Surveys in Operations Research and Management Science 16 (2011) 49–66

Contents lists available at ScienceDirect

Surveys in Operations Research and Management Science

journal homepage: www.elsevier.com/locate/sorms

Review

Modeling and optimization of risk

Pavlo Krokhmal [a], Michael Zabarankin [b], Stan Uryasev [c,*]

[a] Department of Mechanical and Industrial Engineering, University of Iowa, Iowa City, IA 52242, United States
[b] Department of Mathematics, Stevens Institute of Technology, Hoboken, NJ 07030, United States
[c] Risk Management and Financial Engineering Lab, Department of Industrial and Systems Engineering, University of Florida, Gainesville, FL 32611, United States

ARTICLE INFO

Article history:
Received 2 August 2010
Accepted 8 August 2010

ABSTRACT

This paper surveys the most recent advances in the context of decision making under uncertainty, with an emphasis on the modeling of risk-averse preferences using the apparatus of axiomatically defined risk functionals, such as coherent measures of risk and deviation measures, and their connection to utility theory, stochastic dominance, and other more established methods.

Contents

1. Introduction

Decision making and optimization under uncertainty constitute a broad and popular area of operations research and management sciences. Various approaches to modeling of uncertainty are seen in such fields as stochastic programming, simulation, theory of stochastic processes, etc. This survey presents an account of the recent advances in decision making under uncertainty, and, specifically, the methods for modeling and control of risk in the context of their relation to mathematical programming models for dealing with uncertainties, which are broadly classified as stochastic programming methods.

To illustrate the issues pertinent to modeling of uncertainties and risk in the mathematical programming framework, it is instructive to start in the deterministic setting, where a typical decision making or design problem can be formulated in the form

$$\begin{aligned} \max_{\mathbf{x}\in\mathcal{S}} \quad & f(\mathbf{x}) \\ \text{subject to} \quad & g_i(\mathbf{x}) \le 0, \quad i = 1, \dots, k, \end{aligned} \tag{1}$$

with $\mathbf{x}$ being the decision or design vector from $\mathbb{R}^n$ or $\mathbb{Z}^n$. Uncertainty, usually described by a random element $\boldsymbol{\xi}$, leads to

* Corresponding author.
E-mail addresses: krokhmal@engineering.uiowa.edu (P. Krokhmal), Michael.Zabarankin@stevens.edu (M. Zabarankin), uryasev@ufl.edu (S. Uryasev).

1876-7354/$ – see front matter
doi:10.1016/j.sorms.2010.08.001

situations where instead of just $f(\mathbf{x})$ and $g_i(\mathbf{x})$ one has to deal with $f(\mathbf{x}, \boldsymbol{\xi})$ and $g_i(\mathbf{x}, \boldsymbol{\xi})$ (herein the set $\mathcal{S}$ is reserved to representing the deterministic requirements on the decision vector $\mathbf{x}$ that are not affected by uncertainty, such as nonnegativity constraints, etc.). Often it is appropriate to think of $\boldsymbol{\xi}$ as being governed by a probability distribution that is known or can be estimated.

A serious difficulty, however, is that the decision $\mathbf{x}$ must be chosen before the outcome from this distribution can be observed. One cannot then simply replace $f(\mathbf{x})$ by $f(\mathbf{x}, \boldsymbol{\xi})$ in (1), because a choice of $\mathbf{x}$ only produces a random variable $X = f(\mathbf{x}, \boldsymbol{\xi})$ whose realization is not yet known, and it is difficult to make sense of "minimizing a random variable" as such. Likewise, $g_i(\mathbf{x})$ cannot just be replaced by $g_i(\mathbf{x}, \boldsymbol{\xi})$ in (1), at least not without some careful thinking or elaboration. Over the years, a number of approaches have been developed to address these issues; a familiar and commonly used approach is to replace functions $f(\mathbf{x}, \boldsymbol{\xi})$ and $g_i(\mathbf{x}, \boldsymbol{\xi})$ with their expected values, e.g.,

$$f(\mathbf{x}, \boldsymbol{\xi}) \to E_{\boldsymbol{\xi}}[f(\mathbf{x}, \boldsymbol{\xi})].$$

Being intuitively appealing and numerically efficient, this generic method has its limitations, which have long been recognized in literature (see, for example, [1]). In particular, replacing a random objective function with its expected value implies that (i) the decision obtained as a solution of the stochastic programming problem will be employed repeatedly under identical or similar conditions (also known as the "long run" assumption); and (ii) the variability in realizations of the random value $f(\mathbf{x}, \boldsymbol{\xi})$ is not significant. As it poses no difficulty to envisage situations when these two assumptions do not hold, a work-around has to be devised that will allow for coping with models that do not comply with (i) and (ii).

A rather general remedy is to bring the concept of *risk* into the picture, with "risk" broadly defined as a quantitative expression of a system of attitudes, or *preferences* with respect to a set of random outcomes. This general idea has been omnipresent in the field of decision making for quite a long time, tracing as far back as 1738, when Daniel Bernoulli has introduced the concept of utility function (symptomatically, the title of Bernoulli's paper [2] paper translates from Latin as "Exposition of a New Theory on the Measurement of Risk"). Bernoulli's idea represents an integral part of the utility theory of theory of von Neumann and Morgenstern [3], one of the most dominant mathematical paradigms of modern decision making science. Another approach, particularly popular in the investment science, is the Markowitz mean–variance framework that identifies risk with the volatility (variance) of the random outcome of the decision [4].

In this paper, we survey the major developments that stem from these two fundamental approaches, with an emphasis on recent advances associated with measurement and control of risks via the formalism of *risk measures*, and their relation to mathematical programming methods, and, particularly, the stochastic programming framework.

Let us introduce some notations that will be used throughout the paper. The random element $X = X(\mathbf{x}, \omega)$, which depends on the decision vector $\mathbf{x}$ as well as on some random event $\omega \in \Omega$, will denote some performance measure of the decision $\mathbf{x}$ under uncertainty. In relation to the example used in the beginning of this section, the random element X may be taken as $X = f(\mathbf{x}, \boldsymbol{\xi}(\omega))$, where $\boldsymbol{\xi}(\omega)$ is a vector of uncertain (random) parameters. In general, the random quantity $X(\mathbf{x}, \omega)$ will be regarded as a *payoff*, or *profit* function, in the sense that the higher values of X are preferred, while its lower-value realizations must be avoided. This convention is traditional to the risk management literature, which is historically rooted in economic and financial applications.[1] It is also customary to assume that the profit function $X(\mathbf{x}, \omega)$ is concave in the decision vector $\mathbf{x}$, over some appropriate (convex) feasible set of decisions, which facilitates formulation of well-behaved convex mathematical programming models.

In the cases when more formality is required, we will consider X to be an outcome from some probability space $(\Omega, \mathcal{F}, \mathbb{P})$, where Ω is a set of random events, $\mathcal{F}$ is a sigma-algebra, and $\mathbb{P}$ is a probability measure, which belongs to a linear space $\mathcal{X}$ of $\mathcal{F}$-measurable functions $X : \Omega \mapsto \mathbb{R}$. For the purposes of this work, in most cases (unless noted otherwise) it suffices to take $\mathcal{X} = \mathcal{L}^\infty(\Omega, \mathcal{F}, \mathbb{P})$, a space of all bounded functions X, which also includes constants. To cast the corresponding results in the context of stochastic programming, we will follow the traditional method of modeling uncertainty in stochastic programming problems (see, e.g., [1,5–7]) by introducing a finite set of scenarios $\{\omega_1, \ldots, \omega_N\} \subseteq \Omega$, whereby each decision $\mathbf{x}$ results in a range of outcomes $X(\mathbf{x}, \omega_1), \ldots, X(\mathbf{x}, \omega_N)$ that have the respective probabilities $p_1, \ldots, p_N$, where $p_j = \mathbb{P}\{\omega_j\} \in (0, 1)$ and $\sum_{j=1}^N p_j = 1$.

Finally, we would like to mention that this review focuses mostly on models and approaches formulated in a "static", or single-period setting, and does not cover the corresponding "dynamic" or multi-period decision making and risk optimization methods.

In our exposition, we made an attempt to adhere to the historical timeline, whenever appropriate. In Section 2, we briefly recount the most important facts from the topics that are relatively more familiar to the general audience: the expected utility theory, stochastic dominance, Markowitz risk-reward framework, etc., along with some new developments, such as the stochastic dominance constraints. Section 3 discusses some of the most popular downside risk models and related concepts, including Value-at-Risk and probabilistic (chance) constraints. Section 4 deals with the topic of coherent measures of risk and some of the most prominent coherent measures, including the Conditional Value-at-Risk. Finally, Section 5 presents a comprehensive discussion of deviation measures of risk and related topics.

2. Utility theory, stochastic dominance, and risk-reward optimization paradigms

2.1. Utility theory and stochastic dominance

The von Neumann and Morgenstern [3] utility theory of choice under uncertainty represents one of the major pillars of modern decision making science, and plays a fundamental role in economics, finance, operations research, and other related fields (see, among others, [8–11]).

The von Neumann–Morgenstern utility theory argues that when the preference relation $\succeq$ of the decision maker satisfies certain axioms (completeness, transitivity, continuity, and independence), there exists a function $u : \mathbb{R} \mapsto \mathbb{R}$, such that an outcome X is preferred to outcome Y ("$X \succeq Y$") if and only if

$$E[u(X)] \geq E[u(Y)]. \tag{2}$$

Thus, in effect, a decision making problem under uncertainty for a *rational* decision maker reduces to maximization of his/her expected utility:

$$\max\{E[u(X)] \mid X \in \mathcal{X}\}.$$

[1] In engineering literature, the outcome X is often considered as a *cost*, or *loss* function, whose lower values are preferred; obviously, these two interpretations can be reconciled by replacing X with $-X$ and vice versa.

If the function u is non-decreasing and concave, the corresponding preference is said to be *risk averse*. In many applications, however, it is often difficult to obtain an explicit form of the utility function u.

The von Neumann–Morgenstern expected utility approach is closely related to the concepts of *stochastic dominance* [12–14]; see also an account of earlier works in [15]. Namely, a random outcome X is said to dominate outcome Y with respect to the first-order stochastic dominance (FSD) relation, $X \succeq_{(1)} Y$, if

$$\mathbb{P}\{X \le t\} \le \mathbb{P}\{Y \le t\}, \quad \text{or} \quad F_X(t) \le F_Y(t) \quad \text{for all } t \in \mathbb{R}, \tag{3}$$

where F_X and F_Y are the distribution functions of X and Y, respectively. Intuitively, FSD corresponds to the notion that X is preferred over Y if X assumes larger values than Y. The second-order stochastic dominance (SSD) relation is defined as

$$X \succeq_{(2)} Y \Leftrightarrow \int_{-\infty}^{t} F_X(\eta)\,\mathrm{d}\eta \le \int_{-\infty}^{t} F_Y(\eta)\,\mathrm{d}\eta \quad \text{for all } t \in \mathbb{R}, \tag{4}$$

and, in general, the kth order stochastic dominance (kSD) relation is stated in the form

$$X \succeq_{(k)} Y \Leftrightarrow F_X^{(k)}(t) \le F_Y^{(k)}(t) \quad \text{for all } t \in \mathbb{R}, \tag{5}$$

where $F^{(k)}(t)$ is the so-called kth degree distribution function defined recursively as

$$F_X^{(k)}(t) = \int_{-\infty}^{t} F_X^{(k-1)}(\eta)\,\mathrm{d}\eta, \qquad F_X^{(1)}(t) = F_X(t). \tag{6}$$

It follows from the above definition that $X \succeq_{(k-1)} Y$ entails that $X \succeq_{(k)} Y$, provided, of course, that $X, Y \in \mathcal{L}^{k-1}$. The corresponding strict stochastic dominance relations, $X \succ_{(k)} Y$, are defined by requiring that strict inequality in (5) holds for at least one $t \in \mathbb{R}$. For a comprehensive exposition of stochastic dominance, see [16].

Rothschild and Stiglitz [17] have bridged the von Neumann–Morgenstern utility theory with the stochastic dominance principles by showing that X dominating Y in the SSD sense, $X \succeq_{(2)} Y$, is equivalent to relation (2) holding true for all concave non-decreasing functions u; similarly, $X \succeq_{(1)} Y$ if and only if (2) holds for all non-decreasing utility functions u. Strict stochastic dominance means that relation (2) holds strictly for at least one such u.

The *dual utility theory*, also known as *rank-dependent expected utility theory*, was proposed in [18,19] and [20]. It is based on a system of axioms different from those of von Neumann and Morgenstern; in particular, it introduces an axiom dual to the von Neumann–Morgenstern independence axiom, which was brought to question in a number of studies that showed it being violated in actual decision making [21,22]. Then, it follows that a preference relation over uniformly bounded on [0, 1] outcomes satisfies these axioms if and only if there exists a non-decreasing function $v : [0,1] \mapsto [0,1]$, called *dual utility function*, such that $v(0) = 0$ and $v(1) = 1$, and which expresses preference $X \succeq Y$ in terms of *Choquet* integrals [23–25]:

$$\int_0^1 v(\bar{F}_X(t))\,\mathrm{d}t \ge \int_0^1 v(\bar{F}_Y(t))\,\mathrm{d}t. \tag{7}$$

Here, $\bar{F}_X(t)$ is the *decumulative distribution function*, $\bar{F}_X(t) = \mathbb{P}\{X > t\}$. Just as in the expected utility theory [3], the dual utility function v defines the degree of risk aversion of the decision maker; in particular, a concave increasing v introduces an ordering consistent with the second-order stochastic dominance [19].

The deep connections among the expected utility theory, stochastic dominance (particularly, SSD), and dual utility theory have been exploited in numerous developments pertinent to decision making under uncertainty and risk. One of the most recent advances in this context involves optimization problems with stochastic dominance constraints.

2.1.1. Stochastic dominance constraints

Recently, Dentcheva and Ruszczyński [26,27] have introduced optimization problems with stochastic dominance constraints

$$\max\{f(X) \mid X \succeq_{(k)} Y, X \in \mathcal{C}\}, \tag{8}$$

where $Y \in \mathcal{L}^{k-1}$ is a given reference (benchmark) outcome, the objective f is a concave functional on $\mathcal{X}$ and the feasible set $\mathcal{C}$ is convex. Of particular practical significance are the special cases of (8) with $k = 2$ and $k = 1$, corresponding to the second- and first-order stochastic dominance, respectively. Using the equivalent representation for second-order stochastic dominance relation (compare to (4)),

$$X \succeq_{(2)} Y \Leftrightarrow E[(X-\eta)_-] \le E[(Y-\eta)_-] \quad \text{for all } \eta \in \mathbb{R}, \tag{9}$$

where $X_\pm$ denotes the positive (negative) part of X:

$$X_\pm = \max\{0, \pm X\},$$

Dentcheva and Ruszczyński [26] considered the following relaxation of problem (8) with $k = 2$:

$$\begin{aligned} &\max\{f(X) \mid E[(X-\eta)_-] \le E[(Y-\eta)_-] \\ &\quad \text{for all } \eta \in [a,b], X \in \mathcal{C}\}, \end{aligned} \tag{10}$$

where the range of η was restricted to a compact interval $[a, b]$ in order to formulate constraint qualification conditions. In many practical applications, where the reference outcome Y has a discrete distribution over $\{y_1, \dots, y_m\} \subset [a, b]$, formulation (10) admits significant simplifications [26]:

$$\begin{aligned} &\max\{f(X) \mid E[(X-y_i)_-] \le E[(Y-y_i)_-], \\ &\quad i = 1, \dots, m, X \in \mathcal{C}\}. \end{aligned} \tag{11}$$

In the case when X has a discrete distribution $\mathbb{P}\{X = x_i\} = p_i$, $i = 1, \dots, N$, the m constraints in (11) can be represented via $O(Nm)$ linear inequalities by introducing Nm auxiliary variables $w_{ik} \ge 0$:

$$\begin{aligned} &\sum_{i=1}^{N} p_i w_{ik} \le \sum_{j=1}^{m} q_j (y_k - y_j)_+, \quad k = 1, \dots, m, \\ &w_{ik} + x_i \ge y_k, \quad i = 1, \dots, N, \; k = 1, \dots, m, \\ &w_{ik} \ge 0, \quad i = 1, \dots, N, \; k = 1, \dots, m, \end{aligned} \tag{12}$$

where $q_k = \mathbb{P}\{Y = y_k\}$, $k = 1, \dots, m$. In [28], a formulation of SSD constraints was suggested that also employed $O(Nm)$ variables but only $O(N + m)$ inequalities. A cutting plane scheme for SSD constraints (12) based on a cutting plane representation for integrated chance constraints (see Section 3.1) due to [29] was employed in [30].

Using the following characterization of second-order dominance via quantile functions,

$$X \succeq_{(2)} Y \Leftrightarrow F_{(-2)}(X,p) \ge F_{(-2)}(Y,p) \quad \text{for all } p \in [0,1], \tag{13}$$

where $F_{(-2)}(X, p)$ is the *absolute Lorentz function* [31],

$$\begin{aligned} &F_{(-2)}(X,p) = \int_0^p F_{(-1)}(X,t)\,\mathrm{d}t, \quad \text{where} \\ &\quad F_{(-1)}(X,p) = \inf\{\eta \mid \mathbb{P}\{X \le \eta\} \ge p\}, \end{aligned} \tag{14}$$

Dentcheva and Ruszczyński [32] introduced optimization under *inverse stochastic dominance constraints:*

$$\begin{aligned} &\max\{f(X) \mid F_{(-2)}(X,p) \ge F_{(-2)}(Y,p) \text{ for all} \\ &\quad p \in [\alpha, \beta] \subset (0,1), X \in \mathcal{C}\}. \end{aligned} \tag{15}$$

A relationship between (inverse) stochastic dominance constraints and certain class of risk functionals was established in [33], see also Section 4.1. Further extensions of (8)–(10) include nonlinear SSD constraints [34], *robust SSD constraints* where the SSD relation is considered over a set of probability measures [35]. Optimization problems of the form (8) with $k = 1$, corresponding to the (generally non-convex) first-order stochastic dominance constraints, were studied in [27], where it was shown that the SSD constraints can be considered as a convexification of the FSD constraint. Portfolio optimization with second-order stochastic dominance constraints has been considered in [36], see also [37].

2.2. Markowitz risk-reward optimization

The prominent result of Markowitz [4,38], who advocated identification of the portfolio's risk with the volatility (variance) of its returns, represents a cornerstone of the modern theory of risk management. Markowitz's work was also among the first that emphasized the optimizational aspect of risk management problems. In its traditional form, Markowitz's mean–variance (MV) model can be stated using the notations adopted above as the problem of minimization of risk expressed by the variance of decision's payoff $\sigma^2(X(\mathbf{x}, \omega))$ while requiring that the average payoff of the decision exceeds a predefined threshold r_0:

$$\min_{\mathbf{x}\in\mathcal{S}}\{\sigma^2(X(\mathbf{x}, \omega)) \mid E[X(\mathbf{x}, \omega)] \geq r_0\}, \tag{16}$$

where $\mathcal{S} \subset \mathbb{R}^n$ is the set of feasible decisions $\mathbf{x}$. Provided that the feasible set $\mathcal{S}$ is convex and $X(\mathbf{x}, \omega)$ is concave in $\mathbf{x}$ on $\mathcal{S}$, problem (16) is convex, and thus efficiently tractable. The computational tractability of the MV approach, along with its intuitively appealing interpretation, have contributed to widespread popularity of the decision making models of type (16) in finance and economics, as well as in operations research, management science, and engineering. For a survey of developments of the Markowitz MV theory, see, for instance, [39].

In a more general context, Markowitz's work led to formalization of the fundamental view that *a decision under uncertainties may be evaluated in terms of tradeoff between its risk and reward.* [2] Such an approach is different from the expected utility framework; in particular, an SSD efficient outcome is not generally efficient in the risk-reward sense as described below (the original Markowitz model is consistent with the second-order stochastic dominance in the special case when X is normally distributed).

Given a payoff (profit) function $X = X(\mathbf{x}, \omega)$ that is dependent on the decision vector $\mathbf{x}$ and random element $\omega \in \Omega$, let $\rho(X) = \rho(X(\mathbf{x}, \omega))$ represent the measure of risk, and $\pi(X) = \pi(X(\mathbf{x}, \omega))$ be the measure of performance, or reward associated with X. It is natural to presume the reward measure $\pi(X(\mathbf{x}, \omega))$ to be concave in $\mathbf{x}$ over some closed convex set of decisions $\mathcal{S} \subset \mathbb{R}^n$, and the risk measure $\rho(X(\mathbf{x}, \omega))$ to be convex over $\mathcal{S}$. Then, the risk-reward optimization problem generalizing the classical MV model can be formulated as finding the decision $\mathbf{x}$ whose risk is minimal under the condition that the reward exceeds a certain predefined level:

$$\min_{\mathbf{x}\in\mathcal{S}}\{\rho(X(\mathbf{x}, \omega)) \mid \pi(X(\mathbf{x}, \omega)) \geq \pi_0\}. \tag{17}$$

Alternatively, the following two formulations are frequently employed: select the decision $\mathbf{x}$ that maximizes the reward $\pi(\mathbf{x})$ while assuring that the risk does not exceed ρ_0:

$$\min_{\mathbf{x}\in\mathcal{S}}\{-\pi(X(\mathbf{x}, \omega)) \mid \rho(X(\mathbf{x}, \omega)) \leq \rho_0\}, \tag{18}$$

or a weighted combination of risk and reward is optimized:

$$\min_{\mathbf{x}\in\mathcal{S}}\{\rho(X(\mathbf{x}, \omega)) - \lambda\pi(X(\mathbf{x}, \omega)) \mid \lambda \geq 0\}. \tag{19}$$

[2] The term "risk" here has many interpretations; in the context of the original Markowitz's contribution it refers to a dispersion type of uncertainty, and a complementary interpretation refers to risk as a shortfall uncertainty. Both these interpretations are explored in detail in Sections 3–5, correspondingly.

In view of the risk-reward formulations (17)–(19), an outcome $X_1 = X(\mathbf{x}_1, \omega)$ is said to weakly (ρ, π)-dominate outcome $X_2 = X(\mathbf{x}_2, \omega)$, or $X_1 \succeq_{(\rho,\pi)} X_2$, if

$$\rho(X_1) \leq \rho(X_2) \quad \text{and} \quad \pi(X_1) \geq \pi(X_2).$$

Strong (ρ, π)-dominance, $X_1 \succ_{(\rho,\pi)} X_2$, implies that at least one of the inequalities above is strict. An outcome $X_1 = X(\mathbf{x}, \omega)$ corresponding to the decision $\mathbf{x}_1 \in \mathcal{S}$ is considered *efficient*, or (ρ, π)-efficient, if there is no $\mathbf{x}_2 \in \mathcal{S}$ such that $X_2 \succ_{(\rho,\pi)} X_1$, or, equivalently,

$$\rho(X_2) = \rho(X_1) \quad \text{and} \quad \pi(X_2) > \pi(X_1)$$

or

$$\pi(X_2) = \pi(X_1) \quad \text{and} \quad \rho(X_2) < \rho(X_1).$$

Then, the set

$$\mathcal{E} = \{(\rho, \pi) \mid \rho = \rho(X),\ \pi = \pi(X),$$
$$X = X(\mathbf{x}, \omega) \text{ is efficient}, \mathbf{x} \in \mathcal{S}\}$$

is called the *efficient frontier*. In the case when the sets $\{\mathbf{x} \in \mathcal{S} \mid \pi(X(\mathbf{x}, \omega)) \geq \pi_0\}$ and $\{\mathbf{x} \in \mathcal{S} \mid \rho(X(\mathbf{x}, \omega)) \leq \rho_0\}$ have internal points, problems (17)–(19) are equivalent in the sense that they generate the same efficient frontier via varying the parameters λ, ρ_0, and π_0 [40]. The equivalence between problems (17)–(19) is well known for mean–variance [39] and mean-regret [41] efficient frontiers.

Although the original Markowitz approach is still widely used today, it has been acknowledged that variance $\sigma^2(X)$ as a measure of risk in (17)–(19) does not always produce adequate estimates of risk exposure. Part of the criticism is due to the fact that variance $\sigma^2(X) = E\big[(X - E[X])^2\big]$ penalizes equally the "gains" $X > E[X]$ and "losses" $X < E[X]$. Secondly, variance has been found ineffective for measuring the risk of low-probability events. This led to development of the so-called *mean risk* models, where the reward measure in (17)–(19) is taken as the expected value of X, $\pi(X) = E[X]$, for some choice of risk measure ρ [42–44]. In particular, to circumvent the symmetric attitude of variance in (16), a number of the so-called *downside* risk measures have been considered in the literature. Next we outline the most notable developments in this area, including the semivariance risk models, lower partial moments, Value-at-Risk, etc.

Another major development of the classical Markowitz framework is associated with the recent advent of the *deviation measures* that generalize variance as a measure of risk in (16) and are discussed in detail in Section 5.

3. Downside risk measures and optimization models

3.1. Risk measures based on downside moments

The shortcomings of variance $\sigma^2(X)$ as a risk measure have been recognized as far back as by Markowitz himself, who proposed to use *semivariance* $\sigma_-^2(X)$ for a more accurate estimation of risk exposure [38]:

$$\sigma_-^2(X) = E[(X - E[X])_-^2] = \|(X - E[X])_-\|_2^2, \tag{20}$$

where $\|\cdot\|$ is the p-norm in $\mathcal{L}^p$, $p \in [1, \infty]$:

$$\|X\|_p = (E[|X|^p])^{1/p}. \tag{21}$$

Applications of semivariance risk models to decision making under uncertainty in the context of mean risk models have been studied in [43,44,31]. Namely, it was shown in [43] that the *mean risk* model that corresponds to (P3) with $\pi(X) = E[X]$ and $\rho(X) = \sigma_-(X)$ is SSD consistent for $\lambda = 1$, i.e.,

$$\begin{aligned} &X \succeq_{(2)} Y \Rightarrow \pi(X) \geq \pi(Y) \quad \text{and} \\ &\pi(X) - \lambda\rho(X) \geq \pi(Y) - \lambda\rho(Y). \end{aligned} \tag{22}$$

The same relation holds for $\rho(X)$ being selected as the *absolute semideviation*, $\rho(X) = E[(X - E[X])_-]$. In [44], it was shown that a generalization of (22) involving central semi-moments of higher orders holds for the kth order stochastic dominance relation (5). Namely, X dominating Y with respect to the $(k+1)$-order stochastic dominance, $X \succeq_{(k+1)} Y$, implies

$$E[X] \geq E[Y] \quad \text{and} \\ E[X] - \|(X - E[X])_-\|_k \geq E[Y] - \|(Y - E[Y])_-\|_k. \tag{23}$$

The semivariance risk measure $\sigma_-^2(X)$ reflects asymmetric risk preferences; observe, however, that in accordance to its definition (20), the risk is associated with X falling below its expected level, $E[X]$. In many applications, it is desirable to view the risk of X as its shortfall with respect to a certain predefined benchmark level a. Then, if risk is identified with the average shortfall below a target (benchmark) level $a \in \mathbb{R}$, the corresponding *Expected Regret* (ER) measure (see, e.g., [41,45]) is defined as

$$\mathrm{ER}(X) = E[(a - X)_+] = E[(X - a)_-]. \tag{24}$$

The Expected Regret is a special case of the so-called *Lower Partial Moment* measure [46,47]:

$$\mathrm{LPM}_p(X, a) = E\big[(X - a)_-^p\big], \quad p \geq 0,\ a \in \mathbb{R}. \tag{25}$$

A special case of (25) with $p = 2$, a semideviation below a fixed target, was considered by Porter [48], who demonstrated that the corresponding mean risk model is consistent with SSD dominance ordering, i.e. an outcome that is mean risk efficient is also SSD efficient, except for outcomes with identical mean and semivariance. Bawa [46] related the mean risk model with $\rho(X) = \mathrm{LPM}_2(X, a)$ to the third-order stochastic dominance for a class of decreasing absolute risk-averse utility functions. For $p = 0$, LPM (25) can be considered as the "probability of loss", i.e., the probability of X not exceeding the level a, and is related to the Value-at-Risk measure discussed below.

A requirement that risk, when measured by the lower partial moment function $\mathrm{LPM}_p(X, a)$, should not exceed some level $b > 0$, can be expressed as a risk constraint of the form

$$E[(X - a)_-^p] \leq b.$$

In the special case of $p = 1$, the above constraint is known as the Expected Regret constraint, and reduces to

$$E[(X - a)_-] \leq b, \tag{26}$$

which is also known as the *Integrated Chance Constraint* [49]; a more detailed discussion of constraints (26) is presented below. Further, observe that the SSD constraint in (11), corresponding to the case when the reference outcome Y is discretely distributed, can be regarded as a set of Expected Regret constraints (26).

Another popular measure of risk, frequently employed in practice, is the *Maximum Loss*, or *Worst Case Risk* (WCR), which is defined as the maximum loss that can occur over a given time horizon:

$$\mathrm{WCR}(X) = -\operatorname{ess\,inf} X. \tag{27}$$

Obviously, the WCR measure represents the most conservative risk-averse preferences. At the same time, $\mathrm{WCR}(X)$, as a measure of risk, essentially disregards the distributional information of the profit/loss profile X. Despite this, the Worst Case Risk measure, with an appropriately defined function $X(\mathbf{x}, \omega)$, has been successfully applied in many decision making problems under uncertainties, including portfolio optimization [50,40], location theory, machine scheduling, network problems (see a comprehensive exposition in [51]).

The popularity of Worst Case Risk concept (also known as "robust" optimization approach, see [51]) in practical applications can be attributed to its to easy-to-interpret definition, as well as to its amenability to efficient implementation in stochastic programming scenario-based models; namely, for a finite $\Omega = \{\omega_1, \ldots, \omega_N\}$, minimization or bounding of risk using WCR measure can be implemented via constraint of the form

$$\mathrm{WCR}(X(\mathbf{x}, \omega)) \leq y,$$

which, in turn, can be implemented by N inequalities $y \geq -X(\mathbf{x}, \omega_j), j = 1, \ldots, N$, which are convex provided that the profit function $X(\mathbf{x}, \omega)$ is concave in the decision vector $\mathbf{x}$.

3.2. Value-at-Risk and chance constraints

One of the most widely known risk measures in the area of financial risk management is the *Value-at-Risk (VaR)* measure (see, for instance, [52–54], and references therein). Methodologically, if X represents the value of a financial position, then, for instance, its Value-at-Risk at a 0.05 confidence level, denoted as $\mathrm{VaR}_{0.05}(X)$, defines the risk of X as the amount that can be lost with probability no more than 5%, over the given time horizon (e.g., 1 week). Mathematically, Value-at-Risk with a confidence level $\alpha \in (0, 1)$ is defined as the α-quantile of the probability distribution F_X of X:

$$\mathrm{VaR}_\alpha(X) = -\inf\{z \mid \mathbb{P}\{X \leq z\} > \alpha\} = -F_X^{-1}(\alpha). \tag{28}$$

Often, a "lower" α-quantile is used (see, among others, [55–57])

$$\mathrm{VaR}_\alpha^-(X) = -\inf\{z \mid P\{X \leq z\} \geq \alpha\} = -F_{(-1)}(X, \alpha), \tag{29}$$

where $F_{(-1)}$ is defined as in (14). It is easy to see that VaR measure is consistent with the first-order stochastic dominance:

$$X \succeq_{(1)} Y \Rightarrow \mathrm{VaR}_\alpha(X) \geq \mathrm{VaR}_\alpha(Y).$$

In addition, VaR is *comonotonic additive* [58]:

$$\mathrm{VaR}_\alpha(X + Y) = \mathrm{VaR}_\alpha(X) + \mathrm{VaR}_\alpha(Y),$$

for all X, Y that are *comonotone* (see, e.g., [24,20]), namely, for such X and Y, defined on the same probability space, that satisfy

$$(X(\omega_1) - X(\omega_2))(Y(\omega_1) - Y(\omega_2)) \geq 0 \quad \text{a.s.} \\ \text{for every } \omega_1,\ \omega_2 \in \Omega \tag{30}$$

(alternatively, X and Y are comonotonic if and only if there exists Z and increasing real functions f and g such that $X = f(Z), Y = g(Z)$, see [59]).

Due to its intuitive definition and wide utilization by major banking institutions [52], the VaR measure has been adopted as the *de facto* standard for measuring risk exposure of financial positions. However, VaR has turned out to be a technically and methodologically challenging construct for control and optimization of risk. One of the major deficiencies of VaR, from the methodological point of view, is that it does not take into account the extreme losses beyond the α-quantile level. Even more importantly, VaR has been proven to be generally inconsistent with the fundamental risk management principle of *risk reduction via diversification*: it is possible that VaR of a financial portfolio may exceed the sum of VaRs of its components. This is a manifestation of the mathematical fact that, generally, $\mathrm{VaR}_\alpha(X)$ is a non-convex function of X. VaR exhibits convexity in the special case when the distribution of X is *elliptic*; in this case, moreover, minimization of VaR can be considered equivalent to the Markowitz MV model [60]. In addition, $\mathrm{VaR}_\alpha(X)$ is discontinuous with respect to the confidence level α, meaning that small changes in the values of α can lead to significant jumps in the risk estimates provided by VaR.

Being simply a quantile of payoff distribution, the Value-at-Risk concept has its counterparts in the form of *probabilistic*, or *chance* constraints that were first introduced in [61] and since then have been widely used in such disciplines as operations research and stochastic programming [1,5,7], systems reliability theory [62,63], reliability-based design and optimization [64], and others. If the payoff $X = X(\mathbf{x}, \omega)$ is a function of the decision vector $\mathbf{x} \in \mathbb{R}^n$, the chance constraint may stipulate that X should exceed a certain predefined level c with probability at least $\alpha \in (0, 1)$:

$$\mathbb{P}\{X(\mathbf{x}, \omega) \geq c\} \geq \alpha, \tag{31}$$

whereas in the case of $\alpha = 1$ constraint (31) effectively requires that the inequality $X(\mathbf{x}, \omega) \geq c$ holds almost surely (a.s.). For a review of solution methods for chance-constrained stochastic programming problems, see [65]. Using, without loss of generality, definition (29), it is easy to see that probabilistic constraint (31) can be expressed as a constraint on the Value-at-Risk of $X(\mathbf{x}, \omega)$:

$$\mathrm{VaR}_{1-\alpha}(X(\mathbf{x}, \omega)) \leq -c. \tag{32}$$

Chance constraints are well known for their non-convex structure, particularly in the case when the set Ω is discrete, $\Omega = \{\omega_1, \dots, \omega_N\}$. Observe that in this case, even when the set $X(\mathbf{x}, \omega_i) \geq c$ is convex in $\mathbf{x}$ for every $\omega_i \in \Omega$, the chance constraint (31) can be non-convex for $\alpha \in (0, 1)$.

Because of the general non-convexity of constraints (31), a number of convex relaxations of chance constraints have been developed in the literature. One of such relaxations, the *Integrated Chance Constraints* (ICC) [49], see also [66,29], can be derived by considering a parametrized chance constraint

$$\mathbb{P}\{X \leq \xi\} \leq \alpha(\xi), \quad \xi \in \Xi, \tag{33}$$

where $\alpha(\xi)$ is increasing in ξ, which means that smaller values of X are less desirable. Then, assuming that $\Xi = (-\infty, c]$, and integrating (33), one arrives at the integrated chance constraint

$$E[(X - c)_-] = \int_{-\infty}^{c} \mathbb{P}\{X \leq \xi\}\, \mathrm{d}\xi \leq \int_{-\infty}^{c} \alpha(\xi)\, \mathrm{d}\xi =: \beta. \tag{34}$$

Observe that constraints of the form (34) are equivalent to the expected regret, or expected shortfall constraints (26). Other convex approximations to the chance constraints have been obtained by replacing VaR in (32) with a convex risk functional, such as the Conditional Value-at-Risk measure (see below); a Bernstein approximation of chance constraints has been recently proposed by Nemirovski and Shapiro [67].

4. Coherent measures of risk

Historically, development of risk models used in the Markowitz risk-reward framework has been to a large degree application-driven, or "ad hoc", meaning that new risk models have been designed in an attempt to represent particular risk preferences or attitudes in decision making under uncertainty. As a result, some risk models, while possessing certain attractive properties, have been lacking some seemingly fundamental features, which undermined their applicability in many problems. The most notorious example of this is the Value-at-Risk measure, which has been heavily criticized by both academicians and practitioners for its lack of convexity and other shortcomings.

Thus, an axiomatic approach to the construction of risk models has been proposed by Artzner et al. [68], who undertook the task of determining the set of requirements, or axioms that a "good" risk function must satisfy. From a number of such potential requirements they identified four, and called the functionals that satisfied these four requirements *coherent measures of risk*. Since the pioneering work [68], the axiomatic approach has become the dominant framework in risk analysis, and a number of new classes of risk measures, tailored to specific preferences and applications, have been developed in the literature. Examples of such risk measures include *convex risk measures* [69,70], *deviation measures* [71], and others.

Exposition in this section assumes that[3] $\mathcal{X} = \mathcal{L}^\infty(\Omega, \mathcal{F}, P)$ is a space of all bounded $\mathcal{F}$-measurable functions $X : \Omega \mapsto \mathbb{R}$; for a discussion of risk measures on general spaces see, for example, [70]. Then, a *coherent risk measure* is defined as a mapping $\mathcal{R} : \mathcal{X} \mapsto \mathbb{R}$ that satisfies the following four axioms [68,72]:

(A1) *monotonicity:* $X \geq 0$ implies $\mathcal{R}(X) \leq 0$ for all $X \in \mathcal{X}$
(A2) *convexity:* $\mathcal{R}(\lambda X + (1 - \lambda)Y) \leq \lambda\mathcal{R}(X) + (1 - \lambda)\mathcal{R}(Y)$ for all $X, Y \in \mathcal{X}$ and $\lambda \in [0, 1]$
(A3) *positive homogeneity:* $\mathcal{R}(\lambda X) = \lambda\mathcal{R}(X)$ for all $X \in \mathcal{X}$ and $\lambda > 0$
(A4) *translation invariance:* $\mathcal{R}(X + a) = \mathcal{R}(X) - a$ for all $X \in \mathcal{X}$ and $a \in \mathbb{R}$

It must be noted that if the coherent risk measure $\mathcal{R}$ is allowed to take values in the extended real line (see, e.g., [70]), it is necessary to impose additional requirements on $\mathcal{R}$, such as lower semicontinuity and properness. Moreover, certain continuity properties are required for various representation results discussed below; one of the most common such requirements that augment the set of axioms (A1)–(A4) for coherent risk measures is the *Fatou property* (see, for instance, [72,11,73]), e.g., that for any bounded sequence $\{X_n\}$ that converges P-a.s. to some X, the coherent risk measure must satisfy

$$\mathcal{R}(X) \leq \liminf_{n\to\infty} \mathcal{R}(X_n). \tag{35}$$

In order to avoid excessively technical discussion, throughout this section it will be implicitly assumed that the risk measure in question satisfies the appropriate topological conditions, e.g., (35).

The monotonicity axiom (A1) maintains that lower values of X bear more risk. In fact, by combining (A1) with (A2) and (A3) it can be immediately seen that

$$\mathcal{R}(X) \leq \mathcal{R}(Y) \quad \text{whenever } X \geq Y,$$

and, in particular, that $X \geq -a$ implies $\mathcal{R}(X) \leq a$ for all $a \in \mathbb{R}$.

The convexity axiom (A2) is a key property from both the methodological and computational perspectives. In the mathematical programming context, it means that $\mathcal{R}(X(\mathbf{x}, \omega))$ is a convex function of the decision vector $\mathbf{x}$, provided that the profit $X(\mathbf{x}, \omega)$ is concave in $\mathbf{x}$. This, in turn, entails that the minimization of risk over a convex set of decisions $\mathbf{x}$ constitutes a convex programming problem, amenable to efficient solution procedures. Moreover, convexity of coherent risk measures has important implications from the methodological risk management viewpoint: given the positive homogeneity (A3), convexity entails subadditivity

(A2′) *subadditivity:* $\mathcal{R}(X + Y) \leq \mathcal{R}(X) + \mathcal{R}(Y)$ for all $X, Y \in \mathcal{X}$,

which is a mathematical expression of the fundamental risk management principle of *risk reduction via diversification*. Further, convexity allows one to construct coherent measures of risk by combining several coherent functionals using an operation that preserves convexity; for instance,

$$\mathcal{R}(X) = \sum_{i=1}^{k} \lambda_i \mathcal{R}_i(X) \quad \text{and} \quad \mathcal{R}(X) = \max\{\mathcal{R}_1(X), \dots, \mathcal{R}_k(X)\}$$

are coherent, provided that $\mathcal{R}_i(X)$ satisfy (A1)–(A4) and $\lambda_i \geq 0$, $\lambda_1 + \cdots + \lambda_k = 1$.

[3] Although this assumption does not apply to random variables X with unbounded support, e.g., X that are normally distributed, it provides a common ground for most of the results presented in what follows, and allows us to avoid excessively technical exposition.

The positive homogeneity axiom (A3) ensures that if all realizations of X increase or decrease uniformly by a positive factor, the corresponding risk $\mathcal{R}(X)$ scales accordingly. Such a requirement is natural in the context of financial applications, when X represents the monetary payoff of a financial position; obviously, doubling the position value effectively doubles the risk. In some applications, however, such a behavior of $\mathcal{R}$ may not be desirable, and a number of authors have dropped the positive homogeneity from the list of properties required for "nicely behaved" risk measures (see, e.g., [69,70]).

The translation invariance (A4) is also supported by the financial interpretation: if X is a payoff of a financial position, then adding cash to this position reduces its risk by the same amount; in particular, one has

$$\mathcal{R}(X + \mathcal{R}(X)) = 0.$$

Combined with (A3), the translation invariance (A4) also states that the risk of a deterministic payoff factor is given by its negative value:

$$\mathcal{R}(0) = 0, \quad \text{and, in general,} \quad \mathcal{R}(a) = -a \quad \text{for all } a \in \mathbb{R}.$$

It is also worth noting that, given the subadditivity of $\mathcal{R}$, the last condition can be used in place of (A4), see [71]. Finally, we note that, in general, coherent risk measures are inconsistent with utility theory and second-order stochastic dominance, in the sense that if element X is preferred to Y by a risk-averse utility maximizer, $X \succeq_{(2)} Y$, it may happen that X carries a greater risk than Y, $\mathcal{R}(X) > \mathcal{R}(Y)$, when measured by a coherent risk measure; see [74] for an explicit example. To address the issue of consistency with utility theory, the following *SSD isotonicity* property has been considered in addition to or in place of (A1) (see, e.g., [74–76]):

(A1′) *SSD isotonicity:* $\mathcal{R}(X) \le \mathcal{R}(X)$ for all $X, Y \in \mathcal{X}$ such that $X \succeq_{(2)} Y$.

Obviously, (A1′) implies (A1).

According to the above definition (A1)–(A4), the VaR measure (28) is not coherent: although it satisfies axiom (A1), (A3), and (A4), in the general case it fails the all-important convexity (subadditivity) property. On the other hand, the Maximum Loss, or Worst Case Risk measure (27) is coherent; recall that the WCR measure reflects the extremely conservative risk-averse preferences. Interestingly, the class of coherent risk measures also contains the opposite side of the risk preferences spectrum, namely, it is easy to see that $\mathcal{R}(X) = E[-X]$ is coherent, despite representing risk-neutral preferences.

It is worth noting that while the set of axioms (A1)–(A4) has been construed so as to ensure that the risk measure $\mathcal{R}$ satisfying these properties would behave "properly", and produce an "adequate" picture of risk exposure, there exist coherent risk measures that do not represent risk-averse preferences. For example, let space Ω be finite, $\Omega = \{\omega_1, \dots, \omega_N\}$, and, for a fixed j, define the risk measure $\mathcal{R}$ as

$$\mathcal{R}(X) = -X(\omega_j). \tag{36}$$

It is elementary to check that defined in such a manner $\mathcal{R}$ does indeed satisfy axioms (A1)–(A4), and thus is a coherent measure of risk. On the other hand, definition (36) entails that the risk of random outcome X is estimated by *guessing the future*, an approach that rightfully receives much disdain in the field risk management and, generally, decision making under uncertainty. *Averse measures of risk* and their axiomatic foundation are discussed in Section 5.2.

The axiomatic foundation (A1)–(A4), along with a number of other properties considered in subsequent works (see, for instance, [77] for a discussion of interdependencies among various sets of axioms) only postulates the key properties for "well-behaved" measures of risk, but it does not provide functional "recipes" for construction of coherent risk measures. Thus, substantial attention has been paid in the literature to the development of representations for functionals that satisfy (A1)–(A4). One of the most fundamental such representations was presented in the original work [68]. With respect to a coherent risk measure $\mathcal{R}$, the authors introduced the notion of *acceptance set* as a convex cone

$$\mathcal{A}_{\mathcal{R}} = \{X \in \mathcal{X} \mid \mathcal{R}(X) \le 0\}. \tag{37}$$

In the financial interpretation, the cone $\mathcal{A}_{\mathcal{R}}$ contains positions X that comply with capital requirements. The risk preferences introduced by a coherent measure $\mathcal{R}$ are equivalently represented by the acceptance set $\mathcal{A}_{\mathcal{R}}$, and, moreover, $\mathcal{R}$ can be recovered from $\mathcal{A}_{\mathcal{R}}$ as

$$\mathcal{R}(X) = \inf\{c \in \mathbb{R} \mid X + c \in \mathcal{A}\}. \tag{38}$$

Artzner et al. [68] and Delbaen [72] have established that mapping $\mathcal{R} : \mathcal{X} \mapsto \mathbb{R}$ is a coherent measure of risk if and only if

$$\mathcal{R}(X) = \sup_{Q \in \mathcal{Q}} E_Q[-X], \tag{39}$$

where $\mathcal{Q}$ is a closed convex subset of P-absolutely continuous probability measures. For *convex risk measures* (i.e., functionals satisfying (A1), (A2), and (A4)), Föllmer and Schied [69] have generalized the above result:

$$\mathcal{R}(X) = \max_{Q \in \mathcal{Q}} (E_Q[-X] - \alpha(Q)), \tag{40}$$

where α is the penalty function defined for $Q \in \mathcal{Q}$ as

$$\alpha(Q) = \sup_{X \in \mathcal{A}_{\mathcal{R}}} E_Q[-X] = \sup_{X \in \mathcal{X}} (E_Q[-X] - \mathcal{R}(X)), \tag{41}$$

and is therefore the *conjugate function* (see, e.g., [78,79]) of $\mathcal{R}$ on $\mathcal{X}$. A *subdifferential* representation of convex risk measures, which satisfy an additional requirement of $\mathcal{R}(X) \le E[-X]$, was proposed in [75], see also [70]. Representations for coherent and convex risk measures that satisfy an additional property of *law invariance*:

(A5) *law invariance:* $\mathcal{R}(X) \le \mathcal{R}(X)$ for all $X, Y \in \mathcal{X}$ such that $\mathbb{P}\{X \le z\} = \mathbb{P}\{Y \le z\}, z \in \mathbb{R}$,

or, roughly speaking, can be estimated from empirical data, were considered in [80,81,77,82,11,83].

Acerbi [84] has suggested the following *spectral* representation:

$$\mathcal{R}(X) = \int_0^1 \mathrm{VaR}_\lambda(X)\phi(\lambda)\,\mathrm{d}\lambda, \tag{42}$$

where $\phi \in \mathcal{L}^1([0, 1])$ is the "risk spectrum". Then, the functional $\mathcal{R}$ defined by (42) is a coherent risk measure if the risk spectrum ϕ integrates to 1, and is "positive" and "decreasing" (however not pointwise, but in "$\mathcal{L}^1$ sense", see [84] for details).

Differentiability properties of *convex risk measures* that are defined on general probability spaces and satisfy axioms (A1), (A2), and (A4) have been discussed by Ruszczyński and Shapiro [70], who also generalized some of the above representations for convex and coherent measures of risk and presented optimality conditions for optimization problems with risk measures.

Since the pioneering work of Artzner et al. [68], a number of generalizations to the concept of coherent measures of risk have been proposed in the literature, including vector and set-valued coherent risk measures, see, e.g., [85,86]. Dynamic multi-period extensions of coherent and convex measures of risk has been considered in [87,88,73,89].

4.1. Conditional Value-at-Risk and related risk measures

The Conditional Value-at-Risk (CVaR) measure has been designed as a measure of risk that would remedy the shortcomings of VaR (most importantly, its non-convexity) while preserving its intuitive practical meaning. For a random payoff or profit function X that has a continuous distribution, Rockafellar and Uryasev [90] have defined CVaR with a confidence level $\alpha \in (0,1)$ as the conditional expectation of losses that exceed the $\mathrm{VaR}_\alpha(X)$ level:

$$\mathrm{CVaR}_\alpha(X) = \mathrm{CVaR}_\alpha^-(X) = -E[X \mid X \le -\mathrm{VaR}_\alpha(X)]. \tag{43}$$

In accordance with this definition, for example, the 5% Conditional Value-at-Risk, $\mathrm{CVaR}_{0.05}(X)$, represents the *average* of worst case losses that may occur with 5% probability (over a given time horizon). Observe that in such a way CVaR addresses the issue of estimating the amount of losses possible at a given confidence level, whereas the corresponding VaR only provides a lower bound on such a loss. Expression (43) is also known in the literature under the name of *Tail Conditional Expectation* (TCE) [68]. In addition, Artzner et al. [68] introduced a related measure of risk, the *Worst Conditional Expectation* (WCE):

$$\mathrm{WCE}_\alpha(X) = \sup\{E[-X \mid A] \mid A \in \mathcal{F}, \mathbb{P}\{A\} > \alpha\}. \tag{44}$$

It turns out that the quantity (43), which in the general case is known as "lower" CVaR, maintains convexity in the case of continuous X (or, more generally, when the distribution function F_X is continuous at $-\mathrm{VaR}_\alpha(X)$), whereas for general (arbitrary) distributions F_X it does not possess convexity with respect to X. Moreover, neither does the "upper" CVaR defined as the conditional expectation of losses strictly exceeding the $\mathrm{VaR}_\alpha(X)$ level:

$$\mathrm{CVaR}_\alpha^+(X) = -E[X \mid X < -\mathrm{VaR}_\alpha(X)]. \tag{45}$$

In [55], a more intricate definition for Conditional Value-at-Risk for general distributions was introduced, which presented $\mathrm{CVaR}_\alpha(X)$ as a convex combination of $\mathrm{VaR}_\alpha(X)$ and $\mathrm{CVaR}_\alpha^+(X)$:

$$\begin{aligned}\mathrm{CVaR}_\alpha(X) = {}& \lambda_\alpha(X)\mathrm{VaR}_\alpha(X) \\ & + (1-\lambda_\alpha(X))E[-X \mid X < -\mathrm{VaR}_\alpha(X)],\end{aligned} \tag{46}$$

where $\lambda_\alpha(X) = (1-\alpha)^{-1}F_X(-\mathrm{VaR}_\alpha(X))$. Rockafellar and Uryasev [55] have demonstrated that $\mathrm{CVaR}_\alpha(X)$ as defined in (46) is convex in X, and is a coherent measure of risk satisfying the axioms (A1)–(A4). Thus, the following chain of inequalities hold:

$$\begin{aligned}\mathrm{VaR}_\alpha(X) &\le \mathrm{CVaR}_\alpha^-(X) \le \mathrm{WCE}_\alpha(X) \\ &\le \mathrm{CVaR}_\alpha(X) \le \mathrm{CVaR}_\alpha^+(X),\end{aligned} \tag{47}$$

where only $\mathrm{CVaR}_\alpha(X)$ and $\mathrm{WCE}_\alpha(X)$ are coherent in the general case; however, for continuously distributed X the last three inequalities become identities (see, for instance, [55,11] for details).

Besides convexity, $\mathrm{CVaR}_\alpha(X)$ is also continuous in α, which from the risk management perspective means that small variations in the confidence level α result in small changes of risk estimates furnished by the CVaR. In contrast, VaR, as a distribution quantile, is in general discontinuous in α, and therefore can experience jumps due to small variations in α. Furthermore, for the limiting values of α one has

$$\begin{aligned}&\lim_{\alpha\to 1} \mathrm{CVaR}_\alpha(X) = E[-X], \\ &\lim_{\alpha\to 0} \mathrm{CVaR}_\alpha(X) = -\inf X = \mathrm{WCR}(X),\end{aligned} \tag{48}$$

which entails that depending on the choice of the confidence level α, $\mathrm{CVaR}_\alpha(X)$ as a measure of risk can represent a broad spectrum of risk preferences, from the most conservative risk-averse preferences ($\alpha = 0$) to risk-neutrality ($\alpha = 1$).

The functional (46) is also known in the literature under the names of *Expected Shortfall* (ES) [91,56], *Tail VaR* (TVaR) [92], and *Average Value-at-Risk* (AVaR) (see, e.g., [11,82,7], and others). The latter nomenclature is justified by the following representation for CVaR due to Acerbi [84] (compare to (42)):

$$\mathrm{CVaR}_\alpha(X) = \frac{1}{\alpha}\int_0^\alpha \mathrm{VaR}_\lambda(X)\,\mathrm{d}\lambda. \tag{49}$$

Kusuoka [80] has shown that CVaR is the smallest law-invariant coherent risk measure that dominates VaR; at the same time, if the law invariance requirement (A5) is dropped, then the smallest convex (coherent) VaR dominating risk measure does not exist [72,11], i.e.,

$$\begin{aligned}\mathrm{VaR}_\alpha(X) = {}& \min\{\,\mathcal{R}(X) \mid \mathcal{R}(X) \ge \mathrm{VaR}_\alpha(X) \\ & \text{and } \mathcal{R}(X) \text{ is convex (coherent)}\}.\end{aligned}$$

The importance of CVaR measure in the context of coherent and convex measures of risk can be seen from the following representation for law-invariant coherent measures of risk on atomless probability spaces, first obtained by Kusuoka [80]:

$$\mathcal{R}(X) = \sup_{\mu\in\mathcal{M}'\subset\mathcal{M}(0,1]} \mathcal{R}_\mu(X), \tag{50}$$

where $\mathcal{M}(0,1]$ is the set of all probability measures on $(0,1]$, and

$$\mathcal{R}_\mu(X) = \int_{(0,1]} \mathrm{CVaR}_\xi(X)\mu(\mathrm{d}\xi). \tag{51}$$

Moreover, for any given μ the risk measure $\mathcal{R}_\mu$ is law invariant, coherent, and comonotonic. Coherent risk measures of the form (51), dubbed *Weighted VaR* (WVaR), were discussed by Cherny [92], who showed that $\mathcal{R}_\mu$ are strictly subadditive, i.e.,

$$\mathcal{R}_\mu(X+Y) < \mathcal{R}_\mu(X) + \mathcal{R}_\mu(Y),$$

unless X and Y are comonotone. Representation (50) and (51) has its counterpart for convex measures of risk [11]:

$$\mathcal{R}(X) = \sup_{\mu\in\mathcal{M}(0,1]}\left(\int_{(0,1]} \mathrm{CVaR}_\xi(X)\mu(\mathrm{d}\xi) - \beta(\mu)\right), \quad \text{where}$$

$$\beta(\mu) = \sup_{X\in\mathcal{A}_\mathcal{R}} \int_{(0,1]} \mathrm{CVaR}_\xi(X)\mu(\mathrm{d}\xi). \tag{52}$$

In other words, the family of CVaR_α risk measures can be regarded as "building blocks" for law-invariant coherent or convex measures of risk [11]. Furthermore, Inui and Kijima [57] demonstrate that any coherent measure of risk can be represented as a convex combination of CVaR functionals with appropriately chosen confidence levels.

A connection between risk optimization problems with coherent risk measures of form (51) and problems with inverse stochastic dominance constraints (15) has been pointed out by Dentcheva and Ruszczyński [33], who showed that risk-reward optimization problems of the form

$$\max\{f(X) - \lambda\mathcal{R}(X) \mid X \in \mathcal{C}\}, \quad \lambda \ge 0,$$

where $\mathcal{R}(\mathcal{X})$ is a law-invariant risk measure of the form (51), can be regarded as Lagrangian dual of a problem with inverse second-order stochastic dominance constraint (15).

Despite the seemingly complicated definitions (46) and (49), Rockafellar and Uryasev [90,55] have shown that CVaR can be computed as the optimal value of the following optimization problem:

$$\mathrm{CVaR}_\alpha(X) = \min_{\eta\in\mathbb{R}} \Phi_\alpha(X,\eta), \quad \text{where}$$

$$\Phi_\alpha(X,\eta) = \eta + \alpha^{-1}E(X+\eta)_-, \ \alpha \in (0,1). \tag{53a}$$

The importance of representation (53) stems from the fact that the function $\Phi_\alpha(X, \eta)$ is jointly convex in $X \in \mathcal{X}$ and $\eta \in \mathbb{R}$, and thus (53) is a convex programming problem that can be solved very efficiently. Moreover, the optimal value of η that delivers the minimum in (53) is given by $-\mathrm{VaR}_\alpha(X)$, or, more precisely,

$$\mathrm{VaR}_\alpha(X) = \min\{-y \mid y \in \arg\min_{\eta\in\mathbb{R}} \Phi_\alpha(X, \eta)\}. \tag{53b}$$

In fact, the convex (stochastic) programming representation (53) can itself be considered as a definition of CVaR; namely, Pflug [58] demonstrated that coherence properties (A1)–(A4) can be established from (53), and, in addition, that CVaR as the optimal value in (53) satisfies the SSD isotonicity axiom (A1′).

In the case when the profit function $X = X(\mathbf{x}, \omega)$ is concave in the decision vector $\mathbf{x}$ over some closed convex set $\mathcal{S} \subset \mathbb{R}^n$, the result (53) due to Rockafellar and Uryasev [90,55] allows for risk minimization using the Conditional Value-at-Risk measure via an equivalent formulation involving the function Φ_α:

$$\min_{\mathbf{x}\in\mathcal{S}} \mathrm{CVaR}_\alpha(X(\mathbf{x}, \omega)) \Leftrightarrow \min_{(\mathbf{x},\eta)\in\mathcal{S}\times\mathbb{R}} \Phi_\alpha(X(\mathbf{x}, \omega), \eta), \tag{54}$$

(see [55] for details). Furthermore, similar arguments can be employed to handle CVaR constraints in convex programming problems, namely, the risk constraint

$$\mathrm{CVaR}_\alpha(X(\mathbf{x}, \omega)) \le c \tag{55}$$

can be equivalently replaced by (see the precise conditions in [55,93])

$$\Phi_\alpha(X(\mathbf{x}, \omega), \eta) \le c. \tag{56}$$

Convexity of the function $\Phi_\alpha(X, \eta)$ implies convexity of the optimization problems in (54) and constraints (55) and (56). Within the stochastic programming framework, when the uncertain element ω is modeled by a finite set of scenarios $\{\omega_1, \ldots, \omega_N\}$ such that $\mathbb{P}\{\omega_j\} = p_j \in (0, 1)$, constraint (56) can be implemented using $N + 1$ auxiliary variables and $N + 1$ convex constraints (provided that $X(\mathbf{x}, \omega_j)$ are all concave in $\mathbf{x}$):

$$\begin{aligned} &\eta + \alpha^{-1}\sum_{j=1}^{N} p_j w_j \le c,\\ &w_j + X(\mathbf{x}, \omega_j) + \eta \ge 0, \quad j = 1, \ldots, N,\\ &w_j \ge 0, \quad j = 1, \ldots, N. \end{aligned} \tag{57}$$

When $X(\mathbf{x}, \omega_j)$ are linear in $\mathbf{x}$, constraints (57) define a polyhedral set, which allows for formulating many stochastic optimization models involving CVaR objective or constraints as linear programming (LP) problems that can be solved efficiently using many existing LP solver packages. For large-scale problems, further efficiencies in handling constructs of the form (57) have been proposed in the literature, including cutting plane methods [94], smoothing techniques [95], non-differentiable optimization methods [96].

Due to the mentioned fact that CVaR is the smallest coherent law-invariant risk measure dominating VaR, the CVaR constraint (55) can be employed as a convexification of the chance constraint

$$\mathrm{VaR}_\alpha(X(\mathbf{x}, \omega)) \le c. \tag{58}$$

Observe that by virtue of inequalities (47), CVaR constraint (55) is more conservative than (58). Constraints of the form (58) are encountered in many engineering applications, including systems reliability theory [62,63] and reliability-based design and optimization [64]. Specifically, expression (58) with $c = 0$ and $-X(\mathbf{x}, \omega)$ defined as the so-called *limit-state function* is well known in reliability theory, where it represents the probability of the system being "safe", i.e., in the state $X(\mathbf{x}, \omega) \ge 0$. Based on the discussed above properties of the VaR and CVaR measures, Rockafellar and Royset [97] introduced the *buffered failure probability*, which accounts for a degree of "failure" (the magnitude of the negative value of $X(\mathbf{x}, \omega)$), and bounds from above the probability of failure using the CVaR constraint (55). Similarly, application of constraints of the form (55) in place of chance constraints for robust facility location design under uncertainty was considered in [98].

4.2. Risk measures defined on translation invariant hulls

The convex programming representation (53) due to Rockafellar and Uryasev [90,55] can viewed as a special case of more general representations that give rise to classes of coherent (convex) risk measures discussed below.

A constructive representation for coherent measures of risk that can be efficiently applied in stochastic optimization context has been proposed in [76]. Assuming that function $\phi : \mathcal{X} \mapsto \mathbb{R}$ is lower semicontinuous, such that $\phi(\eta) > 0$ for all real $\eta \neq 0$, and satisfies three axioms (A1)–(A3), the optimal value of the following (convex) stochastic programming problem is a coherent measure of risk (similar constructs have been investigated by Ben-Tal and Teboulle [99,100], see discussion below):

$$\mathcal{R}(X) = \inf_\eta\{\eta + \phi(X + \eta)\}. \tag{59}$$

If the function ϕ in (59) satisfies the SSD isotonicity property (A1′), then the corresponding $\mathcal{R}(X)$ is also SSD isotonic. Further, the function defined on the set of optimal solutions of problem (59)

$$\eta(X) = \min\left\{-y \mid y \in \arg\min_{\eta\in\mathbb{R}} \eta + \phi(X + \eta)\right\} \tag{60}$$

exists and satisfies the positive homogeneity and translation invariance axioms (A3), (A4). If, additionally, $\phi(X) = 0$ for every $X \ge 0$, then $\eta(X)$ satisfies the monotonicity axiom (A1), along with the inequality $\eta(X) \le \mathcal{R}(X)$. Observe that representation (53) of Conditional Value-at-Risk measure due to [90,55] constitutes a special case of (59); the former statement on the properties of the function $\eta(X)$ (60) illustrates that the properties of VaR as a risk measure (see (53)) are shared by a larger class of risk measures obtained from representations of the form (59).

Similarly to the CVaR formula due to [90,55], representation (59) can facilitate implementation of coherent risk measures in stochastic programming problems. Namely, for $\mathcal{R}(X)$ that has a representation (59), the following (convex) problems with risk objective and constraints can be equivalently reformulated as

$$\begin{aligned} &\min_{\mathbf{x}\in\mathcal{S}} \mathcal{R}(X(\mathbf{x}, \omega)) \Leftrightarrow \min_{(\mathbf{x},\eta)\in\mathcal{S}\times\mathbb{R}} \{\eta + \phi(X(\mathbf{x}, \omega) + \eta)\},\\ &\min_{\mathbf{x}\in\mathcal{S}}\{g(\mathbf{x}) \mid \mathcal{R}(X(\mathbf{x}, \omega)) \le c\}\\ &\Leftrightarrow \min_{(\mathbf{x},\eta)\in\mathcal{S}\times\mathbb{R}} \{g(\mathbf{x}) \mid \eta + \phi(X(\mathbf{x}, \omega) + \eta) \le c\}, \end{aligned} \tag{61}$$

where the set $\mathcal{S} \subset \mathbb{R}^n$ is convex and closed, and functions $g(\mathbf{x})$ and $-X(\mathbf{x}, \omega)$ are convex on $\mathcal{S}$ (see [76] for details). Representation (59) was used in [76] to introduce a family of *higher moment coherent risk measures* (HMCR) that quantify risk in terms of tail moments of loss distributions,

$$\mathrm{HMCR}_{p,\alpha}(X) = \min_{\eta\in\mathbb{R}} \eta + \alpha^{-1}\|(X + \eta)_-\|_p, \quad p \ge 1,\ \alpha \in (0, 1). \tag{62}$$

Risk measures similar to (62) on more general spaces have been discussed independently by Cheridito and Li [101]. The HMCR family contains, as a special case of $p = 1$, the Conditional Value-at-Risk measure. Another family of coherent measures of risk that employ higher moments of loss distributions has been considered by Fischer [102] and Rockafellar et al. [71], under the name of *risk measures of semi-$\mathcal{L}^p$ type*:

$$\mathcal{R}_{p,\beta}(X) = E[-X] + \beta\|(X - E[X])_-\|_p, \quad p \geq 1, \beta \in [0, 1]. \tag{63}$$

In contrast to risk measures (63), the HMCR measures (62) are *tail* risk measures. By this we mean that in (63) the "tail cutoff" point, about which the partial moments are computed, is always fixed at $E[X]$, whereas in (62) the location of tail cutoff point is determined by $\eta(X) = \eta_{p,\alpha}(X)$ given by (60) with $\phi(X) = \alpha^{-1}\|X_-\|_p$, and is adjustable by means of the parameter α, such that $\eta_{\alpha,p}(X)$ is non-decreasing in α and $\eta_{p,\alpha}(X) \to -\inf X$ as $\alpha \to 0$.

The importance of HMCR measures (62) and semi-$\mathcal{L}^p$ type measures (63) is in measuring the "mass" in the left-hand tail of the payoff distribution. It is widely acknowledged that the "risk" is associated with higher moments of the loss distributions (e.g., "fat tails" are attributable to high kurtosis, etc.). The HMCR measures and semi-$\mathcal{L}^p$ measures are amenable to implementation in stochastic programming models via (convex) p-order conic constraints [103]:

$$t \geq \|\mathbf{w}\|_p \equiv (|w_1|^p + \cdots + |w_N|^p)^{1/p}$$

using transformations analogous to (57).

A comprehensive treatment of expressions of the form (59) was presented in Ben-Tal and Teboulle [99], who revisited the concept of *Optimized Certainty Equivalent* (OCE) introduced earlier by the same authors [100,104]. The concept of certainty equivalents (CE) is well known in utility theory, where it is defined as the deterministic payoff that is equivalent to the stochastic payoff X, given an increasing utility function $u(\cdot)$:

$$\mathrm{CE}_u(X) = u^{-1}(E[u(X)]). \tag{64}$$

Then, the Optimized Certainty Equivalent (OCE) was defined in [100] as the deterministic present value of a (future) income X provided that some part η of it can be consumed right now:

$$S_u(X) = \sup_{\eta}\{\eta + E[u(X - \eta)]\}, \tag{65}$$

or, in other words, as the value of optimal allocation of X between future and present. In [99] it was demonstrated that the OCE $S_u(X)$ has a direct connection to the convex risk measures satisfying (A1), (A2), and (A4) by means of the relation

$$\mathcal{R}(X) = -S_u(X), \tag{66}$$

provided that the utility u is a non-decreasing proper closed concave function, and satisfies $u(0) = 0$ and $1 \in \partial u(0)$, where ∂u is the subdifferential of u. The ranking of random variables induced by the OCE, $S_u(X) \geq S_u(Y)$, is consistent with the second-order stochastic dominance. Although generally the OCE does not satisfy the positive homogeneity property (A3), it is *subhomogeneous*, i.e.,

$$\begin{aligned} &S_u(\lambda X) \geq \lambda S_u(X), \quad \lambda \in [0, 1] \quad \text{and} \\ &S_u(\lambda X) \leq \lambda S_u(X), \quad \lambda > 1. \end{aligned} \tag{67}$$

In [99] it was shown that a positively homogeneous OCE, such that $-S_u(X)$ is a coherent measure of risk, is obtained if and only if the utility u is *strictly risk averse*, $u(t) < t$ for all $t \in \mathbb{R}$, and is a piecewise linear function of the form

$$u(t) = \gamma_1 t_+ + \gamma_2 t_-, \quad \text{for } 0 \leq \gamma_1 < 1 \leq \gamma_2. \tag{68}$$

In addition, Ben-Tal and Teboulle [99] have established an important duality between the concepts of optimized certainty equivalents (convex risk measures) and φ-divergence [105], which is a generalization of the relative entropy, or Kullback–Leibler divergence [106] as a measure of distance between random variables. Namely, for a proper closed convex function φ whose minimum value of 0 is attained at a point $t = 1 \in \mathrm{dom}\varphi$, the φ-divergence of probability measure Q with respect to P, such that Q is absolutely continuous with respect to Q, is defined as

$$I_\varphi(Q, P) = \int_\Omega \varphi\left(\frac{\mathrm{d}Q}{\mathrm{d}P}\right)\mathrm{d}P. \tag{69}$$

Defining the utility via the conjugate φ^* of the function φ as $u(t) = -\phi^*(-t)$, Ben-Tal and Teboulle [99] have shown that the optimized certainty equivalent can be represented as

$$S_u(X) = \inf_{Q \in \mathcal{Q}}\{I_\varphi(Q, P) + E_Q[X]\}, \tag{70}$$

whereby it follows that for the convex risk measure $\mathcal{R}(X) = -S_u(X)$, the penalty term $\alpha(Q)$ in the representation (40) due to Föllmer and Schied [107,11] is equal to the φ-divergence between the probability measures P and Q. Moreover, the following dual representation of φ-divergence via the OCE S_u holds:

$$I_\varphi(P, Q) = \sup_{X \in \mathcal{X}}\{S_u(X) - E_Q[X]\}. \tag{71}$$

A class of *polyhedral risk measures* that are expressed via two-stage linear stochastic programming problems [1,5,7], and thus can be viewed as generalizations of representations (59) and (65), has been proposed by Eichhorn and Römisch [108].

5. Deviation, risk, and error measures

In decision theory and finance, uncertainty in a random variable X is often translated into notions such as *risk, deviation*, and *error* revolving around the standard deviation $\sigma(X)$. By definition, $\sigma(X)$ is a measure of how X deviates from its expected value $E[X]$, i.e., $\sigma(X) = \|X - E[X]\|_2$. It is closely related to measurement of uncertainty in outcomes, i.e., to *deviation*, to aggregated measurement of probable undesirable outcomes (losses), i.e., to *risk*, and to measurement of quality of estimation in statistics, i.e., to *error*. For example, in the classical portfolio theory [4], variance, or equivalently $\sigma(X)$, is used to quantify uncertainty in returns of financial portfolios. Subtracting the expected value of portfolio return from its standard deviation, we obtain a measure which can be interpreted as risk. Therefore, with the standard deviation, we may associate a triplet $\langle\mathcal{D}, \mathcal{R}, \mathcal{E}\rangle$: deviation measure $\mathcal{D}(X) = \sigma(X) \equiv \|X - E[X]\|_2$, risk measure $\mathcal{R}(X) = \sigma(X) - E[X] \equiv \|X - E[X]\|_2 - E[X]$, and error measure $\mathcal{E}(X) = \|X\|_2$.

Another well-known example of such a triplet is the one associated with the mean absolute deviation (MAD), which sometimes is used instead of the standard deviation. In this case, $\mathcal{D}$, $\mathcal{R}$, and $\mathcal{E}$ are defined by: $\mathcal{D}(X) = \|X - E[X]\|_1$, $\mathcal{R}(X) = \|X - E[X]\|_1 - E[X]$, and $\mathcal{E}(X) = \|X\|_1$. Obviously, the triplet $\mathcal{D}(X) = \|X - E[X]\|_p$, $\mathcal{R}(X) = \|X - E[X]\|_p - E[X]$, and $\mathcal{E}(X) = \|X\|_p$ with $p \geq 1$ generalizes the previous two. However, none of these standard triplets are appropriate for applications involving noticeably asymmetric distributions of outcomes.

In financial applications, percentile or *VaR*, defined by (28), emerged as a major competitor to the standard deviation and MAD. However, as a measure of risk, $\mathrm{VaR}_\alpha(X)$ lacks convexity and provides no information of how significant losses in the α-tail could be. These VaR's deficiencies are resolved by *CVaR* [90,55], which evaluates the mean of the α-tail and in general case is defined by (46). Similar to the standard deviation and MAD, CVaR induces the triplet: CVaR-deviation $\mathcal{D}_\alpha(X) = \mathrm{CVaR}_\alpha(X - E[X])$, CVaR measure $\mathcal{R}_\alpha(X) = \mathrm{CVaR}_\alpha(X)$, and asymmetric mean absolute error [109]

$$\mathcal{E}_\alpha(X) = E[X_+ + (\alpha^{-1} - 1)X_-], \quad \alpha \in (0, 1), \tag{72}$$

which relates closely to the one used in a quantile regression [110]. For example, for $\alpha = 1/2$, $\mathcal{E}_\alpha(X)$ reduces to $\mathcal{E}(X) = \|X\|_1$.

Practical needs motivated a search for other triplets which could preserve consistency in risk preferences and could provide adequate analysis of asymmetric distributions in related decision problems. For example, if an agent uses lower semideviation in a portfolio selection problem, it is expected that the agent would use a corresponding error measure in an asset pricing factor model. In response to these needs, Rockafellar et al. [111,71,109] developed a coordinating theory of *deviation measures, error measures*, and *averse measures of risk*, which, in general, are not symmetric with respect to ups and downs of X. Deviation measures [71] quantify "*nonconstancy*" in X and preserve four main properties of the standard deviation (*nonnegativity, positive homogeneity, subadditivity*, and *insensitivity to constant shift*), whereas *error measures* quantify "*nonzeroness*" of X and generalize the mean square error (MSE). The triplets $\langle \mathcal{D}, \mathcal{R}, \mathcal{E} \rangle$ for the standard deviation, MAD and CVaR are, in fact, particular examples of more general relationships

$$\mathcal{R}(X) = \mathcal{D}(X) - E[X], \qquad \mathcal{D}(X) = \min_{c\in\mathbb{R}} \mathcal{E}(X-c) \quad \text{or}$$

$$\mathcal{D}(X) = \mathcal{E}(X - E[X]).$$

In this theory, risk, deviation, and error measures are *lower semicontinuous positively homogeneous convex* functionals satisfying closely related systems of axioms. In view of this fact, the interplay between these measures can be comprehensively analyzed in the framework of convex analysis [78,112]. Rockafellar et al. [113–115, 109] developed the mean-deviation approach to portfolio selection and derived optimality conditions for a linear regression with error measures, while Grechuk et al. [116,117] extended the Chebyshev inequality and the maximum entropy principle for *law-invariant* deviation measures (i.e. those that depend only on the distribution of X).

In what follows, $(\Omega, \mathcal{M}, \mathbb{P})$ is a probability space of elementary events Ω with the sigma-algebra $\mathcal{M}$ over Ω and with a probability measure $\mathbb{P}$ on $(\Omega, \mathcal{M})$. Random variables are measurable functions from $\mathcal{L}^2(\Omega) = \mathcal{L}^2(\Omega, \mathcal{M}, \mathbb{P})$, and the relationships between random variables X and Y, e.g. $X \le Y$ and $X = Y$, are understood to hold in the almost sure sense, i.e. $\mathbb{P}[X \le Y] = 1$ and $\mathbb{P}[X = Y] = 1$. Also, c stands for a real number or a constant random variable, and $\inf X$ and $\sup X$ mean ess $\inf X$ and ess $\sup X$, respectively.

5.1. Deviation measures

Responding to the need for flexibility in treating the ups and downs of a random outcome differently, Rockafellar et al. [71] defined a deviation measure to be a functional $\mathcal{D} : \mathcal{L}^2(\Omega) \to [0, \infty]$ satisfying the axioms

(D1) *Nonnegativity*: $\mathcal{D}(X) = 0$ for constant X, but $\mathcal{D}(X) > 0$ otherwise.
(D2) *Positive homogeneity*: $\mathcal{D}(\lambda X) = \lambda \mathcal{D}(X)$ when $\lambda > 0$.
(D3) *Subadditivity*: $\mathcal{D}(X + Y) \le \mathcal{D}(X) + \mathcal{D}(Y)$ for all X and Y.
(D4) *Lower semicontinuity*: set $\{X \in \mathcal{L}^2(\Omega) | \mathcal{D}(X) \le c\}$ is closed for all $c < \infty$.

It follows from (D1) and (D3) that (see [71])

$$\mathcal{D}(X - c) = \mathcal{D}(X) \quad \text{for all constants } c.$$

Axioms (D1)–(D4) generalize well-known properties of the standard deviation, however, they do not require symmetry, so that in general, $\mathcal{D}(-X) \ne \mathcal{D}(X)$. A deviation measure is called *lower range dominated* if in addition to (D1)–(D4), it satisfies

(D5) *Lower range dominance*: $\mathcal{D}(X) \le E[X] - \inf X$ for all X.

The importance of (D5) will be elucidated in the context of the relationship between deviation measures and coherent risk measures.

Well-known examples of deviation measures include

(a) deviation measures of $\mathcal{L}^p$ type $\mathcal{D}(X) = \|X - E[X]\|_p$, $p \in [1, \infty]$, e.g., the standard deviation $\sigma(X) = \|X - E[X]\|_2$ and mean absolute deviation $\mathrm{MAD}(X) = \|X - E[X]\|_1$,
(b) deviation measures of semi-$\mathcal{L}^p$ type $\mathcal{D}_-(X) = \|[X - E[X]]_-\|_p$ and $\mathcal{D}_+(X) = \|[X - E[X]]_+\|_p$, $p \in [1, \infty]$, e.g., *standard lower* and *upper semideviations*

$$\sigma_-(X) = \|[X - E[X]]_-\|_2, \qquad \sigma_+(X) = \|[X - E[X]]_+\|_2,$$

and *lower* and *upper worst case deviations*:

$$\mathcal{D}(X) = \|[X - E[X]]_-\|_\infty = E[X] - \inf X,$$
$$\mathcal{D}'(X) = \|[X - E[X]]_+\|_\infty = \sup X - E[X]$$

for a bounded random variable X.
(c) *CVaR-deviation* $\mathrm{CVaR}^{\Delta}_{\alpha}(X) = \mathrm{CVaR}_{\alpha}(X - E[X])$ for $\alpha \in [0, 1)$.[4]

In particular, $\mathcal{D}(X) = \|[X - E[X]]_-\|_p$, $p \in [1, \infty]$, and $\mathcal{D}(X) = \mathrm{CVaR}^{\Delta}_{\alpha}(X)$ are lower range dominated.[5]

Proposition 4 in [71] shows that deviation measures can be readily constructed out of given deviation measures $\mathcal{D}_1, \ldots, \mathcal{D}_n$ by the following two operations

$$\mathcal{D}(X) = \sum_{k=1}^{n} \lambda_k \mathcal{D}_k(X),$$

$$\sum_{k=1}^{n} \lambda_k = 1, \quad \lambda_k > 0, \ k = 1, \ldots, n,$$

and

$$\mathcal{D}(X) = \max\{\mathcal{D}_1(X), \ldots, \mathcal{D}_n(X)\}.$$

In both cases, $\mathcal{D}(X)$ is lower range dominated if each $\mathcal{D}_k(X)$ is lower range dominated. For example, taking $\mathcal{D}_k(X) = \mathrm{CVaR}^{\Delta}_{\alpha_k}(X)$ with $\alpha_k \in (0, 1)$, we obtain

$$\mathcal{D}(X) = \sum_{k=1}^{n} \lambda_k \mathrm{CVaR}^{\Delta}_{\alpha_k}(X), \qquad \sum_{k=1}^{n} \lambda_k = 1,$$
$$\lambda_k > 0, \ k = 1, \ldots, n, \tag{73}$$

and

$$\mathcal{D}(X) = \max\{\mathrm{CVaR}^{\Delta}_{\alpha_1}(X), \ldots, \mathrm{CVaR}^{\Delta}_{\alpha_n}(X)\}.$$

Rockafellar et al. [71] extended (73) for the case of continuously distributed λ:

(a) *mixed CVaR-deviation*

$$\mathcal{D}(X) = \int_0^1 \mathrm{CVaR}^{\Delta}_{\alpha}(X) \mathrm{d}\lambda(\alpha), \qquad \int_0^1 \mathrm{d}\lambda(\alpha) = 1, \quad \lambda(\alpha) \ge 0, \tag{74}$$

(b) *worst case mixed CVaR deviation*

$$\mathcal{D}(X) = \sup_{\lambda\in\Lambda} \int_0^1 \mathrm{CVaR}^{\Delta}_{\alpha}(X) \mathrm{d}\lambda(\alpha) \tag{75}$$

for some collection Λ of weighting nonnegative measures λ on $(0, 1)$ with $\int_0^1 \mathrm{d}\lambda(\alpha) = 1$.

[4] $\mathrm{CVaR}^{\Delta}_{1}(X) = -E[X] + E[X] = 0$ is not a deviation measure, since it vanishes for all r.v.'s (not only for constants).

[5] Indeed, $\|[X - E[X]]_-\|_p \le \|[X - E[X]]_-\|_\infty = E[X] - \inf X$ for $p \in [1, \infty]$, and $\mathrm{CVaR}^{\Delta}_{\alpha}(X) = E[X] - \mathrm{CVaR}_{\alpha}(X) \le E[X] - \inf X$.

These deviation measures provide a powerful modeling tool for customizing agent's risk preferences, where the weights $\lambda_1, \ldots, \lambda_n$ and the weighting measure $\lambda(\alpha)$ can be considered as discrete and continuous *risk profiles*, respectively.

Also, Proposition 5 in [71] proves that if $\int_0^1 \alpha^{-1} \mathrm{d}\lambda(\alpha) < \infty$, the deviation measure (74) can be represented in the equivalent form

$$\mathcal{D}(X) = \int_0^1 \mathrm{VaR}_\alpha(X - E[X])\phi(\alpha)\mathrm{d}\alpha,$$

$$\phi(\alpha) = \int_\alpha^1 \alpha^{-1}\mathrm{d}\lambda(\alpha),$$

where $\phi(\alpha)$ is left-continuous and nonincreasing with $\phi(0^+) < \infty$, $\phi(1^-) = 0$, and $\int_0^1 \phi(\alpha)\mathrm{d}\alpha = 1$ and plays a role similar to that of a *dual utility* function in [20,118].

5.1.1. Risk envelopes and risk identifiers

Deviation measures have dual characterization in terms of *risk envelopes* $\mathcal{Q} \subset \mathcal{L}^2(\Omega)$ defined by the properties

(Q1) $\mathcal{Q}$ is nonempty, closed and convex,
(Q2) for every nonconstant X there is some $Q \in \mathcal{Q}$ such that $E[XQ] < E[X]$,
(Q3) $E[Q] = 1$ for all $Q \in \mathcal{Q}$.

Rockafellar et al. [71, Theorem 1] showed that there is a one-to-one correspondence between deviation measures and risk envelopes:

$$\mathcal{D}(X) = E[X] - \inf_{Q\in\mathcal{Q}} E[XQ],$$

$$\mathcal{Q} = \{Q \in \mathcal{L}^2(\Omega) \mid \mathcal{D}(X) \ge E[X] - E[XQ] \text{ for all } X\},$$

and a deviation measure $\mathcal{D}$ is lower range dominated if and only if the corresponding risk envelope $\mathcal{Q}$ satisfies

(Q4) $Q \ge 0$ for all $Q \in \mathcal{Q}$.

Remarkably, with (Q4), a risk envelope $\mathcal{Q}$ can be viewed as a set of probability measures providing alternatives for the given probability measure $\mathbb{P}$. In this case, the corresponding deviation measure $\mathcal{D}(X) = E[X] - \inf_{Q\in\mathcal{Q}} E[XQ] \equiv E_{\mathbb{P}}[X] - \inf_{Q\in\mathcal{Q}} E_Q[X]$ estimates the difference of what the agent can expect under $\mathbb{P}$ and under the worst probability distribution.

The elements of $\mathcal{Q}$ at which $E[XQ]$ attains infimum for a given X are called the *risk identifiers* for X:

$$\mathcal{Q}(X) = \arg\min_{Q\in\mathcal{Q}} E[XQ].$$

In view of the one-to-one correspondence between deviation measures and risk envelopes, the risk identifiers can also be defined for each $\mathcal{D}$ through the corresponding risk envelope $\mathcal{Q}$:

$$\mathcal{Q}_{\mathcal{D}}(X) = \{Q \in \mathcal{Q} \mid \mathcal{D}(X) = E[(E[X] - X)Q] \equiv \mathrm{covar}(-X, Q)\},$$

and we say that $\mathcal{Q}_{\mathcal{D}}(X)$ is the risk identifier for X with respect to a deviation measure $\mathcal{D}$. In this case, the meaning of the risk identifiers is especially elucidating: they are those elements of $\mathcal{Q}$ that "track the downside of X as closely as possible" (see [71,114] for details).

For the standard deviation, standard lower semideviation, and CVaR-deviation, the corresponding risk envelopes and risk identifiers are given by

$$\mathcal{D}(X) = \sigma(X), \qquad \mathcal{Q} = \{Q \mid E[Q] = 1,\ \sigma(Q) \le 1\},$$

$$\mathcal{Q}_{\mathcal{D}}(X) = \left\{1 - \frac{X - E[X]}{\sigma(X)}\right\},$$

$$\mathcal{D}(X) = \sigma_-(X), \qquad \mathcal{Q} = \{Q \mid E[Q] = 1, \|Q - \inf Q\|_2 \le 1\},$$

$$\mathcal{Q}_{\mathcal{D}}(X) = \left\{1 - \frac{E[Y] - Y}{\sigma_-(X)}\right\},$$

where $Y = [X - E[X]]_-$, and

$$\mathcal{D}(X) = \mathrm{CVaR}_\alpha^\Delta(X), \qquad \mathcal{Q} = \{Q \mid E[Q] = 1,\ 0 \le Q \le 1/\alpha\}$$

with $\mathcal{Q}_{\mathcal{D}}(X)$ being the set of elements such that $E[Q] = 1$ and

$$Q(\omega)\begin{cases} = \alpha^{-1} & \text{on } \{\omega \mid X(\omega) < -\mathrm{VaR}_\alpha(X)\}, \\ \in [0, \alpha^{-1}] & \text{on } \{\omega \mid X(\omega) = -\mathrm{VaR}_\alpha(X)\}, \\ = 0 & \text{on } \{\omega \mid X(\omega) > -\mathrm{VaR}_\alpha(X)\}. \end{cases}$$

Observe that for σ and σ_-, $\mathcal{Q}_{\mathcal{D}}$ is a *singleton*. For $\mathcal{Q}$ and $\mathcal{Q}_{\mathcal{D}}(X)$ of other deviation measures and for operations with risk envelopes, the reader may refer to [111,71,114].

From the optimization perspective, $\mathcal{Q}_{\mathcal{D}}(X)$ is closely related to *subgradients* of $\mathcal{D}$ at X, which are elements $Z \in \mathcal{L}^2(\Omega)$ such that $\mathcal{D}(Y) \ge \mathcal{D}(X) + E[(Y - X)Z]$ for all $Y \in \mathcal{L}^2(\Omega)$. In fact, Proposition 1 in [114] states that for a deviation measure $\mathcal{D}$, the subgradient set $\partial\mathcal{D}(X)$ at X is related to the risk identifier $\mathcal{Q}_{\mathcal{D}}(X)$ by $\partial\mathcal{D}(X) = 1 - \mathcal{Q}_{\mathcal{D}}(X)$. In general, risk identifiers along with risk envelopes play a central role in formulating optimality conditions and devising optimization procedures in applications involving deviation measures. For example, if X is discretely distributed with $\mathbb{P}\{X = x_k\} = p_k$, $k = 1, \ldots, n$, then with the risk envelope representation, the CVaR-deviation and mixed CVaR-deviation are readily restated in the linear programming form

$$\mathrm{CVaR}_\alpha^\Delta(X) = E[X] - \min_{q_k}\left\{\sum_{k=1}^n q_k p_k x_k \mid 0 \le q_k \le 1/\alpha,\ \sum_{k=1}^n q_k p_k = 1\right\},$$

$$\sum_{i=1}^m \lambda_i \mathrm{CVaR}_{\alpha_i}^\Delta(X) = E[X] - \min_{q_{ik}}\left\{\sum_{i=1,\,k=1}^{m,n} \lambda_i q_{ik} p_k x_k \mid 0 \le q_{ik} \le 1/\alpha_i,\ \sum_{k=1}^n q_{ik} p_k = 1\right\}.$$

5.1.2. Mean-deviation approach to portfolio selection

As an important financial application, Rockafellar et al. [113–115] solved and analyzed a Markowitz-type portfolio selection problem [4,119] with a deviation measure $\mathcal{D}$:

$$\min_{X\in\mathcal{X}} \mathcal{D}(X) \quad \text{s.t. } E[X] \ge r_0 + \Delta,$$

where X is the portfolio rate of return, $\mathcal{X}$ is the set of feasible portfolios, and Δ is the desirable gain over the risk-free rate r_0. For example, if a portfolio has an initial value 1 with the capital portions $x_0, x_1, \ldots, x_n$ allocated into a risk-free instrument with the constant rate of return r_0 and into risky instruments with uncertain rates of return $r_1, \ldots, r_n$, then $\mathcal{X} = \{X | X = \sum_{k=0}^n x_k r_k, \sum_{k=0}^n x_k = 1\}$ and $E[X] = x_0 r_0 + \sum_{k=1}^n x_k E[r_k]$. In this case, the portfolio selection problem reduces to finding optimal weights $(x_0^*, x_1^*, \ldots, x_n^*)$.

Theorem 3 in [113] proves that for the *nonthreshold* (noncritical) values of r_0, there exists a master fund of either positive or negative type having the expected rate of return $r_0 + \Delta^*$ with $\Delta^* > 0$, such that the optimal investment policy is to invest the amount Δ/Δ^* in the master fund and the amount $1 - \Delta/\Delta^*$ in the risk-free instrument when there exists a master fund of positive type and to invest $-\Delta/\Delta^*$ in the master fund and $1 + \Delta/\Delta^*$ in the risk-free instrument when there exists a master fund of negative type. For the *threshold* values of r_0, there exists a master fund of threshold type with zero price, so that in this case, the optimal investment policy is to invest the whole capital in the risk-free instrument and to open a position of magnitude Δ in the master fund through long and short positions. This result generalizes the classical *one fund theorem* [120,121] stated for the case of the standard deviation when a master fund of positive type (market portfolio) exists.

Theorem 5 in [114] shows that conditions on the existence of the master funds introduced in [113] generalize the well-known *capital asset pricing model (CAPM)* [121–123]:

$$E[r_i] - r_0 = \begin{cases} \beta_i(E[X^*] - r_0), & \text{when there exists a master fund of positive type,} \\ \beta_i(E[X^*] + r_0), & \text{when there exists a master fund of negative type,} \\ \beta_i E[X^*], & \text{when there exists a master fund of threshold type,} \end{cases}$$

where X^* is the master fund's rate of return, and

$$\beta_i = \frac{\operatorname{covar}(-r_i, Q^*)}{\mathcal{D}(X^*)}, \qquad Q^* \in \mathcal{Q}(X^*),\ i = 1, \dots, n.$$

For example, $\beta_i = \operatorname{covar}(r_i, X^*)/\sigma^2(X^*)$ for the standard deviation, whereas

$$\beta_i = \frac{\operatorname{covar}(-r_i, [X^* - E[X^*]_-])}{\sigma_-^2(X^*)}$$

for the standard lower semideviation, and

$$\beta_i = \frac{E[(E[r_i] - r_i)Q^*]}{\mathrm{CVaR}_\alpha^\Delta(X^*)}, \quad Q^* \in \mathcal{Q}_{\mathrm{CVaR}_\alpha^\Delta}(X^*)$$

for the CVaR-deviation. When $\mathbb{P}\{X^* = -\mathrm{VaR}_\alpha(X^*)\} = 0$, the last formula can be expressed in terms of conditional probabilities

$$\beta_i = \frac{E[E[r_i] - r_i \mid X^* \le -\mathrm{VaR}_\alpha(X^*)]}{E[E[X^*] - X^* \mid X^* \le -\mathrm{VaR}_\alpha(X^*)]}.$$

It should be mentioned that in general, β's may not be uniquely defined because of either a master fund is not unique or $\mathcal{Q}_{\mathcal{D}}(X^*)$ is not a singleton. For β's with other deviation measures, see [114].

Interpretation of these *CAPM-like* relations in the sense of the classical CAPM relies on the existence of a market equilibrium for investors using a deviation measure other than the standard deviation. Rockafellar et al. [115] proved that indeed, when investors' utility functions depend only on the mean and deviation of portfolio's return and satisfy some additional conditions, the market equilibrium exists even if different groups of investors use different deviation measures. This result justifies viewing of the generalized β's in the classical sense and shows that the CAPM-like relations can also serve as one factor predictive models for expected rates of return of risky instruments.

5.1.3. Chebyshev inequalities with deviation measures

In engineering applications dealing with safety and reliability as well as in the actuarial science, risk if often interpreted as the probability of a dread event or disaster. Minimizing the probability of a highly undesirable event is known as the *safety first principle*, which was originally introduced by Roy [124] in the context of portfolio selection. When the probability distribution function of a random variable X is unknown or very complex, the probability that X falls below a certain threshold ξ can be estimated in terms of the mean $\mu = E[X]$ and variance $\sigma^2(X) < \infty$ of X by the one-sided Chebyshev inequality[6]

$$\mathbb{P}\{X \le \xi\} \le \frac{1}{1 + (\mu - \xi)^2/\sigma^2(X)}, \quad \xi \le \mu.$$

Estimates similar to this one are also used in non-convex decision making problems involving chance constraints [125]. The Chebyshev inequality can be improved if the standard deviation is replaced by another deviation measure.

[6] The two-sided Chebyshev inequality is stated as $\mathbb{P}\{|X - EX| \ge a\} \le \sigma^2(X)/a^2$, $a > 0$.

The problem of generalizing the one-sided Chebyshev inequality for *law-invariant* deviation measures, e.g. σ, σ_-, MAD, $\mathrm{CVaR}_\alpha^\Delta$, etc., is formulated as follows: for law-invariant $\mathcal{D} : \mathcal{L}^p(\Omega) \to [0, \infty]$, $1 \le p < \infty$, find a function $g_{\mathcal{D}}(d)$ such that

$$\mathbb{P}\{X \le \mu - a\} \le g_{\mathcal{D}}(\mathcal{D}(X)) \quad \text{for all } X \in \mathcal{L}^p(\Omega) \text{ and } a > 0 \qquad (76)$$

under the conditions: (i) $g_{\mathcal{D}}$ is independent of the distribution of X; and (ii) $g_{\mathcal{D}}$ is the least upper bound in (76), i.e., for every $d > 0$, there is a random variable X such that (76) becomes the equality with $\mathcal{D}(X) = d$. For the two-sided Chebyshev inequality, the problem is formulated similarly, see [116].

Grechuk et al. [116] showed that (76) reduces to an auxiliary optimization problem

$$u_{\mathcal{D}}(\alpha) = \inf_{X \in \mathcal{L}^p(\Omega)} \mathcal{D}(X)$$
$$\text{s.t. } X \in U = \{X \mid E[X] = 0, \mathbb{P}\{X \le -a\} \ge \alpha\}, \qquad (77)$$

and that the function $g_{\mathcal{D}}$ is determined by

$$g_{\mathcal{D}}(d) = \sup\{\alpha \mid u_{\mathcal{D}}(\alpha) \le d\}.$$

Proposition 3 in [116] proves that (76) is equivalent to minimizing $\mathcal{D}$ over a subset of U, whose elements are undominated random variables with respect to *convex ordering*,[7] and that the later problem reduces to finite parameter optimization.

For the mean absolute deviation, standard lower semideviation, and CVaR-deviation, the one-sided Chebyshev inequality is given by

$$\mathbb{P}\{X \le \xi\} \le \frac{\mathrm{MAD}(X)}{2(\mu - \xi)}, \quad \xi < \mu,$$

$$\mathbb{P}\{X \le \xi\} \le \frac{\sigma_-(X)^2}{(\mu - \xi)^2}, \quad \xi < \mu,$$

$$\mathbb{P}\{X \le \xi\} \le \frac{\alpha\, \mathrm{CVaR}_\alpha^\Delta(X)}{\alpha\, \mathrm{CVaR}_\alpha^\Delta(X) + (1 - \alpha)(\mu - \xi)}, \quad \xi \le \mathrm{CVaR}_\alpha(X).$$

Examples of one-sided and two-sided Chebyshev inequalities with other deviation measures as well as generalizations of the Rao–Blackwell and Kolmogorov inequalities with law-invariant deviation measures are discussed in [116].

5.1.4. Maximum entropy principle with deviation measures

Entropy maximization is a fundamental principle originated from the information theory and statistical mechanics (see [126]) and finds its application in financial engineering and decision making under risk [127–129]. The principle determines the least-informative (or most unbiased) probability distribution for a random variable X given some prior information about X. For example, if only mean μ and variance σ^2 of X are available, e.g. through estimation, the probability distribution with continuous probability density $f_X : \mathbb{R} \mapsto \mathbb{R}_0^+$ that maximizes the Shannon differential entropy

$$S(X) = -\int_{-\infty}^{\infty} f_X(t) \log f_X(t) \mathrm{d}t$$

is the normal distribution with the mean μ and variance σ^2.

Let $\mathcal{X} \subseteq \mathcal{L}^1(\Omega)$ be the set of random variables with continuous probability densities on $\mathbb{R}$. Then the most unbiased probability distribution of a random variable $X \in \mathcal{X}$ with known mean and

[7] X dominates Y with respect to convex ordering if $E[f(X)] \ge E[f(Y)]$ for any convex function $f : \mathbb{R} \mapsto \mathbb{R}$, which is equivalent to the conditions $E[X] = E[Y]$ and $\int_{-\infty}^{x} F_X(t)\mathrm{d}t \le \int_{-\infty}^{x} F_Y(t)\mathrm{d}t$ for all $x \in \mathbb{R}$, where F_X and F_Y are cumulative probability distribution functions of X and Y, respectively.

law-invariant deviation $\mathcal{D} : \mathcal{L}^p(\Omega) \mapsto [0, \infty]$, $p \in [1, \infty]$, of X can be found from the maximum entropy principle:

$$\max_{X \in \mathcal{X}} S(X) \quad \text{s.t. } E[X] = \mu, \quad \mathcal{D}(X) = d. \tag{78}$$

Boltzmann's theorem [130, Theorem 12.1.1] shows that if for given measurable functions $h_1, \ldots, h_n$, constants $a_1, \ldots, a_n$, and a closed support set $V \subseteq \mathbb{R}$, there exist $\lambda_1, \ldots, \lambda_n$, and $c > 0$ such that the probability density function

$$f_X(t) = c \exp\left(\sum_{j=1}^{n} \lambda_j h_j(t)\right), \quad t \in V \tag{79}$$

satisfies the constraints

$$\int_V f_X(t)\mathrm{d}t = 1, \qquad \int_V h_j(t) f_X(t)\mathrm{d}t = a_j, \quad j = 1, \ldots, n, \tag{80}$$

then among all continuous probability density functions on V, (79) maximizes $S(X)$ subject to (80).

With this theorem, solutions to (78) for the standard deviation, mean absolute deviation, standard lower semideviation, and lower range deviation $E[X] - \inf X$ readily follows. For example,

(a) $f_X(t) = \exp(-|t - \mu|/\mathrm{d})/(2\mathrm{d})$ for $\mathcal{D}(X) = \mathrm{MAD}(X)$ and $V = \mathbb{R}$.
(b) $f_X(t) = \exp((\mu - t)/\mathrm{d} - 1)/\mathrm{d}$, $t \geq \mu - \mathrm{d}$, for $\mathcal{D}(X) = E[X] - \inf X$ and $V = [\mu - \mathrm{d}, \infty)$.
(c) $f_X(t) = c\exp(\lambda_1 t + \lambda_2 [t - \mu]_-^2)$ for $\mathcal{D}(X) = \sigma_-(X)$, where c, λ_1, and λ_2 are found from the conditions $\int_{-\infty}^{\infty} f_X(t)\,\mathrm{d}t = 1$, $\int_{-\infty}^{\infty} t\, f_X(t)\,\mathrm{d}t = \mu$, and $\int_{-\infty}^{\mu} (t - \mu)^2 f_X(t)\,\mathrm{d}t = \mathrm{d}$.

However, not all deviation measures can be represented in the form of the constraints in (80). For this case, Grechuk et al. [117] proved that a law-invariant deviation measure $\mathcal{D} : \mathcal{L}^p(\Omega) \mapsto \mathbb{R}$ can be represented in the form

$$\mathcal{D}(X) = \sup_{g(s) \in G} \int_0^1 g(s)\mathrm{d}(q_X(s)), \tag{81}$$

where $q_X(\alpha) = \inf\{t | F_X(t) > \alpha\}$ is the quantile of X, and G is a set of positive concave functions $g : (0, 1) \mapsto \mathbb{R}^+$. If $\mathcal{D}$ is *comonotone*, i.e., $\mathcal{D}(X + Y) = \mathcal{D}(X) + \mathcal{D}(Y)$ for any two comonotone $X \in \mathcal{L}^p(\Omega)$ and $Y \in \mathcal{L}^p(\Omega)$, then G in (81) is a singleton. For example, $\mathrm{CVaR}_\alpha^\Delta(X)$ is comonotone, and its set G has a single function defined by $g(s) = (1/\alpha - 1)s$ for $s \in [0, \alpha]$, and $g(s) = 1 - s$ for $s \in (\alpha, 1]$. With (81) and (78) reduces to a calculus of variations problem, which in the case of comonotone $\mathcal{D}$ has a closed form solution, see [117]. For example, a solution to (78) with $\mathcal{D}(X) = \mathrm{CVaR}_\alpha^\Delta(X)$ is given by $f_X((x - \mu)/d)/d$, where

$$f_X(t) = \begin{cases} (1-\alpha)\exp\left(\dfrac{1-\alpha}{\alpha}\left(t - \dfrac{2\alpha - 1}{1-\alpha}\right)\right), & t \leq \dfrac{2\alpha - 1}{1-\alpha}, \\ (1-\alpha)\exp\left(-\left(t - \dfrac{2\alpha - 1}{1-\alpha}\right)\right), & t \geq \dfrac{2\alpha - 1}{1-\alpha}. \end{cases}$$

Grechuk et al. [117] made the following conclusions:

(i) A solution $X \in \mathcal{X}$ to (78) has a log-concave distribution, i.e., $\ln f_X(t)$ is concave.
(ii) For any log-concave $f_X(t)$, there exists comonotone $\mathcal{D}$ such that a solution to (78) is $f_X(t)$.

Conclusion (ii) solves *the inverse problem*: if agent's solution to (78) is known (estimated) then agent's risk preferences can be recovered from the comonotone deviation measure corresponding to this solution through (78), see [117] for details. Other examples of distributions that maximize either Shannon or Renyi differential entropy subject to constraints on the mean and deviation are discussed in [117].

5.2. Averse measures of risk

Rockafellar et al. [111,71] introduced *averse measures of risk* as functionals $\mathcal{R} : \mathcal{L}^2(\Omega) \to (-\infty; \infty]$ satisfying[8]

(R1) *Risk aversion*: $\mathcal{R}(c) = -c$ for constants c, but $\mathcal{R}(X) > E[-X]$ for nonconstant X.
(R2) *Positive homogeneity*: $\mathcal{R}(\lambda X) = \lambda \mathcal{R}(X)$ when $\lambda > 0$.
(R3) *Subadditivity*: $\mathcal{R}(X + Y) \leq \mathcal{R}(X) + \mathcal{R}(Y)$ for all X and Y.
(R4) *Lower semicontinuity*: set $\{X \in \mathcal{L}^2(\Omega) | \mathcal{R}(X) \leq c\}$ is closed for all $c < \infty$.

Axiom (R1) requires an additional explanation. It follows from $\mathcal{R}(c) = -c$ and (R3) that $\mathcal{R}$ is constant translation invariant, i.e.,

$$\mathcal{R}(X + c) = R(X) - c,$$

see [71]. On the other hand, $\mathcal{R}(c) = -c$ implies $\mathcal{R}(E[X]) = -E[X]$, and $\mathcal{R}(X) > E[-X]$ can be restated as $\mathcal{R}(X) > \mathcal{R}(E[X])$ for $X \neq c$, which is the risk aversion property in terms of $\mathcal{R}$ (a risk-averse agent always prefers $E[X]$ over nonconstant X).

Averse measures of risk and *coherent risk measures* in the sense of [68] (see Section 4) share three main properties: subadditivity, positive homogeneity, and constant translation invariance. The key difference between these two classes of risk measures is that averse measures of risk are not required to be monotone (and the monotonicity axiom (A1) in Section 4 does not follow from (R1)–(R4)), while coherent risk measures are not, in general, risk averse, i.e. do not satisfy (R1). Nevertheless, the axioms of risk aversion and monotonicity are not incompatible, and the two classes have nonempty intersection: *coherent-averse measures of risk*; see [111,71] for details.

Theorem 2 in [71] establishes a one-to-one correspondence between deviation measures and averse measures of risk through the relationships:

$$\mathcal{R}(X) = \mathcal{D}(X) - E[X], \qquad \mathcal{D}(X) = \mathcal{R}(X - E[X]), \tag{82}$$

and shows that $\mathcal{R}$ is a coherent-averse measure of risk if and only if $\mathcal{D}$ is lower range dominated, i.e. satisfies (D5). This result provides a simple recipe for constructing averse measures of risk:

(a) *Risk measures of $\mathcal{L}^p(\Omega)$ type*

$$\mathcal{R}(X) = \lambda \|X - E[X]\|_p - E[X], \quad p \in [1, \infty], \ \lambda > 0,$$

e.g. $\mathcal{R}(X) = \lambda\,\sigma(X) - E[X]$ and $\mathcal{R}(X) = \lambda\,\mathrm{MAD}(X) - E[X]$.

(b) *Risk measures of semi-$\mathcal{L}^p(\Omega)$ type*

$$\mathcal{R}(X) = \lambda \|[X - E[X]]_-\|_p - E[X], \quad p \in [1, \infty], \ \lambda > 0,$$

e.g. $\mathcal{R}(X) = \lambda\,\sigma_-(X) - E[X]$.

(c) *Risk measures of CVaR type*: (i) $\mathcal{R}(X) = \mathrm{CVaR}_\alpha(X)$; (ii) *mixed CVaR*

$$\mathcal{R}(X) = \int_0^1 \mathrm{CVaR}_\alpha(X)\mathrm{d}\lambda(\alpha),$$

where $\int_0^1 \mathrm{d}\lambda(\alpha) = 1$ and $\lambda(\alpha) \geq 0$; and (iii) *worst case mixed CVaR*

$$\mathcal{R}(X) = \sup_{\lambda \in \Lambda} \int_0^1 \mathrm{CVaR}_\alpha(X)\mathrm{d}\lambda(\alpha),$$

where Λ is a set of weighting nonnegative measures λ on $(0, 1)$ with $\int_0^1 \mathrm{d}\lambda(\alpha) = 1$. These measures correspond to the CVaR-deviation, mixed CVaR deviation (74), and worst case mixed CVaR deviation (75), respectively.

[8] In [111], these measures are originally called *strict expectation bounded risk measures*, and then in the subsequent work [109], they are named averse measures of risk to reflect the concept more accurately.

Among these, only risk measures of CVaR type and risk measures of semi-$\mathcal{L}^p(\Omega)$ type with $\lambda \in (0,1]$ are coherent. Also, the mixed CVaR can be equivalently represented in the form (42), see [71, Proposition 5].

Another major implication of Theorem 2 in [71] is that all optimization procedures available for deviation measures can be readily applied to averse measures of risk. In particular, $\mathcal{R}$ and $\mathcal{D}$ corresponding through (82) have the same risk envelope and risk identifier and

$$\mathcal{R}(X) = -\inf_{Q\in\mathcal{Q}} E[XQ],$$

$$\mathcal{Q} = \{Q \in \mathcal{L}^2(\Omega) \mid \mathcal{R}(X) \geq -E[XQ] \text{ for all } X\},$$

where in addition $\mathcal{R}$ is coherent if and only if the corresponding risk envelope $\mathcal{Q}$ satisfies (Q4).

As coherent risk measures, averse measures of risk can be also characterized in terms of *acceptance sets*: a random variable X is accepted or belongs to an acceptance set $\mathcal{A}$ if its risk is nonpositive, i.e. $\mathcal{R}(X) \leq 0$. In view of the property $\mathcal{R}(c) = -c$ for constants c, $\mathcal{R}(X)$ can be interpreted as the minimal cash reserve (possibly negative) making $X + \mathcal{R}(X)$ acceptable. Theorem 2 in [111] shows that there is a one-to-one correspondence between averse measures of risk $\mathcal{R}$ and acceptance sets $\mathcal{A}$:

$$\mathcal{A} = \{X \mid \mathcal{R}(X) \leq 0\}, \qquad \mathcal{R}(X) = \inf\{c \mid X + c \in \mathcal{A}\}, \tag{83}$$

where each $\mathcal{A}$ is a subset of $\mathcal{L}^2(\Omega)$ and satisfies

(A1) $\mathcal{A}$ is closed and contains positive constants c,
(A2) $0 \in \mathcal{A}$, and $\lambda X \in \mathcal{A}$ whenever $X \in \mathcal{A}$ and $\lambda > 0$,
(A3) $X + Y \in \mathcal{A}$ for any $X \in \mathcal{A}$ and $Y \in \mathcal{A}$,
(A4) $E[X] > 0$ for every $X \not\equiv 0$ in $\mathcal{A}$.

In addition, $\mathcal{R}$ is coherent if and only if $\mathcal{A}$ contains all nonnegative X. With this theorem, examples of acceptance sets for averse measures of risk are straightforward:

(a) $\mathcal{A} = \{X | \lambda \, \|X - E[X]\|_p \leq E[X]\}$ for the risk measures of $\mathcal{L}^p(\Omega)$ type with $p \in [1,\infty]$, $\lambda > 0$.
(b) $\mathcal{A} = \{X | \lambda \, \|[X - E[X]]_-\|_p \leq E[X]\}$ for the risk measures of semi-$\mathcal{L}^p(\Omega)$ type with $p \in [1,\infty]$, $\lambda > 0$.
(c) $\mathcal{A} = \{X | \mathrm{CVaR}_\alpha(X) \leq 0\}$ for $\mathcal{R}(X) = \mathrm{CVaR}_\alpha(X)$, $\alpha \in [0,1]$.

In view of (83), Rockafellar et al. [111] interpreted $\mathcal{R}$ as *$\mathcal{A}$-effective infimum* of X: $\mathcal{R}(X) = \mathcal{A}\text{-}\inf X = \inf_{X+c\in\mathcal{A}} c$, and restated $\mathcal{D}$ corresponding to $\mathcal{R}$ through (82) as $\mathcal{D}(X) = E[X] - \mathcal{A}\text{-}\inf X$. This provides an interesting interpretation of $\mathcal{D}$: for each X, $\mathcal{D}(X)$ is the least upper bound of the difference between what is expected and what is accepted under given $\mathcal{A}$. For detailed discussion of these and other issues concerning averse measures of risk, the reader may refer to [111,71].

5.3. Error measures

The third important concept characterizing uncertainty in a random outcome is error measures introduced by Rockafellar et al. [111,71,109] as functionals $\mathcal{E} : \mathcal{L}^2(\Omega) \to [0,\infty]$ satisfying

(E1) *Nonnegativity*: $\mathcal{E}(0) = 0$, but $\mathcal{E}(X) > 0$ for $X \neq c$; also, $\mathcal{E}(c) < \infty$ for constants c.
(E2) *Positive homogeneity*: $\mathcal{E}(\lambda X) = \lambda\mathcal{E}(X)$ when $\lambda > 0$.
(E3) *Subadditivity*: $\mathcal{E}(X + Y) \leq \mathcal{E}(X) + \mathcal{E}(Y)$ for all X and Y.
(E4) *Lower semicontinuity*: set $\{X \in \mathcal{L}^2(\Omega) | \mathcal{E}(X) \leq c\}$ is closed for all $c < \infty$.

Error measures can be viewed as *norms* on $\mathcal{L}^p(\Omega)$, e.g. $\mathcal{E}(X) = \|X\|_2$, however, as deviation measures and averse measures of risk, they are not required to be symmetric $\mathcal{E}(-X) \neq \mathcal{E}(X)$ to allow treating gains and losses differently. An example of *asymmetric* error measure is given by

$$\mathcal{E}_{a,b,p}(X) = \|aX_+ + bX_-\|_p, \quad a \geq 0,\ b \geq 0,\ 1 \leq p \leq \infty, \tag{84}$$

where a and b are not both zero. Observe that for $a = 1$ and $b = 1$, (84) reduces to $\mathcal{L}^p$ norms $\|X\|_p$, whereas for $a = 1,\ b = 0$ and $a = 0,\ b = 1$, it simplifies to $\|X_+\|_p$ and $\|X_-\|_p$, respectively. Another example is the *asymmetric mean absolute error* (72) discussed in [110] in the context of the quantile regression.

Functionals $\mathcal{D}$, $\mathcal{R}$, and $\mathcal{E}$ share the same three properties: positive homogeneity, subadditivity, and lower semicontinuity. The only difference comes from axioms (D1), (R1), and (E1) on how the functionals treat constants. In fact, any two of (D1), (R1), and (E1) are incompatible, i.e. there is no functional satisfying any two of these axioms. Unlike the relationships (82), there is no one-to-one correspondence between deviation measures and error measures. Nevertheless, a simple relationship between these two classes can be established through *penalties relative to expectation*

$$\mathcal{D}(X) = \mathcal{E}(X - E[X]). \tag{85}$$

The relationship (85) is only a particular example of such a correspondence. Another subclass of deviation measures can be obtained from error measures by *error projection*, which in the case of infinite dimensional $\mathcal{L}^2(\Omega)$ requires an additional assumption on $\mathcal{E}$.

An error measure $\mathcal{E}$ is *nondegenerate* if there exists $\delta > 0$ such that $\mathcal{E}(X) \geq \delta\,|E[X]|$ for all X. For example, the asymmetric mean absolute error (72) is nondegenerate, whereas $\mathcal{E}_{a,b,p}(X)$ is nondegenerate for $a > 0, b > 0, 1 \leq p \leq \infty$ with $\delta = \min\{a,b\}$; see [109]. Theorem 2.1 in [109] proves that for a nondegenerate error measure $\mathcal{E}$,

$$\mathcal{D}(X) = \inf_{c\in\mathbb{R}} \mathcal{E}(X - c) \tag{86}$$

is the deviation measure called *the deviation of X projected from $\mathcal{E}$*, and

$$\mathcal{S}(X) = \arg\min_{c\in\mathbb{R}} \mathcal{E}(X - c) \tag{87}$$

is *the statistics of X associated with $\mathcal{E}$*. In general, $\mathcal{S}(X)$ is an interval $[\mathcal{S}^-(X), \mathcal{S}^+(X)]$ of constants such that $\mathcal{S}^-(X) = \min\{c | c \in \mathcal{S}(X)\}$ and $\mathcal{S}^+(X) = \max\{c | c \in \mathcal{S}(X)\}$.

Well-known examples of the relationships (86) and (87) include

$$\mathcal{E}(X) = \|X\|_2, \quad \mathcal{D}(X) = \|X - E[X]\|_2 = \sigma(X), \quad \mathcal{S}(X) = E[X],$$
$$\mathcal{E}(X) = \|X\|_1, \quad \mathcal{D}(X) = \|X - \mathrm{med}\,(X)\|_1, \quad \mathcal{S}(X) = \mathrm{med}\,(X),$$

where med (X) is the median of X (possibly an interval), and

$$\mathcal{E}_\alpha(X) = E[X_+ + (\alpha^{-1} - 1)X_-], \qquad \mathcal{D}(X) = \mathrm{CVaR}^\Delta_\alpha(X),$$
$$\mathcal{S}(X) = [q^-_\alpha(X), q^+_\alpha(X)],$$

where $q^-_\alpha(X) = \inf\{t | F_X(t) \geq \alpha\}$ and $q^+_\alpha(X) = \sup\{t | F_X(t) \leq \alpha\}$ with $F_X(t)$ being the cumulative probability distribution function of X. Observe that for $\mathcal{E}(X) = \|X\|_2$, deviations (85) and (86) coincide, whereas for $\mathcal{E}(X) = \|X\|_1$, they are different.

Theorem 2.2 in [109] proves that if for $k = 1,\ldots,n$, $\mathcal{D}_k$ is a measure of deviation, and $\mathcal{E}_k$ is a nondegenerate measure of error that projects to $\mathcal{D}_k$, then, for any weights $\lambda_k > 0$ with $\sum_{k=1}^n \lambda_k = 1$,

$$\mathcal{E}(X) = \inf_{\substack{C_1,\ldots,C_n \\ \lambda_1 C_1+\cdots+\lambda_n C_n=0}} \{\lambda_1\mathcal{E}_1(X - C_1) + \cdots + \lambda_r\mathcal{E}_n(X - C_n)\}$$

defines a nondegenerate measure of error which projects to the deviation measure

$$\mathcal{D}(X) = \lambda_1\mathcal{D}_1(X) + \cdots + \lambda_n\mathcal{D}_n(X)$$

with the associated statistic

$$\mathcal{S}(X) = \lambda_1 \mathcal{S}_1(X) + \cdots + \lambda_n \mathcal{S}_n(X).$$

An immediate consequence of this remarkable result is that for any choice of probability thresholds $\alpha_k \in (0, 1)$ and weights $\lambda_k > 0$ with $\sum_{k=1}^n \lambda_k = 1$,

$$\mathcal{E}(X) = E[X] + \inf_{\substack{C_1,\ldots,C_n \\ \lambda_1 C_1 + \cdots + \lambda_n C_n = 0}} \left\{ \frac{\lambda_1}{\alpha_1} E[\max\{0, C_1 - X\}] + \cdots + \frac{\lambda_n}{\alpha_n} E[\max\{0, C_n - X\}] \right\}$$

is a nondegenerate error measure which projects to the mixed CVaR deviation measure $\mathcal{D}$ in (73) with the associated statistic

$$\mathcal{S}(X) = \lambda_1 q_{\alpha_1}(X) + \cdots + \lambda_n q_{\alpha_n}(X),$$
$$q_{\alpha_k}(X) = [q^-_{\alpha_k}(X), q^+_{\alpha_k}(X)].$$

Example 2.5 in [109] shows that for a given deviation measure $\mathcal{D}$, a nondegenerate error measure can be obtained by *inverse projection*

$$\mathcal{E}(X) = \mathcal{D}(X) + |E[X]|,$$

which through (86) projects back to $\mathcal{D}$ with the associated statistics $\mathcal{S}(X) = E[X]$. Consequently, there could be more than one error measure projecting to the same deviation measure, e.g. $\mathcal{E}(X) = \|X\|_2$ and $\mathcal{E}(X) = \|X - E[X]\|_2 + |E[X]|$ both project to $\mathcal{D}(X) = \sigma(X)$, and an arbitrary nondegenerate error measure $\mathcal{E}$ can be modified as $\mathcal{E}'(X) = \inf_{c\in\mathbb{R}} \mathcal{E}(X - c) + |E[X]| \equiv \mathcal{E}(X - E[X]) + |E[X]|$ to have $E[X]$ as the associated statistics.

It is left to mention that for a given error measure $\mathcal{E}(X)$, the representations (85) and (86) along with the relationships (82) provide two ways for constructing (different) averse measures of risk

$$\mathcal{R}(X) = \mathcal{E}(X - E[X]) - E[X], \qquad \mathcal{R}(X) = \inf_{c\in\mathbb{R}} \mathcal{E}(X - c) - E[X].$$

Remarkably, for the asymmetric mean absolute error (72), the second formula can be restated as $\mathcal{R}(X) = \inf_{c\in\mathbb{R}} E[\alpha^{-1}[X - c]_- - c)$, which coincides with the well-known optimization formula (53) for CVaR. This finishes the discussion about the relationships between three classes of measures $\mathcal{D}$, $\mathcal{R}$, and $\mathcal{E}$. For other examples of such relationships, in particular for the error measure corresponding to the mixed CVaR-deviation (73), see [111,71,109].

One of the important applications of error measures in risk analysis, statistics, and decision making under uncertainty is a generalized linear regression: *approximate a random variable* $Y \in \mathcal{L}^2(\Omega)$ *by a linear combination* $c_0 + c_1X_1 + \cdots + c_nX_n$ *of given random variables* $X_k \in \mathcal{L}^2(\Omega)$, $k = 1, \ldots, n$, i.e., minimize the error $Z(c_0, c_1, \ldots, c_n) = Y - (c_0 + c_1X_1 + \cdots + c_nX_n)$ with respect to $c_0, c_1, \ldots, c_n$:

$$\min_{c_0,c_1,\ldots,c_n} \mathcal{E}(Z(c_0, c_1, \ldots, c_n)). \tag{88}$$

Observe that because of possible asymmetry of $\mathcal{E}$, $\mathcal{E}(-Z) \neq \mathcal{E}(Z)$.

Theorem 3.2 in [109] proves that error minimization (88) can be decomposed into

$$\min_{c_1,\ldots,c_n} \mathcal{D}\left(Y - \sum_{k=1}^n c_kX_k\right) \quad \text{and} \quad c_0 \in \mathcal{S}\left(Y - \sum_{k=1}^n c_kX_k\right), \tag{89}$$

where $\mathcal{D}$ is the deviation of X projected from $\mathcal{E}$, and $\mathcal{S}$ is the statistics of X associated with $\mathcal{E}$. As an immediate consequence of this important result, we obtain the following examples:

(a) Classical linear regression (least squares) $\min_{c_0,c_1,\ldots,c_n} \|Y - (c_0 + \sum_{k=1}^n c_kX_k)\|_2$ is equivalent to

$$\min_{c_1,\ldots,c_n} \sigma\left(Y - \sum_{k=1}^n c_kX_k\right) \quad \text{and} \quad c_0 = E\left[Y - \sum_{k=1}^n c_kX_k\right].$$

(b) Median regression $\min_{c_0,c_1,\ldots,c_n} \left\|Y - \left(c_0 + \sum_{k=1}^n c_kX_k\right)\right\|_1$ is equivalent to

$$\min_{c_1,\ldots,c_n} E\left|Y - \sum_{k=1}^n c_kX_k - \operatorname{med}\left(Y - \sum_{k=1}^n c_kX_k\right)\right| \quad \text{and}$$

$$c_0 = \operatorname{med}\left(Y - \sum_{k=1}^n c_kX_k\right).$$

(c) Quantile regression $\min_{c_0,c_1,\ldots,c_n} E[Z(c_0, c_1, \ldots, c_n)_+ + (\alpha^{-1} - 1)Z(c_0, c_1, \ldots, c_n)_-]$, $\alpha \in (0, 1)$, reduces to

$$\min_{c_1,\ldots,c_n} \mathrm{CVaR}^{\Delta}_\alpha\left(Y - \sum_{k=1}^n c_kX_k\right) \quad \text{and}$$

$$c_0 = -\mathrm{VaR}_\alpha\left(Y - \sum_{k=1}^n c_kX_k\right).$$

Example (a) confirms the well-known fact that the least squares regression is equivalent to minimizing variance of $Y - \sum_{k=1}^n c_kX_k$ with the constant term c_0 (intercept) set to the mean of $Y - \sum_{k=1}^n c_kX_k$, whereas Example (b) shows that the linear regression with $\mathcal{E}(\cdot) = \|\cdot\|_1$ does not reduce to minimization of the mean absolute deviation and that c_0 is not the mean of $Y - \sum_{k=1}^n c_kX_k$. The theory of error measures elucidates that this is possible in Example (a) because for $\mathcal{E}(\cdot) = \|\cdot\|_2$, the deviation from the penalties relative to expectation, i.e. (85), coincides with the deviation from error projection, i.e. with (86). Examples of the linear regression with other error measures, including the so-called *mixed quantile regression*, *risk-acceptable regression*, and *unbiased linear regression* with general deviation measures, as well as optimality conditions for (89) are available in [109].

References

[1] A. Prékopa, Stochastic Programming, Kluwer Academic Publishers, 1995.
[2] D. Bernoulli, Specimen teoriae novae de mensura sortis, Commentarii Academiae Scientarium Imperialis Petropolitanae V (1738) 175–192. English translation: exposition of a new theory on the measurement of risk, Econometrica, 22 (1954) 23–36.
[3] J. von Neumann, O. Morgenstern, Theory of Games and Economic Behavior, Princeton University Press, Princeton, NJ, 1953, edition.
[4] H.M. Markowitz, Portfolio selection, Journal of Finance 7 (1952) 77–91.
[5] J.R. Birge, F. Louveaux, Introduction to Stochastic Programming, Springer, New York, 1997.
[6] P. Kall, J. Mayer, Stochastic Linear Programming: Models, Theory, and Computation, Springer, 2005.
[7] A. Shapiro, D. Dentcheva, A. Ruszczyński, Lectures on Stochastic Programming: Modeling and Theory, SIAM, Philadelphia, PA, 2009.
[8] P.C. Fishburn, Utility Theory for Decision-Making, Wiley, New York, 1970.
[9] P.C. Fishburn, Non-Linear Preference and Utility Theory, Johns Hopkins University Press, Baltimore, 1988.
[10] E. Karni, D. Schmeidler, Utility theory with uncertainty, in: Hildenbrand, Sonnenschein (Eds.), Handbook of Mathematical Economics, vol. IV, North-Holland, Amsterdam, 1991.
[11] H. Föllmer, A. Schied, Stochastic Finance: An Introduction in Discrete Time, 2nd ed., Walter de Gruyter, Berlin, 2004.
[12] J.P. Quirk, R. Saposnik, Admissibility and measurable utility functions, The Review of Economic Studies 29 (1962) 140–146.
[13] P.C. Fishburn, Decision and Value Theory, John Wiley & Sons, New York, 1964.
[14] J. Hadar, W.R. Russell, Rules for ordering uncertain prospects, The American Economic Review 59 (1969) 25–34.
[15] H. Levy, Stochastic dominance and expected utility: survey and analysis, Management Science 38 (1992) 555–593.
[16] A. Müller, D. Stoyan, Comparison Methods for Stochastic Models and Risks, John Wiley & Sons, Chichester, 2002.

[17] M. Rothschild, J. Stiglitz, Increasing risk I: a definition, Journal of Economic Theory 2 (1970) 225–243.
[18] J. Quiggin, A theory of anticipated utility, Journal of Economic Behavior and Organization 3 (1982) 225–243.
[19] J. Quiggin, Generalized Expected Utility Theory—The Rank-Dependent Expected Utility Model, Kluwer Academic Publishers, Dordrecht, 1993.
[20] M.E. Yaari, The dual theory of choice under risk, Econometrica 55 (1987) 95–115.
[21] M. Allais, Le comportement de l'homme rationnel devant le risque: critique des postulats et axiomes de l'école américaine, Econometrica 21 (1953) 503–546.
[22] D. Ellsberg, Risk, ambiguity, and the savage axioms, Quarterly Journal of Economics 75 (1961) 643–669.
[23] G. Choquet, Theory of capacities, Annales de l'Institut Fourier Grenoble 5 (1955) 131–295.
[24] D. Schmeidler, Integral representation without additivity, Proceedings of the Anerican Mathematical Society 97 (1986) 255–261.
[25] D. Schmeidler, Subjective probability and expected utility without additivity, Econometrica 57 (1989) 571–587.
[26] D. Dentcheva, A. Ruszczyński, Optimization with stochastic dominance constraints, SIAM Journal on Optimization 14 (2003) 548–566.
[27] D. Dentcheva, A. Ruszczyński, Semi-infinite probabilistic optimization: first order stochastic dominance constraints, Optimization 53 (2004) 583–601.
[28] J. Luedtke, New formulations for optimization under stochastic dominance constraints, SIAM Journal on Optimization 19 (2008) 1433–1450.
[29] W.K. Klein Haneveld, M.H. van der Vlerk, Integrated chance constraints: reduced forms and an algorithm, Computational Management Science 3 (2006) 245–269.
[30] C.I. Fábián, G. Mitra, D. Roman, Processing second-order stochastic dominance models using cutting-plane representations, Mathematical Programming (2009).
[31] W. Ogryczak, A. Ruszczyński, Dual stochastic dominance and related mean-risk models, SIAM Journal on Optimization 13 (2002) 60–78.
[32] D. Dentcheva, A. Ruszczyński, Inverse stochastic dominance constraints and rank dependent expected utility theory, Mathematical Programming 108 (2006) 297–311.
[33] D. Dentcheva, A. Ruszczyński, Duality between coherent risk measures and stochastic dominance constraints in risk-averse optimization, Pacific Journal of Optimization 4 (2008) 433–446.
[34] D. Dentcheva, A. Ruszczyński, Optimality and duality theory for stochastic optimization problems with nonlinear dominance constraints, Mathematical Programming 99 (2004) 329–350.
[35] D. Dentcheva, A. Ruszczyński, Robust stochastic dominance and its application to risk-averse optimization, Mathematical Programming 123 (2010) 85–100.
[36] D. Dentcheva, A. Ruszczyński, Portfolio optimization with stochastic dominance constraints, in: Risk Management and Optimization in Finance, Journal of Banking & Finance 30 (2006) 433–451.
[37] D. Roman, K. Darby-Dowman, G. Mitra, Portfolio construction based on stochastic dominance and target return distributions, Mathematical Programming 108 (2006) 541–569.
[38] H.M. Markowitz, Portfolio Selection, Wiley and Sons, New York, 1959.
[39] M.C. Steinbach, Markowitz revisited: mean–variance models in financial portfolio analysis, SIAM Review 43 (2001) 31–85.
[40] P. Krokhmal, S. Uryasev, G. Zrazhevsky, Risk management for hedge fund portfolios: a comparative analysis of linear rebalancing strategies, Journal of Alternative Investments 5 (2002) 10–29.
[41] R.S. Dembo, D. Rosen, The practice of portfolio replication: a practical overview of forward and inverse problems, Annals of Operations Research 85 (1999) 267–284.
[42] H.M. Markowitz, Mean-Variance Analysis in Portfolio Choice and Capital Markets, Blackwell, Oxford, 1987.
[43] W. Ogryczak, A. Ruszczyński, From stochastic dominance to mean-risk models: semideviations as risk measures, European Journal of Operational Research 116 (1999) 33–50.
[44] W. Ogryczak, A. Ruszczyński, On consistency of stochastic dominance and mean-semideviation models, Mathematical Programming 89 (2001) 217–232.
[45] C. Testuri, S. Uryasev, On relation between expected regret and conditional value-at-risk, in: Z. Rachev (Ed.), Handbook of Numerical Methods in Finance, Birkhauser, 2004, pp. 361–373.
[46] V.S. Bawa, Optimal rules for ordering uncertain prospects, Journal of Financial Economics 2 (1975) 95–121.
[47] P.C. Fishburn, Mean-risk analysis with risk associated with below-target returns, The American Economic Review 67 (1977) 116–126.
[48] R.B. Porter, Semivariance and stochastic dominance: a comparison, American Economic Review 64 (1974) 200–204.
[49] W.K. Klein Haneveld, Duality in Stochastic Linear and Dynamic Programming, in: Lecture Notes in Economics and Mathematical Systems, vol. 274, Springer, Berlin, 1986.
[50] M.R. Young, A minimax portfolio selection rule with linear programming solution, Management Science 44 (1998) 673–683.
[51] P. Kouvelis, G. Yu, Robust Discrete Optimization and its Applications, Kluwer Academic Publishers, Dodrecht, 1997.
[52] JP Morgan, Riskmetrics, JP Morgan, New York, 1994.
[53] P. Jorion, Value at Risk: The New Benchmark for Controlling Market Risk, McGraw-Hill, 1997.
[54] D. Duffie, J. Pan, An overview of value-at-risk, Journal of Derivatives 4 (1997) 7–49.
[55] R.T. Rockafellar, S. Uryasev, Conditional value-at-risk for general loss distributions, Journal of Banking & Finance 26 (2002) 1443–1471.
[56] D. Tasche, Expected shortfall and beyond, Journal of Banking & Finance 26 (2002) 1519–1533.
[57] K. Inui, M. Kijima, On the significance of expected shortfall as a coherent risk measure, in: Risk Measurement, Journal of Banking & Finance 29 (2005) 853–864.
[58] G. Pflug, Some remarks on the value-at-risk and the conditional value-at-risk, in: S. Uryasev (Ed.), Probabilistic Constrained Optimization: Methodology and Applications, Kluwer Academic Publishers, Dordrecht, 2000, pp. 272–281.
[59] S.S. Wang, V.R. Young, H.H. Panjer, Axiomatic characterization of insurance prices, Insurance: Mathematics and Economics 21 (1997) 173–183. In Honor of Prof. J.A. Beekman.
[60] P. Embrechts (Ed.), Extremes and Integrated Risk Management, Risk Books, London, 2000.
[61] A.C.W.W. Cooper, G.H. Symonds, Cost horizons and certainty equivalents: an approach to stochastic programming of heating oil, Management Science 4 (1958) 235–263.
[62] M. Rausand, A. Høyland, System Reliability Theory: Models, Statistical Methods, and Applications, 2nd ed., Wiley, Hoboken, NJ, 2004.
[63] B. Epstein, I. Weissman, Mathematical Models for Systems Reliability, CRC Press, Boca Raton, FL, 2008.
[64] S.K. Choi, R.V. Grandhi, R.A. Canfield, Reliability-Based Structural Design, Springer, London, 2007.
[65] D. Dentcheva, Optimization models with probabilistic constraints, in: G. Calafiore, F. Dabbene (Eds.), Probabilistic and Randomized Methods for Design under Uncertainty, Springer, London, 2006, pp. 49–98.
[66] M.H. van der Vlerk, Integrated chance constraints in an alm model for pension funds. Working paper, 2003.
[67] A. Nemirovski, A. Shapiro, Convex approximations of chance constrained programs, SIAM Journal on Optimization 17 (2006) 969–996.
[68] P. Artzner, F. Delbaen, J.M. Eber, D. Heath, Coherent measures of risk, Mathematical Finance 9 (1999) 203–228.
[69] H. Föllmer, A. Schied, Robust preferences and convex measures of risk, in: K. Sandmann, P.J. Schönbucher (Eds.), Advances in Finance and Stochastics: Essays in Honour of Dieter Sondermann, Springer, 2002, pp. 39–56.
[70] A. Ruszczyński, A. Shapiro, Optimization of convex risk functions, Mathematics of Operations Research 31 (2006) 433–452.
[71] R.T. Rockafellar, S. Uryasev, M. Zabarankin, Generalized deviations in risk analysis, Finance and Stochastics 10 (2006) 51–74.
[72] F. Delbaen, Coherent risk measures on general probability spaces, in: K. Sandmann, P.J. Schönbucher (Eds.), Advances in Finance and Stochastics: Essays in Honour of Dieter Sondermann, Springer, 2002, pp. 1–37.
[73] P. Cheridito, F. Delbaen, M. Kupper, Coherent and convex monetary risk measures for bounded càdlàg processes, Stochastic Processes and their Applications 112 (2004) 1–22.
[74] E. De Giorgi, Reward-risk portfolio selection and stochastic dominance, Journal of Banking & Finance 29 (2005) 895–926.
[75] G.C. Pflug, Subdifferential representations of risk measures, Mathematical Programming 108 (2006) 339–354.
[76] P. Krokhmal, Higher moment coherent risk measures, Quantitative Finance 7 (2007) 373–387.
[77] M. Frittelli, E.R. Gianin, Putting order in risk measures, Journal of Banking & Finance 26 (2002) 1473–1486.
[78] R.T. Rockafellar, Convex Analysis, in: Princeton Mathematics, vol. 28, Princeton University Press, 1970.
[79] C. Zălinescu, Convex Analysis in General Vector Spaces, World Scientific, Singapore, 2002.
[80] S. Kusuoka, On law invariant risk measures, Advances in Mathematical Economics 3 (2001) 83–95.
[81] S. Kusuoka, A remark on law invariant convex risk measures, Advances in Mathematical Economics 10 (2007) 91–100.
[82] M. Frittelli, E.R. Gianin, Law invariant convex risk measures, Advances in Mathematical Economics 7 (2005) 33–46.
[83] R.A. Dana, A representation result for concave Schur concave functions, Mathematical Finance 15 (2005) 613–634.
[84] C. Acerbi, Spectral measures of risk: a coherent representation of subjective risk aversion, Journal of Banking & Finance 26 (2002) 1487–1503.
[85] E. Jouini, M. Meddeb, N. Touzi, Vector-valued coherent risk measures, Finance and Stochastics 8 (2004) 531–552.
[86] A.H. Hamel, A duality theory for set-valued functions I: Fenchel conjugation theory, Set-Valued and Variational Analysis 17 (2009) 153–182.
[87] P. Artzner, F. Delbaen, J.M. Eber, D. Heath, H. Ku, (2002). Coherent multiperiod risk measurement, prprint.
[88] P. Artzner, F. Delbaen, J.M. Eber, D. Heath, H. Ku, Coherent multiperiod risk adjusted values and bellman's principle, Annals of Operations Research 152 (2007) 5–22.
[89] A. Ruszczyński, A. Shapiro, Conditional risk mappings, Mathematics of Operations Research 31 (2006) 544–561.
[90] R.T. Rockafellar, S. Uryasev, Optimization of conditional value-at-risk, Journal of Risk 2 (2000) 21–41.

[91] C. Acerbi, D. Tasche, On the coherence of expected shortfall, Journal of Banking & Finance 26 (2002) 1487–1503.
[92] A.S. Cherny, Weighted V@R and its properties, Finance and Stochastics 10 (2006) 367–393.
[93] P. Krokhmal, J. Palmquist, S. Uryasev, Portfolio optimization with conditional value-at-risk objective and constraints, Journal of Risk 4 (2002) 43–68.
[94] A. Künzi-Bay, J. Mayer, Computational aspects of minimizing conditional value-at-risk, Computational Management Science 3 (2006) 3–27.
[95] S. Alexander, T. Coleman, Y. Li, Minimizing cvar and var for a portfolio of derivatives, Journal of Banking & Finance 30 (2006) 583–605.
[96] C. Lim, H.D. Sherali, S. Uryasev, Portfolio optimization by minimizing conditional value-at-risk via nondifferentiable optimization, Computation Optimization and Applications (2008).
[97] R.T. Rockafellar, J. Royset, On buffered failure probability in design and optimization of structures, Reliability Engineering & System Safety 95 (2010) 499–510.
[98] G. Chen, M. Daskin, M. Shen, S. Uryasev, The α-reliable mean-excess regret model for stochastic facility location modeling, Naval Research Logistics 53 (2006) 617–626.
[99] A. Ben-Tal, M. Teboulle, An old–new concept of convex risk measures: an optimized certainty equivalent, Mathematical Finance 17 (2007) 449–476.
[100] A. Ben-Tal, M. Teboulle, Expected utility, penalty functions, and duality in stochastic nonlinear programming, Management Science 32 (1986) 1145–1466.
[101] P. Cheridito, T. Li, Risk measures on Orlicz hearts, Mathematical Finance 19 (2009) 189–214.
[102] T. Fischer, Risk capital allocation by coherent risk measures based on one-sided moments, Insurance: Mathematics and Economics 32 (2003) 135–146.
[103] P. Krokhmal, P. Soberanis, Risk optimization with p-order conic constraints: a linear programming approach, European Journal of Operational Research 201 (2010) 653–671.
[104] A. Ben-Tal, M. Teboulle, Portfolio theory for the recourse certainty equivalent maximizing investor, Annals of Operations Research 31 (1991) 479–499.
[105] I. Csiszar, Information-type measures of difference of probability distributions and indirect observations, Studia Scientiarum Mathematicarum Hungarica 2 (1967) 299–318.
[106] S. Kullback, R.A. Leibler, On information and sufficiency, The Annals of Mathematical Statistics 22 (1951) 79–86.
[107] H. Föllmer, A. Schied, Convex measures of risk and trading constraints, Finance and Stochastics 6 (2002) 429–447.
[108] A. Eichhorn, W. Römisch, Polyhedral measures of risk in stochastic programming, SIAM Journal on Optimization 16 (2005) 69–95.
[109] R.T. Rockafellar, S. Uryasev, M. Zabarankin, Risk tuning with generalized linear regression, Mathematics of Operations Research 33 (2008) 712–729.
[110] R. Koenker, G. Bassett, Regression quantiles, Econometrica 46 (1978) 33–50.
[111] R.T. Rockafellar, S. Uryasev, M. Zabarankin, 2002. Deviation measures in risk analysis and optimization. Technical Report 2002-7. ISE Department, University of Florida. Gainesville, FL.
[112] R.T. Rockafellar, Coherent approaches to risk in optimization under uncertainty, in: Tutorials in Operations Research INFORMS 2007, INFORMS, 2007, pp. 38–61.
[113] R.T. Rockafellar, S. Uryasev, M. Zabarankin, Master funds in portfolio analysis with general deviation measures, Journal of Banking & Finance 30 (2006) 743–778.
[114] R.T. Rockafellar, S. Uryasev, M. Zabarankin, Optimality conditions in portfolio analysis with general deviation measures, Mathematical Programming 108 (2006) 515–540.
[115] R.T. Rockafellar, S. Uryasev, M. Zabarankin, Equilibrium with investors using a diversity of deviation measures, Journal of Banking & Finance 31 (2007) 3251–3268.
[116] B. Grechuk, A. Molyboha, M. Zabarankin, Chebyshev's inequalities with law invariant deviation measures, Probability in the Engineering and Informational Sciences 24 (2010) 145–170.
[117] B. Grechuk, A. Molyboha, M. Zabarankin, Maximum entropy principle with general deviation measures, Mathematics of Operations Research 34 (2009) 445–467.
[118] A. Roell, Risk aversion in Quiggin and Yaari's rank-order model of choice under uncertainty, The Economic Journal, Issue Supplement: Conference papers 97 (1987) 143–159.
[119] H.M. Markowitz, Foundations of portfolio theory, Journal of Finance 46 (1991) 469–477.
[120] J. Tobin, Liquidity preference as behavior towards risk, The Review of Economic Studies 25 (1958) 65–86.
[121] W.F. Sharpe, Capital asset prices: a theory of market equilibrium under conditions of risk, Journal of Finance 19 (1964) 425–442.
[122] W.F. Sharpe, Capital asset prices with and without negative holdings, Journal of Finance 46 (1991) 489–509.
[123] R.R. Grauer, Introduction to asset pricing theory and tests, in: R. Roll (Ed.), The International Library of Critical Writings in Financial Economics, Edward Elgar Publishing Inc., 2001.
[124] A.D. Roy, Safety first and the holding of assets, Econometrica 20 (1952) 431–449.
[125] P. Bonami, M.A. Lejeune, An exact solution approach for portfolio optimization problems under stochastic and integer constraints, Operations Research 57 (2009) 650–670.
[126] E.T. Jaynes, Information theory and statistical mechanics, Physical Review 106 (1957) 620–630.
[127] J.M. Cozzolino, M.J. Zahner, The maximum-entropy distribution of the future market price of a stock, Operations Research 21 (1973) 1200–1211.
[128] M.U. Thomas, A generalized maximum entropy principle, Operations Research 27 (1979) 1188–1196.
[129] J.J. Buckley, Entropy principles in decision making under risk, Risk Analysis 5 (1979) 303–313.
[130] T.M. Cover, J.A. Thomas, Elements of Information Theory, 2nd ed., Wiley, New York, 2006.