,\lambda)$*, and hence there is no duality gap between problems* (61) *and* (76)*, and* $(\bar{X},\bar{x})$ *and* $\bar{\lambda}$ *are optimal solutions of problems* (68) *and* (76)*, respectively.*

Let us return to the question of decomposing problem (78). Suppose that ρ is real valued and conditions (A1)–(A3) are satisfied, and hence by Theorem 2 representation (13) holds. Then

$$\inf_{X \in \mathcal{M}_S} L(X,\lambda) = \inf_{X \in \mathcal{M}_S} \sup_{\zeta \in \mathcal{P}} \left\{ \mathbb{E}[\zeta F_X + \lambda^T X] - \rho^*(\zeta) \right\}. \tag{81}$$

Suppose, further, that the 'inf' and 'sup' operators at the right hand side of the above equation (81) can be interchanged (note that the function inside the parentheses in the right hand side of (81) is convex in X and concave in ζ). Then

$$\begin{aligned} \inf_{X \in \mathcal{M}_S} L(X,\lambda) &= \sup_{\zeta \in \mathcal{P}} \inf_{X \in \mathcal{M}_S} \left\{ \mathbb{E}[\zeta F_X + \lambda^T X] - \rho^*(\zeta) \right\} \\ &= \sup_{\zeta \in \mathcal{P}} \left\{ \mathbb{E}\Big(\inf_{x \in S} [\zeta(\omega) f(x,\omega) + \lambda(\omega)^T x] \Big) - \rho^*(\zeta) \right\}, \end{aligned}$$

where the last equality follows by the interchangeability principle. Therefore, we obtain that, under the specified assumptions, the optimal value of the dual problem (76) is equal to $\sup_{\mathbb{E}[\lambda]=0, \zeta \in \mathcal{P}} D(\lambda,\zeta)$, where

$$D(\lambda,\zeta) := \mathbb{E}\left\{ \inf_{x \in S} [\zeta(\omega) f(x,\omega) + \lambda(\omega)^T x] \right\} - \rho^*(\zeta). \tag{82}$$

If, moreover, there is no duality gap between problems (61) and (76), then the following duality relation holds

$$\inf_{x \in S} \rho[F(x)] = \sup_{\substack{\lambda \in \mathcal{M}^*,\, \zeta \in \mathcal{P} \\ \mathbb{E}[\lambda]=0}} D(\lambda,\zeta). \tag{83}$$

Note the separable structure of the right hand side of (82). That is, in order to calculate $D(\lambda,\zeta)$ one needs to solve the minimization problem inside the parentheses at the right hand side of (82) separately for every $\omega \in \Omega$, and then to take the expectation of the optimal values calculated.

10 Two-Stage Programming

Suppose now that the function $f(x,\omega)$ is given in the form

$$f(x,\omega) := \inf_{y\in\mathcal{G}(x,\omega)} g(x,y,\omega), \tag{84}$$

where $g : \mathbb{R}^n \times \mathbb{R}^m \times \Omega \to \mathbb{R}$ is a random lower semicontinuous function and $\mathcal{G} : \mathbb{R}^n \times \Omega \rightrightarrows \mathbb{R}^m$ is a closed valued measurable multifunction. Note that it follows that the optimal value function $f(x,\omega)$ is measurable, and moreover is random lower semicontinuous provided that $\mathcal{G}(\cdot,\omega)$ are locally uniformly bounded. We refer to the corresponding problem (61) as a *two-stage program*. For example, if the set S is polyhedral,

$$g(x,y,\omega) := c^T x + q(\omega)^T y, \tag{85}$$

$$\mathcal{G}(x,\omega) := \big\{y : T(\omega)x + W(\omega)y = h(\omega),\ y \geq 0\big\}, \tag{86}$$

and $\rho(Z) \equiv \mathbb{E}[Z]$, then problem (61) becomes a two-stage linear stochastic programming problem.

It is important to note that it is implicitly assumed here that for every $x \in S$ the optimal value $f(x,\omega)$ is *finite* for all $\omega \in \Omega$. In particular, this requires the second stage problem to be feasible (i.e., $\mathcal{G}(x,\omega) \neq \emptyset$) for every $\omega \in \Omega$. That is, it requires the considered two-stage problem to have a relatively complete recourse.

Suppose that ρ satisfies conditions (A1) and (A2). Then by the interchangeability formula (70) we have that, for a fixed $x \in S$,

$$\rho(F(x)) = \inf_{\substack{Y\in\mathcal{M}\\ Y(\cdot)\in\mathcal{G}(x,\cdot)}} \rho\left[\Gamma_Y(x)\right],$$

where $[\Gamma_Y(x)](\omega) := g(x,Y(\omega),\omega)$ and $\mathcal{M} := \mathcal{L}_p(\Omega,\mathcal{F},P,\mathbb{R}^m)$. Consequently the first stage problem (61) is equivalent to the problem

$$\operatorname*{Min}_{x\in S,\, Y\in\mathcal{M}} \ \rho\left[\Gamma_Y(x)\right] \ \text{ s.t. } \ Y(\omega) \in \mathcal{G}(x,\omega) \ a.e.\ \omega \in \Omega. \tag{87}$$

Note again that the key property ensuring equivalence of problems (61) and (87) is the monotonicity condition (A2).

If the set $\Omega = \{\omega_1,\dots,\omega_K\}$ is finite, we can identify space $\mathcal{L}_p(\Omega,\mathcal{F},P,\mathbb{R}^m)$ with the finite dimensional space $\mathbb{R}^{mK}$ of vectors $Y = (y_1,\dots,y_K)$. In that case $\Gamma_Y(x) = (g(x,y_1,\omega_1),\dots,g(x,y_K,\omega_K)) \in \mathbb{R}^K$ and ρ is a function from $\mathbb{R}^K$ to $\mathbb{R}$. Then problem (87) can be written in the form

$$\operatorname*{Min}_{x\in\mathbb{R}^n,\, Y\in\mathbb{R}^{mK}} \ \rho\left[\Gamma_Y(x)\right] \ \text{ s.t. } \ x \in S,\ y_k \in \mathcal{G}(x,\omega_k),\ k = 1,\dots,K. \tag{88}$$

In particular, if the function g and mapping $\mathcal{G}$ are given in the form (85) and (86), respectively, then problem (88) takes the form

$$\begin{array}{l} \operatorname*{Min}\limits_{x\in S,\, Y\in\mathbb{R}^{mK}} \ \rho\big(c^T x + q_1^T y_1,\dots,c^T x + q_K^T y_K\big) \\ \text{subject to } \ T_k x + W_k y_k = h_k,\ y_k \geq 0,\ \ k = 1,\dots,K, \end{array} \tag{89}$$

where $q_k := q(\omega_k)$, $T_k := T(\omega_k)$, $W_k := W(\omega_k)$ and $h_k := h(\omega_k)$. If, further, condition (A3) is satisfied, then

$$\rho\big(c^T x + q_1^T y_1, \dots, c^T x + q_K^T y_K\big) = c^T x + \rho\big(q_1^T y_1, \dots, q_K^T y_K\big).$$

Assume now that condition (A4) also holds true. Then the set $\mathcal{A}$ of probability measures, constituting the domain of the conjugate function ρ^*, can be identified with a certain convex closed subset of the simplex in $\mathbb{R}^K$:

$$\mathcal{A} \subset \Big\{ p \in \mathbb{R}^K : \textstyle\sum_{k=1}^K p_k = 1,\ p_k \geq 0,\ k = 1, \dots, K \Big\}.$$

In this case we can rewrite problem (89) as follows

$$\begin{array}{ll} \underset{x\in S,\, Y\in\mathbb{R}^{mK}}{\text{Min}} & \left(c^T x + \max\limits_{p\in\mathcal{A}} \displaystyle\sum_{k=1}^K p_k q_k^T y_k \right) \\ \text{subject to } T_k x + W_k y_k = h_k,\ y_k \geq 0, \quad k = 1, \dots, K. & \end{array}$$

In the following sections of this chapter we shall extend this observation to multistage problems.

11 Conditional Risk Mappings

In order to construct dynamic models of risk we need to extend the concept of a risk function. In multi-stage (dynamic) stochastic programming the main theoretical tool is the concept of *conditional expectation*. That is, let $(\Omega, \mathcal{F}_2, P)$ be a probability space, $\mathcal{F}_1$ be a sigma subalgebra of $\mathcal{F}_2$, i.e., $\mathcal{F}_1 \subset \mathcal{F}_2$, and $\mathcal{X}_i$, $i = 1, 2$, be spaces of all $\mathcal{F}_i$-measurable and P-integrable functions $Z : \Omega \to \mathbb{R}$. The conditional expectation $\mathbb{E}[\,\cdot\,|\mathcal{F}_1]$ is defined as a mapping from $\mathcal{X}_2$ into $\mathcal{X}_1$ such that

$$\int_A \mathbb{E}[Z|\mathcal{F}_1](\omega)\, dP(\omega) = \int_A Z(\omega) dP(\omega), \text{ for all } A \in \mathcal{F}_1 \text{ and } Z \in \mathcal{X}_2.$$

The approach that we adopt here is aimed at extending this concept to risk mappings. Our presentation is based on [24]. Let $(\Omega, \mathcal{F}_2)$ be a measurable space, $\mathcal{F}_1$ be a sigma subalgebra of $\mathcal{F}_2$, and $\mathcal{Z}_i$, $i = 1, 2$, be linear spaces of $\mathcal{F}_i$-measurable functions $Z : \Omega \to \mathbb{R}$. We assume that $\mathcal{Z}_1 \subset \mathcal{Z}_2$ and each space $\mathcal{Z}_i$ is sufficiently large such that it includes all $\mathcal{F}_i$-measurable step functions, i.e., condition (C) is satisfied. Also we assume that with each $\mathcal{Z}_i$ is paired a dual space $\mathcal{Z}_i^*$ of finite signed measures on $(\Omega, \mathcal{F}_i)$. In applications we typically use spaces $\mathcal{Z}_i := \mathcal{L}_p(\Omega, \mathcal{F}_i, P)$ and $\mathcal{Z}_i^* := \mathcal{L}_q(\Omega, \mathcal{F}_i, P)$ for some (reference) probability measure P. At this moment, however, this is not essential and is not assumed. Let $\rho : \mathcal{Z}_2 \to \mathcal{Z}_1$ be a mapping, referred to as *risk mapping*. Consider the following conditions[21]:

---

[21] Recall that the relation $Z_1 \preceq Z_2$ denotes the inequality $Z_1(\omega) \leq Z_2(\omega)$ for all $\omega \in \Omega$.

(M1) *Convexity:*

$$\rho(\alpha Z_1 + (1-\alpha)Z_2) \preceq \alpha\rho(Z_1) + (1-\alpha)\rho(Z_2)$$

for all $Z_1, Z_2 \in \mathcal{Z}_2$ and all $\alpha \in [0,1]$.

(M2) *Monotonicity:* If $Z_1, Z_2 \in \mathcal{Z}_2$ and $Z_2 \succeq Z_1$, then $\rho(Z_2) \succeq \rho(Z_1)$.

(M3) *Translation Equivariance:* If $Y \in \mathcal{Z}_1$ and $Z \in \mathcal{Z}_2$, then

$$\rho(Z+Y) = \rho(Z) + Y.$$

(M4) *Positive Homogeneity:* If $\alpha > 0$ and $Z \in \mathcal{Z}_2$, then $\rho(\alpha Z) = \alpha\rho(Z)$.

If the sigma algebra $\mathcal{F}_1$ is trivial, i.e., $\mathcal{F}_1 = \{\emptyset, \Omega\}$, then any $\mathcal{F}_1$-measurable function is constant over Ω, and hence the space $\mathcal{Z}_1$ can be identified with $\mathbb{R}$. In that case ρ maps $\mathcal{Z}_2$ into the real line $\mathbb{R}$, and conditions (M1)–(M4) coincide with the respective conditions (A1)–(A4). In order to emphasize that the risk mapping ρ is associated with spaces $\mathcal{Z}_1$ and $\mathcal{Z}_2$ we sometimes write it as $\rho_{\mathcal{Z}_2|\mathcal{Z}_1}$. We say that the risk mapping ρ is a *conditional* risk mapping if it satisfies conditions (M1)–(M3).

Remark 1. Note that if $Y \in \mathcal{Z}_1$, then we have by condition (M3) that

$$\rho(Y) = \rho(0+Y) = Y + \rho(0).$$

If, moreover, ρ is positively homogeneous (i.e., condition (M4) holds), then $\rho(0) = 0$. Therefore, if conditions (M1)–(M4) hold, then $\rho(Y) = Y$ for any $Y \in \mathcal{Z}_1$.

For $\omega \in \Omega$, we associate with a risk mapping ρ the function

$$\rho_\omega(Z) := [\rho(Z)](\omega), \quad Z \in \mathcal{Z}_2. \tag{90}$$

Note that since it is assumed that all functions of the space $\mathcal{Z}_1$ are real valued, it follows that ρ_ω maps $\mathcal{Z}_2$ into $\mathbb{R}$, i.e., $\rho_\omega(\cdot)$ is also real valued. Conditions (M1), (M2) and (M4) simply mean that function ρ_ω satisfies the respective conditions (A1), (A2) and (A4) for every $\omega \in \Omega$. Condition (M3) implies (but is not equivalent) condition (A3) for the functions ρ_ω, $\omega \in \Omega$.

We say that the mapping ρ is lower semicontinuous if for every $\omega \in \Omega$ the corresponding function ρ_ω is lower semicontinuous. With each function $\rho_\omega : \mathcal{Z}_2 \to \mathbb{R}$ is associated its conjugate function $\rho^*_\omega : \mathcal{Z}^*_2 \to \overline{\mathbb{R}}$, defined in (7). Note that although ρ_ω is real valued, it can happen that $\rho^*_\omega(\mu) = +\infty$ for some $\mu \in \mathcal{Z}^*_2$.

By $\mathcal{P}_{\mathcal{Z}^*_i}$ we denote the set of all probability measures on $(\Omega, \mathcal{F}_i)$ which are in $\mathcal{Z}^*_i$. Moreover, with each $\omega \in \Omega$ we associate a set of probability measures $\mathcal{P}_{\mathcal{Z}^*_2|\mathcal{F}_1}(\omega) \subset \mathcal{P}_{\mathcal{Z}^*_2}$ formed by all $\nu \in \mathcal{P}_{\mathcal{Z}^*_2}$ such that for every $B \in \mathcal{F}_1$ it holds that

$$\nu(B) = \begin{cases} 1, & \text{if } \omega \in B, \\ 0, & \text{if } \omega \notin B. \end{cases} \tag{91}$$

Note that ω is fixed here and B varies in $\mathcal{F}_1$. Condition (91) simply means that for every ω and every $B \in \mathcal{F}_1$ we know whether B happened or not. In particular, if $\mathcal{F}_1 = \{\emptyset, \Omega\}$, then $\mathcal{P}_{\mathcal{Z}_2^*|\mathcal{F}_1}(\omega) = \mathcal{P}_{\mathcal{Z}_2^*}$ for all $\omega \in \Omega$.

We can now formulate the basic duality result for conditional risk mappings (cf., [24]) which can be viewed as an extension of Theorem 2. Recall that $\langle \mu, Z \rangle = \mathbb{E}_\mu[Z]$ for $\mu \in \mathcal{P}_{\mathcal{Z}_i^*}$ and $Z \in \mathcal{Z}_i$.

Theorem 5. *Let $\rho = \rho_{\mathcal{Z}_2|\mathcal{Z}_1}$ be a lower semicontinuous conditional risk mapping satisfying conditions* (M1)–(M3)*. Then for every $\omega \in \Omega$ it holds that*

$$\rho_\omega(Z) = \sup_{\mu \in \mathcal{P}_{\mathcal{Z}_2^*|\mathcal{F}_1}(\omega)} \big\{ \langle \mu, Z \rangle - \rho_\omega^*(\mu) \big\}, \quad \forall\, Z \in \mathcal{Z}_2. \tag{92}$$

Moreover, if ρ is positively homogeneous (i.e., condition (M4) *holds), then for every $\omega \in \Omega$ there is a closed convex set $\mathcal{A}(\omega) \subset \mathcal{P}_{\mathcal{Z}_2^*|\mathcal{F}_1}(\omega)$ such that*

$$\rho_\omega(Z) = \sup_{\mu \in \mathcal{A}(\omega)} \langle \mu, Z \rangle, \quad \forall\, Z \in \mathcal{Z}_2. \tag{93}$$

Conversely, suppose that a mapping $\rho : \mathcal{Z}_2 \to \mathcal{Z}_1$ can be represented in form (92) *for some $\rho^* : \mathcal{Z}_2^* \times \Omega \to \overline{\mathbb{R}}$. Then ρ is lower semicontinuous and satisfies conditions* (M1)–(M3)*.*

Remark 2. As it was mentioned in the discussion following Theorem 1, if $\mathcal{Z}_2$ is a Banach lattice (e.g., $\mathcal{Z}_2 := \mathcal{L}_p(\Omega, \mathcal{F}_2, P)$) and ρ satisfies conditions (M1) and (M2), then for any $\omega \in \Omega$ the corresponding function $\rho_\omega : \mathcal{Z}_2 \to \mathbb{R}$ is continuous, and hence is lower semicontinuous. Therefore, in the case of $\mathcal{Z}_2 := \mathcal{L}_p(\Omega, \mathcal{F}_2, P)$, the assumption of lower semicontinuity of ρ in the above theorem holds true automatically.

Remark 3. The concept of conditional risk mappings is closely related to the concept of conditional expectations. Let P be a probability measure on $(\Omega, \mathcal{F}_2)$ and suppose that every $Z \in \mathcal{Z}_2$ is P-integrable. For $Z \in \mathcal{Z}_2$, define

$$\rho(Z) := \mathbb{E}[Z|\mathcal{F}_1].$$

Suppose, further, that the space $\mathcal{Z}_1$ is large enough so that it contains $\mathbb{E}[Z|\mathcal{F}_1]$ for all $Z \in \mathcal{Z}_2$. Then $\rho : \mathcal{Z}_2 \to \mathcal{Z}_1$ is a well defined[22] mapping. The conditional expectation mapping ρ satisfies conditions (M1)–(M3) and is linear, and hence is positively homogeneous. The representation (93) holds with $\mathcal{A}(\omega) = \{\mu(\omega)\}$ being a singleton and $\mu_\omega = \mu(\omega)$ being a probability measure on $(\Omega, \mathcal{F}_2)$ for every $\omega \in \Omega$. Moreover, for any $A \in \mathcal{F}_2$ it holds that

$$\mu_\omega(A) = \mathbb{E}[\mathbb{1}_A|\mathcal{F}_1](\omega) = [P(A|\mathcal{F}_1)](\omega).$$

That is, $\mu(\cdot)$ is the conditional probability of P with respect to $\mathcal{F}_1$. Note that $\mathbb{E}[Z|\mathcal{F}_1](\omega) = \mathbb{E}_{\mu_\omega}[Z]$.

[22] Note that the function $\mathbb{E}[Z|\mathcal{F}_1](\cdot)$ is defined up to a set of P-measure zero, i.e., two versions of $\mathbb{E}[Z|\mathcal{F}_1](\cdot)$ can be different on a set of P-measure zero.

The family of conditional risk mappings is closed under the operation of taking maximum. That is, let $\{\rho^\nu = \rho^\nu_{\mathcal{Z}_2|\mathcal{Z}_1}\}_{\nu\in\mathcal{I}}$ be a family of conditional risk mappings satisfying assumptions (M1)–(M3). Suppose, further, that for every $Z \in \mathcal{Z}_2$ the function

$$[\rho(Z)](\cdot) := \sup_{\nu\in\mathcal{I}} \big[\rho^\nu(Z)\big](\cdot) \tag{94}$$

belongs to the space $\mathcal{Z}_1$, and hence ρ maps $\mathcal{Z}_2$ into $\mathcal{Z}_1$. It is then straightforward to verify that the max-function ρ also satisfies assumptions (M1)–(M3). Moreover, if ρ^ν, $\nu \in \mathcal{I}$, are lower semicontinuous and/or positively homogeneous, then ρ is also lower semicontinuous and/or positively homogeneous. In particular, let $\rho^\nu(Z) := \mathbb{E}_\nu[Z|\mathcal{F}_1]$, $\nu \in \mathcal{I}$, where $\mathcal{I}$ is a family of probability measures on $(\Omega, \mathcal{F}_2)$. Suppose that the corresponding max-function

$$[\rho(Z)](\cdot) := \sup_{\nu\in\mathcal{I}} \mathbb{E}_\nu[Z|\mathcal{F}_1](\cdot) \tag{95}$$

is well defined, i.e., ρ maps $\mathcal{Z}_2$ into $\mathcal{Z}_1$. Then ρ is a lower semicontinuous positively homogeneous conditional risk mapping. It is possible to show that, under certain regularity conditions, the converse is also true, i.e., a positively homogeneous conditional risk mapping can be represented in form (95) (cf., [24]).

12 Multistage Optimization Problems

In this section we discuss optimization of risk measures in a dynamical setting. We use the following framework. Let $(\Omega, \mathcal{F})$ be a measurable space and $\mathcal{F}_1 \subset \mathcal{F}_2 \subset \cdots \subset \mathcal{F}_T$ be a sequence of sigma algebras such that $\mathcal{F}_1 = \{\emptyset, \Omega\}$ and $\mathcal{F}_T = \mathcal{F}$. Let $\mathcal{Z}_1 \subset \mathcal{Z}_2 \subset \cdots \subset \mathcal{Z}_T$ be a corresponding sequence of linear spaces of $\mathcal{F}_t$ measurable functions, $t = 1, \ldots, T$, and let $\rho_{\mathcal{Z}_t|\mathcal{Z}_{t-1}} : \mathcal{Z}_t \to \mathcal{Z}_{t-1}$ be conditional risk mapings satisfying assumptions (M1)–(M3). Also consider a sequence of functions $Z_t \in \mathcal{Z}_t$, $t = 1, \ldots, T$. By the definition of spaces $\mathcal{Z}_t$, each function $Z_t : \Omega \to \mathbb{R}$ is $\mathcal{F}_t$-measurable, and since the sigma algebra $\mathcal{F}_1$ is trivial, $Z_1(\omega)$ is constant and the space $\mathcal{Z}_1$ can be identified with $\mathbb{R}$.

Consider the composite mappings[23] $\rho_{\mathcal{Z}_{t-1}|\mathcal{Z}_{t-2}} \circ \rho_{\mathcal{Z}_t|\mathcal{Z}_{t-1}}$. Let us observe that conditions (M1)–(M4) are preserved by such compositions. That is, if conditions (M1) and (M2) (and also (M3), (M4)) hold for mappings $\rho_{\mathcal{Z}_{t-1}|\mathcal{Z}_{t-2}}$ and $\rho_{\mathcal{Z}_t|\mathcal{Z}_{t-1}}$, then these conditions hold for their composition as well. Therefore the assumption that the mappings $\rho_{\mathcal{Z}_t|\mathcal{Z}_{t-1}}$, $t = 2, \ldots, T$, satisfy conditions (M1)–(M3) implies that the risk functions

$$\rho_t := \rho_{\mathcal{Z}_2|\mathcal{Z}_1} \circ \cdots \circ \rho_{\mathcal{Z}_t|\mathcal{Z}_{t-1}} : \mathcal{Z}_t \to \mathbb{R}, \;\; t = 2, \ldots, T,$$

[23] The composite mapping $\rho_{\mathcal{Z}_{t-1}|\mathcal{Z}_{t-2}} \circ \rho_{\mathcal{Z}_t|\mathcal{Z}_{t-1}} : \mathcal{Z}_t \to \mathcal{Z}_{t-2}$ maps $Z_t \in \mathcal{Z}_t$ into $\rho_{\mathcal{Z}_{t-1}|\mathcal{Z}_{t-2}}\big[\rho_{\mathcal{Z}_t|\mathcal{Z}_{t-1}}(Z_t)\big]$.

satisfy conditions (A1)–(A3). Moreover, consider the space $\mathcal{Z} := \mathcal{Z}_1 \times \cdots \times \mathcal{Z}_T$ and $Z := (Z_1, \ldots, Z_T) \in \mathcal{Z}$. Define function $\tilde{\rho} : \mathcal{Z} \to \mathbb{R}$ as follows:

$$\begin{aligned} \tilde{\rho}(Z) := Z_1 + \rho_{\mathcal{Z}_2|\mathcal{Z}_1}\Big[Z_2 + \rho_{\mathcal{Z}_3|\mathcal{Z}_2}\Big(Z_3 + \ldots \\ \cdots + \rho_{\mathcal{Z}_{T-1}|\mathcal{Z}_{T-2}}\big[Z_{T-1} + \rho_{\mathcal{Z}_T|\mathcal{Z}_{T-1}}(Z_T)\big]\Big)\Big]. \end{aligned} \tag{96}$$

By assumption (M3) we have

$$Z_{T-1} + \rho_{\mathcal{Z}_T|\mathcal{Z}_{T-1}}(Z_T) = \rho_{\mathcal{Z}_T|\mathcal{Z}_{T-1}}(Z_{T-1} + Z_T),$$

and so on for $t = T-1, \ldots, 2$. Therefore we obtain that

$$\tilde{\rho}(Z) = \rho_T(Z_1 + \cdots + Z_T). \tag{97}$$

Thus, condition (M3) allows us to switch between the cumulative formulation (97) and nested formulation (96).

Remark 4. As it was mentioned above, we have that if $\rho_{\mathcal{Z}_{t-1}|\mathcal{Z}_{t-2}}$ and $\rho_{\mathcal{Z}_t|\mathcal{Z}_{t-1}}$ are positively homogeneous risk mappings, then the composite mapping $\rho_{\mathcal{Z}_t|\mathcal{Z}_{t-2}} := \rho_{\mathcal{Z}_{t-1}|\mathcal{Z}_{t-2}} \circ \rho_{\mathcal{Z}_t|\mathcal{Z}_{t-1}}$ is also a positively homogeneous risk mapping. By virtue of Theorem 5, with these mappings are associated closed convex sets $\mathcal{A}_{t-1,t-2}(\omega)$, $\mathcal{A}_{t,t-1}(\omega)$ and $\mathcal{A}_{t,t-2}(\omega)$, depending on $\omega \in \Omega$, such that the corresponding representation (93) holds, provided that these mappings are lower semicontinuous. It is possible to show (cf., [24]) that $\mathcal{A}_{t,t-2}(\omega)$ is formed by all measures $\mu \in \mathcal{Z}_t^*$ representable in the form

$$\mu(A) = \int_\Omega [\mu_2(\tilde{\omega})](A)\, d\mu_1(\tilde{\omega}), \quad A \in \mathcal{F}_t, \tag{98}$$

where $\mu_1 \in \mathcal{A}_{t-1,t-2}(\omega)$ and $\mu_2(\cdot) \in \mathcal{A}_{t,t-1}(\cdot)$ is a weakly* $\mathcal{F}_t$-measurable selection. Unfortunately, this formula is not very constructive and in general it could be quite difficult to calculate the dual representation of the composite mapping explicitly.

Remark 5. Consider the composite function $\tilde{\rho}(\cdot)$. As we mentioned in Remark 4, it could be difficult to write it explicitly. The situation simplifies considerably if we assume a certain type of "between stages independence" condition. That is, suppose that $Z \in \mathcal{Z}$ is such that the functions $\big[\rho_{\mathcal{Z}_t|\mathcal{Z}_{t-1}}(Z_t)\big](\omega)$, $t = T, \ldots, 2$, are constants, i.e., independent of ω. Then by condition (M3) we have that

$$\tilde{\rho}(Z) = Z_1 + \rho_{\mathcal{Z}_2|\mathcal{Z}_1}(Z_2) + \cdots + \rho_{\mathcal{Z}_T|\mathcal{Z}_{T-1}}(Z_T). \tag{99}$$

We discuss this further in Example 12 of the following section.

Now let us formulate a multistage optimization problem involving risk mappings. Suppose that we are given functions $f_t : \mathbb{R}^{n_t} \times \Omega \to \mathbb{R}$ and multifunctions $\mathcal{G}_t : \mathbb{R}^{n_{t-1}} \times \Omega \rightrightarrows \mathbb{R}^{n_t}$, $t = 1, \ldots, T$. We assume that the functions $f_t(x_t, \omega)$ are $\mathcal{F}_t$-random lower semicontinuous[24], and the multifunctions

[24] Recall that it is said that function $f_t(x_t, \omega)$ is $\mathcal{F}_t$-random lower semicontinuous if its epigraphical mapping is closed valued and $\mathcal{F}_t$-measurable.

$\mathcal{G}_t(x_{t-1},\cdot)$ are closed valued and $\mathcal{F}_t$-measurable. Note that since the sigma algebra $\mathcal{F}_1$ is trivial, the function $f_1 : \mathbb{R}^{n_1} \to \mathbb{R}$ does not depend on $\omega \in \Omega$, and by the definition $\mathcal{G}_1(\omega) \equiv G_1$, where G_1 is a fixed closed subset of $\mathbb{R}^{n_1}$. Let $\mathcal{M}_t$, $t = 1,\dots,T$, be linear spaces of $\mathcal{F}_t$-measurable functions $X_t : \Omega \to \mathbb{R}^{n_t}$, and $\mathcal{M} := \mathcal{M}_1 \times \cdots \times \mathcal{M}_T$. With functions f_t we associate mappings $F_t : \mathcal{M}_t \to \mathcal{Z}_t$ defined as follows

$$[F_t(X_t)](\omega) := f_t(X_t(\omega),\omega), \quad X_t \in \mathcal{M}_t.$$

Since $\mathcal{F}_1$ is trivial, the space $\mathcal{M}_1$ can be identified with $\mathbb{R}^{n_1}$, and hence $F_1(X_1) = f_1(X_1)$.

Consider the problem

$$\begin{array}{ll} \operatorname*{Min}_{X \in \mathcal{M}} & \rho_T\big(F_1(X_1) + \cdots + F_T(X_T)\big), \\ \text{s.t.} & X_t(\omega) \in \mathcal{G}_t(X_{t-1}(\omega),\omega),\ \omega \in \Omega,\ t = 1,\dots,T. \end{array} \tag{100}$$

We refer to (100) as a multistage risk optimization problem. By (96) and (97) we can write the equivalent nested formulation:

$$\begin{aligned} \rho_T\big(F_1(X_1) + \cdots + F_T(X_T)\big) = F_1(X_1) + \rho_{\mathcal{Z}_2|\mathcal{Z}_1}\Big[F_2(X_2) + \\ \cdots + \rho_{\mathcal{Z}_{T-1}|\mathcal{Z}_{T-2}}\big[F_{T-1}(X_{T-1}) + \rho_{\mathcal{Z}_T|\mathcal{Z}_{T-1}}\big(F_T(X_T)\big)\big]\Big]. \end{aligned} \tag{101}$$

Since $X \in \mathcal{M}$, it is assumed that $X_t(\omega)$ are $\mathcal{F}_t$-measurable, and hence (100) is adapted to the filtration $\mathcal{F}_t$, $t = 1,\dots,T$.

As an example consider the following linear setting. Suppose that[25]

$$\begin{gathered} f_t(x_t,\omega) := c_t(\omega)\cdot x_t, \\ \mathcal{G}_t(x_{t-1},\omega) := \big\{x_t \in \mathbb{R}^{n_t} : B_t(\omega)x_{t-1} + A_t(\omega)x_t = b_t(\omega),\ x_t \geq 0\big\}, \end{gathered} \tag{102}$$

where $c_t(\omega), b_t(\omega)$ are vectors and $B_t(\omega), A_t(\omega)$ are matrices of appropriate dimensions. It is assumed that the corresponding vector-valued functions

$$\xi_t(\omega) := (c_t(\omega), B_t(\omega), A_t(\omega), b_t(\omega)), \quad t = 1,\dots,T,$$

are adapted to the filtration $\mathcal{F}_t$, i.e., $\xi_t(\omega)$ is $\mathcal{F}_t$-measurable, $t = 1,\dots,T$. Then the nested formulation of the corresponding multistage risk optimization problem can be written as follows

$$\begin{aligned} \operatorname*{Min}_{x_1 \in G_1} \Big(c_1\cdot x_1 + \rho_{\mathcal{Z}_2|\mathcal{Z}_1}\Big[\inf\nolimits_{x_2 \in \mathcal{G}_2(x_1,\omega)} \big(c_2(\omega)\cdot x_2 + \cdots \\ + \rho_{\mathcal{Z}_{T-1}|\mathcal{Z}_{T-2}}\big[\inf\nolimits_{x_{T-1} \in \mathcal{G}_2(x_{T-2},\omega)} c_{T-1}(\omega)\cdot x_{T-1} \\ + \rho_{\mathcal{Z}_T|\mathcal{Z}_{T-1}}[\inf\nolimits_{x_T \in \mathcal{G}_T(x_{T-1},\omega)} c_T(\omega)\cdot x_T]\big]\big)\Big]\Big). \end{aligned} \tag{103}$$

The precise meaning of the nested formulation of problem (100) is explained by dynamic programming equations as follows. Define the (cost-to-go) function

[25] In order to avoid notational confusion we denote here by $a\cdot b$ the standard scalar product of two vectors $a, b \in \mathbb{R}^n$.

$$\mathcal{Q}_T(x_{T-1},\omega) := \left[\rho_{\mathcal{Z}_T|\mathcal{Z}_{T-1}}\big(V_T(x_{T-1})\big)\right](\omega), \tag{104}$$

where

$$[V_T(x_{T-1})](\omega) := \inf_{x_T\in\mathcal{G}_T(x_{T-1},\omega)} f_T(x_T,\omega). \tag{105}$$

And so on for $t = T-1,\dots,2$,

$$\mathcal{Q}_t(x_{t-1},\omega) := \left[\rho_{\mathcal{Z}_t|\mathcal{Z}_{t-1}}\big(V_t(x_{t-1})\big)\right](\omega), \tag{106}$$

where

$$\big[V_t(x_{t-1})\big](\omega) := \inf_{x_t\in\mathcal{G}_t(x_{t-1},\omega)} \Big\{ f_t(x_t,\omega) + \mathcal{Q}_{t+1}(x_t,\omega)\Big\}. \tag{107}$$

Of course, equations (106) and (107) can be combined into one equation:

$$\big[V_t(x_{t-1})\big](\omega) = \inf_{x_t\in\mathcal{G}_t(x_{t-1},\omega)} \Big\{ f_t(x_t,\omega) + \left[\rho_{\mathcal{Z}_{t+1}|\mathcal{Z}_t}\big(V_{t+1}(x_t)\big)\right](\omega)\Big\}. \tag{108}$$

Finally, at the first stage we solve the problem

$$\inf_{x_1\in G_1} \mathcal{Q}_2(x_1). \tag{109}$$

The optimal value and the set of optimal solutions of problem (109) provide the optimal value and the first-stage set of optimal solutions of the multistage program (100).

It should be mentioned that for the dynamic programming equations (108) to be well defined we need to ensure that $V_t(x_{t-1}) \in \mathcal{Z}_t$ for every considered x_{t-1}. Note that since the function $f_T(x_t,\omega)$ is $\mathcal{F}_T$-random lower semicontinuous and $\mathcal{G}_T(x_{T-1},\cdot)$ is closed valued and $\mathcal{F}_T$-measurable, it follows that $[V_T(x_{T-1})](\cdot)$ is $\mathcal{F}_T$-measurable (e.g., [19, Theorem 14.37]). Still one has to verify that $V_T(x_{T-1}) \in \mathcal{Z}_T$ in order for $\mathcal{Q}_T(x_{T-1},\omega)$ to be well defined. It will follow then that $\mathcal{Q}_T(x_{T-1},\cdot)$ is $\mathcal{F}_{t-1}$-measurable. In order to continue the process for $t = T-1$, it should be verified further that the function $\mathcal{Q}_T(x_{T-1},\omega)$ is $\mathcal{F}_{T-1}$-random lower semicontinuous. And so on for $t = T-2,\dots,2$. Finally, for $t = 2$ the function $\mathcal{Q}_2(x_1,\cdot)$ is $\mathcal{F}_1$-measurable, and hence does not depend on ω. Let us emphasize that the key assumption ensuring equivalence of the two formulations of the risk optimization problem is the monotonicity condition (M2) (cf., [24]).

Remark 6. In some cases the function $[V_T(\cdot)](\omega)$, where V_T is defined in (105), is convex for all $\omega \in \Omega$. This happens, for example, if $f_T(\cdot,\omega)$ is convex for all $\omega \in \Omega$, and $\mathcal{G}_T$ is defined by linear constraints of form (102). If $[V_T(\cdot)](\omega)$ is convex, then conditions (M1) and (M2) ensure that the corresponding function $\mathcal{Q}_T(\cdot,\omega)$ is also convex. Similarly, the convexity property propagates to the functions $\mathcal{Q}_t(\cdot,\omega)$, $t = T-1,\dots,2$. In particular, in the linear case, where f_t and $\mathcal{G}_t$ are defined in (102), the functions $\mathcal{Q}_t(\cdot,\omega)$, $t = T,\dots,2$, are convex for all $\omega \in \Omega$. In convex cases it makes sense to talk about subdifferentials[26] $\partial \mathcal{Q}_t(x_{t-1}, \omega)$. In principle, these subdifferentials can be written in a recursive form by using equations (106) and (107) and the analysis of Section 6.

13 Examples of Risk Mappings and Multistage Problems

In this section we adopt the framework of Sections 11 and 12 with $\mathcal{Z}_t := \mathcal{L}_p(\Omega, \mathcal{F}_t, P)$ and $\mathcal{Z}_t^* := \mathcal{L}_q(\Omega, \mathcal{F}_t, P)$, $t = 1, \ldots, T$. As before, unless stated otherwise, all expectations and probability statements are made with respect to the probability measure P. As it was already mentioned in Section 11, the conditional expectation

$$\rho_{\mathcal{Z}_t|\mathcal{Z}_{t-1}}(Z_t) := \mathbb{E}[Z_t|\mathcal{F}_{t-1}]$$

provides a (relatively simple) example of a conditional risk mapping. For that choice of conditional risk mappings, we have that

$$\big(\rho_{\mathcal{Z}_{t-1}|\mathcal{Z}_{t-2}} \circ \rho_{\mathcal{Z}_t|\mathcal{Z}_{t-1}}\big)(Z_t) = \mathbb{E}[Z_t|\mathcal{F}_{t-2}],$$

and $\rho_T(\cdot) = \mathbb{E}[\,\cdot\,]$. In that case (101) becomes the standard formulation of a multistage stochastic programming problem and (104)–(109) represent well known dynamic programming equations.

Now let us discuss analogues of some examples of risk functions considered Section 4.

Example 12. Consider the following extension of the mean-upper-semideviation risk function (of order $p \in [1, +\infty)$) discussed in Example 3. For $Z_t \in \mathcal{Z}_t$ and $c_t \geq 0$ define

$$\rho_{\mathcal{Z}_t|\mathcal{Z}_{t-1}}(Z_t) := \mathbb{E}[Z_t|\mathcal{F}_{t-1}] + c_t\, \sigma_p(Z_t|\mathcal{F}_{t-1}), \tag{110}$$

where

$$\sigma_p(Z_t|\mathcal{F}_{t-1}) := \Big(\mathbb{E}\Big[\big[Z_t - \mathbb{E}[Z_t|\mathcal{F}_{t-1}]\big]_+^p \big|\mathcal{F}_{t-1}\Big]\Big)^{1/p}. \tag{111}$$

If the sigma algebra $\mathcal{F}_{t-1}$ is trivial, then $\mathbb{E}[\,\cdot\,|\mathcal{F}_{t-1}] = \mathbb{E}[\,\cdot\,]$ and $\sigma_p(Z_t|\mathcal{F}_{t-1})$ becomes the upper semideviation of Z_t of order p. For a while we keep t fixed and we use the notation ρ for the above mapping $\rho_{\mathcal{Z}_t|\mathcal{Z}_{t-1}}$.

By using the analysis of Example 3 it is possible to show that ρ satisfies conditions (M1),(M3) and (M4), and also condition (M2), provided that $c_t \in [0, 1]$. Indeed, clearly ρ is positively homogeneous, i.e., condition (M4) holds. Condition (M3) can be verified directly. That is, if $Y \in \mathcal{Z}_{t-1}$ and $Z_t \in \mathcal{Z}_t$, then

[26] These subdifferentials are taken with respect to x_{t-1} for a fixed value $\omega \in \Omega$.

$$
\begin{aligned}
\rho(Z_t+Y) &= \mathbb{E}[Z_t+Y|\mathcal{F}_{t-1}] + c_t\Big(\mathbb{E}\Big[\big(Z_t+Y-\mathbb{E}[Z_t+Y|\mathcal{F}_{t-1}]\big)_+^p \,\big|\, \mathcal{F}_{t-1}\Big]\Big)^{1/p}\\
&= \mathbb{E}[Z_t|\mathcal{F}_{t-1}] + Y + c_t\Big(\mathbb{E}\Big[\big(Z_t-\mathbb{E}[Z_t|\mathcal{F}_{t-1}]\big)_+^p \,\big|\, \mathcal{F}_{t-1}\Big]\Big)^{1/p}\\
&= \rho(Z_t)+Y.
\end{aligned}
$$

For $\omega \in \Omega$ consider the function $\rho_\omega(\cdot) = [\rho(\cdot)](\omega)$. Consider also the conditional probability of P with respect to $\mathcal{F}_{t-1}$, denoted $\mu(\omega)$ or μ_ω (see Remark 3). Recall that $\mathbb{E}[Z_t|\mathcal{F}_{t-1}](\omega) = \mathbb{E}_{\mu_\omega}[Z_t]$, and hence

$$
\rho_\omega(Z_t) = \mathbb{E}_{\mu_\omega}[Z_t] + c_t\Big(\mathbb{E}_{\mu_\omega}\Big[\big(Z_t-\mathbb{E}_{\mu_\omega}[Z_t]\big)_+^p\Big]\Big)^{1/p}. \tag{112}
$$

For a fixed ω the function ρ_ω coincides with the risk function analyzed in Example 3 with μ_ω playing the role of the corresponding probability measure. Consequently, ρ_ω is convex, i.e., condition (M1) holds, and condition (M2) follows, provided that $c_t \in [0,1]$.

We have that $\mu_\omega \in \mathcal{P}_{\mathcal{Z}_t^*|\mathcal{F}_{t-1}}(\omega)$ and its conditional probability density $g_\omega = d\mu_\omega/dP$ has the following properties: $g_\omega \in \mathcal{Z}_t^*$, $g_\omega \geq 0$, for any $A \in \mathcal{F}_t$, the function $\omega \mapsto \int_A g_\omega(\tilde\omega)\,dP(\tilde\omega)$ is $\mathcal{F}_{t-1}$-measurable and, moreover, for any $B \in \mathcal{F}_{t-1}$ it holds that

$$
\int_B \int_A g_\omega(\tilde\omega)\,dP(\tilde\omega)\,dP(\omega) = P(A\cap B).
$$

By the analysis of Example 3 it follows that the representation

$$
\rho_\omega(Z_t) = \sup_{\zeta_t \in \mathcal{A}_t(\omega)} \mathbb{E}[\zeta_t Z_t], \tag{113}
$$

holds with

$$
\mathcal{A}_t(\omega) = \big\{\zeta_t' \in \mathcal{Z}_t^* : \zeta_t' = g_\omega\big(1+\zeta_t - \mathbb{E}[g_\omega \zeta_t]\big),\ \|\zeta_t\|_q \leq c_t,\ \zeta_t \succeq 0\big\}. \tag{114}
$$

In order to write the corresponding multistage problem in form (100) we need to describe the composite function $\tilde\rho$ defined in (96). In general a description of $\tilde\rho$ is quite messy. Let us consider the following two particular cases. Suppose that $p=1$ and all c_t are zero except one, say c_T. That is, $\rho_{\mathcal{Z}_t|\mathcal{Z}_{t-1}}(\cdot) := \mathbb{E}[\,\cdot\,|\mathcal{F}_{t-1}]$, for $t=2,\dots,T-1$. In that case

$$
\begin{aligned}
&\rho_{\mathcal{Z}_{T-1}|\mathcal{Z}_{T-2}}\big[Z_{T-1} + \rho_{\mathcal{Z}_T|\mathcal{Z}_{T-1}}(Z_T)\big] = \mathbb{E}\big[Z_{T-1} + \rho_{\mathcal{Z}_T|\mathcal{Z}_{T-1}}(Z_T)\big|\mathcal{F}_{T-2}\big]\\
&\quad = \mathbb{E}\big[Z_{T-1}+Z_T\big|\mathcal{F}_{T-2}\big] + c_T\mathbb{E}\big[\big(Z_T - \mathbb{E}[Z_T|\mathcal{F}_{T-1}]\big)_+\big|\mathcal{F}_{T-2}\big],
\end{aligned}
$$

and

$$
\tilde\rho(Z) = \mathbb{E}\Big[Z_1+\cdots+Z_T + c_T\big[Z_T - \mathbb{E}[Z_T|\mathcal{F}_{T-1}]\big]_+\Big]. \tag{115}
$$

Another case where calculations are simplified considerably is under the "between stages independence" condition (compare with Remark 5). That is, suppose that the objective functions f_t and the constraint mappings $\mathcal{G}_t$,

$t = 2, \dots, T$, are given in the form $f_t(x_t, \xi_t(\omega))$ and $\mathcal{G}_t(x_{t-1}, \xi_t(\omega))$, respectively, where $\xi_t(\omega)$ are random vectors defined on a probability space $(\Omega, \mathcal{F}, P)$. That is the case, for example, if f_t and $\mathcal{G}_t$ are defined in the form (102). With some abuse of notation we simply write $f_t(x_t, \xi_t)$ and $\mathcal{G}_t(x_{t-1}, \xi_t)$ for the corresponding random functions and mappings. It will be clear from the context when ξ_t is viewed as a random vector and when as its particular realization.

Assume that sigma algebra $\mathcal{F}_t$ is generated by $(\xi_1(\omega), \dots, \xi_t(\omega))$, $t = 1, \dots, T$. Assume also that ξ_1 is not random, and hence the sigma algebra $\mathcal{F}_1$ is trivial. Assume further the following condition, referred to as the *between stages independence* condition:

(I) For every $t \in \{2, \dots, T\}$, random vector ξ_t is (stochastically) independent of $(\xi_1, \dots, \xi_{t-1})$.

Then the minimum in the right hand side of (105) is a function of x_{T-1} and ξ_T, and hence is independent of the random vector $(\xi_1, \dots, \xi_{T-1})$. It follows then that the corresponding cost-to-go function $\mathcal{Q}_T(x_{T-1})$, defined in (105), is independent of ω. By continuing this process backwards we obtain that, under the between stages independence condition, the cost-to-go functions are independent of ω and the corresponding dynamic programming equations can be written in the form

$$\mathcal{Q}_t(x_{t-1}) = \rho_{\mathcal{Z}_t|\mathcal{Z}_{t-1}}(V_t(x_{t-1})), \tag{116}$$

$$V_t(x_{t-1})(\xi_t) = \inf_{x_t \in \mathcal{G}_t(x_{t-1}, \xi_t)} \big\{ f_t(x_t, \xi_t) + \mathcal{Q}_{t+1}(x_t) \big\}, \tag{117}$$

with

$$\rho_{\mathcal{Z}_t|\mathcal{Z}_{t-1}}(V_t(x_{t-1})) = \mathbb{E}\big[V_t(x_{t-1})\big] + c_t \left(\mathbb{E}\Big[\big(V_t(x_{t-1}) - \mathbb{E}[V_t(x_{t-1})]\big)_+^p \Big] \right)^{1/p}.$$

Also in that case the optimization in problem (100) should be performed over functions $X_t(\xi_t)$ and (compare with (99))

$$\begin{aligned} &\rho_T(F_1(X_1) + F_2(X_2) + \cdots + F_T(X_T)) = \\ &\qquad F_1(X_1) + \rho_{\mathcal{Z}_2|\mathcal{Z}_1}(F_2(X_2)) + \cdots + \rho_{\mathcal{Z}_T|\mathcal{Z}_{T-1}}(F_T(X_T)). \end{aligned} \tag{118}$$

Example 13. Consider the framework of Example 6. Let $v : \mathbb{R} \to \mathbb{R}$ be a convex real valued function such that the function $z + v(z)$ is monotonically nondecreasing on $\mathbb{R}$. Define

$$\big[\rho_{\mathcal{Z}_t|\mathcal{Z}_{t-1}}(Z_t)\big](\omega) := \inf_{Y \in \mathcal{Z}_{t-1}} \mathbb{E}\big[Z_t + v(Z_t - Y)|\mathcal{F}_{t-1}\big](\omega). \tag{119}$$

Of course, a certain care should be exercised in verification that the right hand side of equation (119) gives a well defined mapping. For a while we will keep t fixed and use notation $\rho = \rho_{\mathcal{Z}_t|\mathcal{Z}_{t-1}}$. Since the function $(Z_t, Y) \mapsto \mathbb{E}\big[Z_t + v(Z_t - Y)|\mathcal{F}_{t-1}\big](\omega)$ is convex, it follows that $\rho_\omega(\cdot)$ is convex, i.e., the condition (M1) holds. Since $z + v(z)$ is nondecreasing, condition (M2) holds as well. It is also straightforward to verify that condition (M3) holds here by making change of variables $Z_t \mapsto Z_t - Y$. Let us calculate the conjugate function ρ^*_ω. In a way similar to (30) we have for $\zeta_t \in \mathcal{Z}^*_t$,

$$\begin{aligned}\rho^*_\omega(\zeta_t) &= \sup_{Z_t \in \mathcal{Z}_t} \big\{\mathbb{E}[\zeta_t Z_t] - \rho_\omega(Z_t)\big\}\\ &= \sup_{Z_t \in \mathcal{Z}_t} \big\{\mathbb{E}[\zeta_t Z_t] + \sup\nolimits_{Y \in \mathcal{Z}_{t-1}} \mathbb{E}\big[-Z_t - v(Z_t - Y)|\mathcal{F}_{t-1}\big](\omega)\big\}\\ &= \sup_{Z_t \in \mathcal{Z}_t} \big\{\mathbb{E}[\zeta_t (Z_t + Y)] + \sup\nolimits_{Y \in \mathcal{Z}_{t-1}} \mathbb{E}\big[-Z_t - Y - v(Z_t)|\mathcal{F}_{t-1}\big](\omega)\big\},\end{aligned}$$

and hence

$$\begin{aligned}\rho^*_\omega(\zeta_t) = &\sup_{Z_t \in \mathcal{Z}_t} \big\{\mathbb{E}[\zeta_t Z_t] - \mathbb{E}\big[Z_t + v(Z_t)|\mathcal{F}_{t-1}\big](\omega)\big\}\\ &+ \sup_{Y \in \mathcal{Z}_{t-1}} \mathbb{E}\big[Y(\zeta_t - 1)|\mathcal{F}_{t-1}\big](\omega).\end{aligned} \tag{120}$$

Since $Y \in \mathcal{Z}_{t-1}$, and hence $Y(\omega)$ is $\mathcal{F}_{t-1}$-measurable, we have

$$\mathbb{E}\big[Y(\zeta_t - 1)|\mathcal{F}_{t-1}\big](\omega) = Y(\omega)\big(\mathbb{E}[\zeta_t|\mathcal{F}_{t-1}](\omega) - 1\big).$$

Therefore, the second maximum in the right hand side of (120) is equal to zero if $\mathbb{E}[\zeta_t|\mathcal{F}_{t-1}](\omega) = 1$, and to $+\infty$ otherwise. It follows that the domain of ρ^*_ω is included (this inclusion can be strict) in the set

$$\mathcal{A}^*_t(\omega) := \big\{\zeta_t \in \mathcal{Z}^*_t : \mathbb{E}[\zeta_t|\mathcal{F}_{t-1}](\omega) = 1\big\}. \tag{121}$$

Note that for any $B \in \mathcal{F}_{t-1}$ and $\zeta_t \in \mathcal{A}^*_t(\omega)$ it holds that $\int_B \zeta_t \, dP$ is equal to 1 if $\omega \in B$, and to 0 if $\omega \notin B$, i.e., $\mathcal{A}^*_t(\omega)$ is a subset of $\mathcal{P}_{\mathcal{Z}^*_t|\mathcal{F}_{t-1}}(\omega)$.

Consider the conditional probability of P with respect to $\mathcal{F}_{t-1}$, denoted $\mu(\omega)$ or μ_ω (see Remark 3). We have that $\mu_\omega \in \mathcal{P}_{\mathcal{Z}^*_t|\mathcal{F}_{t-1}}(\omega)$ and let $g_\omega = d\mu_\omega/dP$ be its conditional probability density (properties of g_ω were discussed in the previous example). Recall that $\mathbb{E}[Z_t|\mathcal{F}_{t-1}](\omega) = \mathbb{E}_{\mu_\omega}[Z_t]$, and hence

$$\mathbb{E}\big[Z_t + v(Z_t)|\mathcal{F}_{t-1}\big](\omega) = \mathbb{E}[g_\omega(Z_t + v(Z_t))].$$

By using this and since by the interchangeability formula the maximum over Z_t at the right hand side of (120) can be taken inside the integral, we obtain

$$\rho^*_\omega(\zeta_t) = \begin{cases} \mathbb{E}\left[\sup_{z_t \in \mathbb{R}} \big\{(\zeta_t - g_\omega)z_t - g_\omega v(z_t)\big\}\right], & \text{if } \zeta_t \in \mathcal{A}^*_t(\omega),\\ +\infty, & \text{otherwise.}\end{cases} \tag{122}$$

By Theorem 5 we have then that

$$\rho_\omega(Z_t) = \sup_{\zeta_t \in \mathcal{A}^*_t(\omega)} \big\{\mathbb{E}[\zeta_t Z_t] - \rho^*_\omega(\zeta_t)\big\}. \tag{123}$$

In particular, let $\mathcal{Z}_t := \mathcal{L}_1(\Omega, \mathcal{F}_t, P)$, $\mathcal{Z}^*_t := \mathcal{L}_\infty(\Omega, \mathcal{F}_t, P)$ and take

$$v(z) := \varepsilon_1[-z]_+ + \varepsilon_2[z]_+,$$

where $\varepsilon_1 \in [0,1]$ and $\varepsilon_2 \geq 0$ (compare with Example 7). This function $v(z)$ is convex positively homogeneous, and the corresponding function $z + v(z)$ is nondecreasing. The maximum inside the expectation in the right hand side of (122) is equal to zero if $-\varepsilon_1 g_\omega \leq \zeta_t - g_\omega \leq g_\omega \varepsilon_2$, and to $+\infty$ otherwise. It follows that the corresponding risk mapping ρ satisfies conditions (M1)–(M4), and

$$\rho_\omega(Z_t) = \sup_{\zeta_t \in \mathcal{A}_t(\omega)} \mathbb{E}[\zeta_t Z_t], \tag{124}$$

where $\eta_1 := 1 - \varepsilon_1$, $\eta_2 := 1 + \varepsilon_2$,

$$\mathcal{A}_t(\omega) = \left\{ \zeta_t \in \mathcal{Z}_t^* : \begin{array}{l} \eta_1 g_\omega(\tilde{\omega}) \leq \zeta_t(\tilde{\omega}) \leq \eta_2 g_\omega(\tilde{\omega}),\ a.e.\ \tilde{\omega} \in \Omega, \\ \mathbb{E}[\zeta_t | \mathcal{F}_{t-1}](\omega) = 1 \end{array} \right\}. \tag{125}$$

The between stages independence condition can be introduced here in a way similar to the previous example. Under this condition formulas (116), (117) and (118) will hold here as well.

References

1. C. Acerbi and D. Tasche, On the coherence of expected shortfall, *Journal of Banking and Finance*, 26 (2002), 1487–1503.
2. P. Artzner, F. Delbaen, J.-M. Eber, and D. Heath, Coherent measures of risk, *Mathematical Finance*, 9 (1999), 203–228.
3. P. Artzner, F. Delbaen, J.-M. Eber, D. Heath and H. Ku, Coherent multiperiod risk measurement, Manuscript, ETH Zürich, 2003.
4. A. Eichhorn, W. Römisch. Polyhedral risk measures in stochastic programming, Stochastic Programming E-Print Series, 2004.
5. H. Föllmer and A. Schied, Convex measures of risk and trading constraints, *Finance and Stochastics*, 6 (2002), 429–447.
6. W.K. Klein Haneveld, *Duality in Stochastic Linear and Dynamic Programming*, Lecture Notes in Economics and Mathematical Systems 274, Springer-Verlag, New York, 1986.
7. M.O. Lorenz, Methods of measuring concentration of wealth, *Journal of the American Statistical Association*, 9 (1905), 209–219.
8. H.M. Markowitz, Portfolio selection, *Journal of Finance*, 7 (1952), 77–91.
9. H.M. Markowitz, *Mean–Variance Analysis in Portfolio Choice and Capital Markets*, Blackwell, Oxford, 1987.
10. A. Müller and D. Stoyan, *Comparison Methods for Stochastic Models and Risks*, Wiley, Chichester, 2002.
11. W. Ogryczak and A. Ruszczyński, From stochastic dominance to mean–risk models: semideviations as risk measures, *European Journal of Operational Research*, 116 (1999), 33–50.
12. W. Ogryczak and A. Ruszczyński, On consistency of stochastic dominance and mean–semideviation models, *Mathematical Programming*, 89 (2001), 217–232.

13. W. Ogryczak and A. Ruszczyński, Dual stochastic dominance and related mean–risk models, *SIAM Journal on Optimization*, 13 (2002), 60–78.
14. G. Pflug and A. Ruszczyński, A risk measure for income processes, in: G. Szegö (Ed.), *Risk Measures for the 21st Century*, John Wiley & Sons, 2004.
15. A. Prékopa, *Stochastic Programming*, Kluwer, Dordrecht, Boston, 1995.
16. F. Riedel, Dynamic coherent risk measures, Working Paper 03004, Department of Economics, Stanford University, 2003.
17. R.T. Rockafellar, *Conjugate Duality and Optimization*, Regional Conference Series in Applied Mathematics, SIAM, Philadelphia, 1974.
18. R.T. Rockafellar and R.J-B. Wets, Stochastic convex programming: basic duality, *Pacific J. Math.*, 62 (1976), 173–195.
19. R.T. Rockafellar and R.J-B. Wets, *Variational Analysis*, Springer-Verlag, Berlin, 1998.
20. R.T. Rockafellar and S.P. Uryasev, Optimization of conditional value-at-risk, *The Journal of Risk*, 2 (2000), 21-41.
21. R.T. Rockafellar, S. Uryasev and M. Zabarankin, Deviation measures in risk analysis and optimization, Research Report 2002-7, Department of Industrial and Systems Engineering, University of Florida.
22. A. Ruszczyński and A. Shapiro, Eds., *Stochastic Programming*, in: Handbooks in OR & MS, Vol. 10, North-Holland Publishing Company, Amsterdam, 2003.
23. A. Ruszczyński and A. Shapiro, Optimization of convex risk functions, *E-print available at:* `http://www.optimization-online.org`, 2004.
24. A. Ruszczyński and A. Shapiro, Conditional risk mappings, *E-print available at:* `http://www.optimization-online.org`, 2004.
25. J. von Neumann and O. Morgenstern, *Theory of Games and Economic Behavior*, Princeton University Press, Princeton, 1947.
26. J. Žáčková, On minimax solutions of stochastic linear programming problems, *Čas. Pěst. Mat.*, 91 (1966), 423-430.