DEVIATION MEASURES IN RISK ANALYSIS AND OPTIMIZATION

***R. Tyrrell Rockafellar*[1], *Stanislav Uryasev*[2], *Michael Zabarankin*[2]**

RESEARCH REPORT # 2002-7

Risk Management and Financial Engineering Lab
Center for Applied Optimization
Department of Industrial and Systems Engineering
University of Florida, Gainesville, FL 32611

Version: June 8, 2003

Correspondence should be addressed to: Stanislav Uryasev

Abstract

General deviation measures, which include standard deviation as a special case but need not be symmetric with respect to ups and downs, are defined and shown to correspond to risk measures in the sense of Artzner, Delbaen, Eber and Heath when those are applied to the difference between a random variable and its expectation, instead of to the random variable itself. A property called expectation-boundedness of the risk measure is uncovered as essential for this correspondence. It is shown to be satisfied by conditional value-at-risk and by worst-case risk, as well as various mixtures, although not by ordinary value-at-risk.

Interpretations are developed in which inequalities that are “acceptably sure”, relative to a designated acceptance set, replace inequalities that are “almost sure” in the usual sense being violated only with probability zero. Acceptably sure inequalities fix the standard for a particular choice of a deviation measure. This is explored in examples that rely on duality with an associated risk envelope, comprised of alternative probability densities.

The role of deviation measures and risk measures in optimization is analyzed, and the possible influence of “acceptably free lunches” is thereby brought out. Optimality conditions based on concepts of convex analysis, but relying on the special features of risk envelopes, are derived in support of a variety of potential applications, such as portfolio optimization and variants of linear regression in statistics.

Keywords: *risk management, deviation measures, coherent risk measures, value-at-risk, conditional value-at-risk, portfolio optimization, convex analysis.*

[1]University of Washington, Department of Mathematics, Box 354350, Seattle, WA 98195-4350;
E-mail: *rtr@math.washington.edu*. Research supported by NSF grant *DMS-0104055*
[2]University of Florida, ISE Department, P.O. Box 116595, 303 Weil Hall, Gainesville, FL 32611-6595;
E-mails: *uryasev@ufl.edu*, *zabarank@ufl.edu*; URLs: *www.ise.ufl.edu/uryasev*, *http://plaza.ufl.edu/zabarank*

1 Introduction

The minimization of variance, or equivalently standard deviation, is a universally familiar feature of classical portfolio theory. It has been subjected to criticism, however, because standard deviation does not adequately account for the phenomenon of "fat tails" in loss distributions, and moreover penalizes ups and downs equally. In everyday applications, a more common tool than standard deviation is value-at-risk, VaR. But that too has been controversial because of mathematical shortcomings (lack of convexity and monotonicity, as well as reasonable continuity) and its inability to respond to the magnitude of the possible losses below the threshold it identifies. A related concept, conditional value-at-risk, CVaR (also known by other names such as expected shortfall and tail-VaR in some contexts), has proved to be superior in these respects and therefore more suited for the optimization of choices in financial management.

For sound methodology in finance, a theory relating these different approaches and allowing their effects to be compared is essential. A very important step toward such a comprehensive theory was made in the paper of Artzner, Delbaen, Eber and Heath [3], which introduced the notion of a *coherent risk measure*, demonstrating that VaR did not provide coherency. Although extensive motivation was given in [3], "coherency" has not fully taken hold in the community concerned with applications. One obstacle has been the axiom in [3] that deals with the effect of adding a constant to the return of a financial random variable. Many have balked at that axiom, despite its explanation in [3].

The reason for this difficulty may well be confusion over the meaning of "loss". In [3], loss refers to a *negative outcome*, whereas for many practitioners it is assumed to refer to a *shortfall relative to expectation*. This is compounded by the fact that, in applications, VaR and CVaR are typically invoked to measure such shortfall, instead of absolute loss itself.

There is more to this than might appear on the surface. On a space of random variables X, it is one thing to apply a risk measure to $X - C$ for some fixed constant C, and quite another to apply it to $X - EX$. A basic aim of this paper is to make clear that risk measures, more or less in the sense of [3], when applied to $X - EX$, yield a separate class functionals that can aptly be called *deviation measures*. Deviation measures satisfy different axioms and have their own significance, with standard deviation just being one example that happens to be symmetric.

The first part of the paper is devoted to laying this out and providing current examples, aimed at enhancing the theory of risk measures and deviation measures simultaneously. A distinction between our effort and the one undertaken in [3] is that we do not limit our attention to coherent risk measures only. Although we believe in the virtues of coherency, the deviation measures corresponding to coherent risk measures fail to include standard deviation, an indispensable example even if flawed for some purposes in finance. On the other hand, not all coherent risk measures yield deviation measures that make sense. A key new concept has to be brought in: that of an *expectation-bounded* risk measure.

Next in the paper, we review the connection between risk measures and "acceptance sets" such as were introduced in [3], but in doing so we again go beyond the territory of coherency. Moreover we furnish fresh interpretations that add support to the use of risk measures and deviation measures by relating them in an elementary, yet conceptually fruitful manner to "sureness valuations". We emphasize in this way that "almost sure" inequalities, which are customary in probability, might well be replaced in some situations in finance by "acceptably sure" inequalities. Such weaker inequalities could lead to fresh developments in other directions. Under coherency, they correspond to a tempered form of worst-case analysis of the future and can readily be employed in optimization.

In the final part of the paper, we draw general consequences from this theory for problems of minimizing risk or deviation. Contrary to what many people seem to take for granted, these two kinds of minimization need not always turn out to be the same. The situations in which they may diverge are in fact ones where a new kind of "free lunch" appears that, for a particular investor, may at least be acceptably sure, if not almost sure in the usual probabilistic sense.

We go on then to lay down rules for optimality in a general setting for minimization. These rules involve the "subgradients" and "normals" of convex analysis and support a broad range of applications. We identify the subgradients in relation to risk envelopes and the normals in the case of linear constraints.

2 Deviation and Risk

We begin by axiomatizing what we mean by a deviation measure, and then go on to explain how deviation measures are related to risk measures.

We consider a state space Ω, the elements ω of which represent future states, or individual scenarios (perhaps just finitely many), and suppose it to be supplied with a probability measure P and the other technicalities that make it a legitimate probability space (namely a field $\mathcal{M}$ of sets, the measurable sets). The probability measure P stands for the "true" distribution of future states, but it also could be a pricing measure derived from market considerations, or a representation of the subjective probabilities. Whatever the interpretation, the mathematics will be the same.

We treat as random variables (r.v.'s) the functions $X : \Omega \to I\!R$ that belong to the linear space $\mathcal{L}^2 = \mathcal{L}^2(\Omega, \mathcal{M}, P)$, i.e., the (measurable) functions for which the mean and variance

$$\begin{aligned} &\mu(X) = EX = \textstyle\int_\Omega X(\omega)dP(\omega), \\ &\sigma^2(X) = E[X - EX]^2 = \textstyle\int_\Omega [X(\omega) - \mu(X)]^2 dP(\omega), \end{aligned}$$

exist (i.e., these integrals are well defined). The inner product in $\mathcal{L}^2$ between any X and Y is $E[XY]$, which can be identified with $\operatorname{covar}(X,Y) + \mu(X)\mu(Y)$.

In particular, of course, this space contains all constant r.v.'s, $X \equiv C$. To assist in working with such r.v.'s, the letter C will always stand for a constant in the real numbers $I\!R$, and when C or a number like 0, 1, or -1 appears in the position of some r.v. X, it will signify the r.v. has that constant value almost surely. Similarly, an inequality like $X \geq C$ or $X \leq C$, or $X_1 \leq X_2$ is to be viewed in the sense of holding almost surely. The *essential* infimum and supremum of X will be denoted simply by $\inf X$ and $\sup X$:

$$\inf X = \begin{cases} \text{highest } C \text{ such that } C \leq X & \text{if a constant } C \leq X \text{ exists,} \\ -\infty & \text{otherwise,} \end{cases}$$

$$\sup X = \begin{cases} \text{lowest } C \text{ such that } C \geq X & \text{if a constant } C \geq X \text{ exists,} \\ \infty & \text{otherwise.} \end{cases}$$

The subspace of $\mathcal{L}^2$ consisting of all X such that $EX = 0$ will be denoted by $\mathcal{L}^2_0$; it is the orthogonal complement of the one-dimensional subspace of $\mathcal{L}^2$ consisting of the constant r.v.'s

Definition 1 (general deviation measures). *By a deviation measure on $\mathcal{L}^2$ will be meant any functional $\mathcal{D} : \mathcal{L}^2 \to [0, \infty]$ satisfying*

(D1) $\mathcal{D}(X + C) = \mathcal{D}(X)$ *for all X and constants C,*
(D2) $\mathcal{D}(0) = 0$, *and* $\mathcal{D}(\lambda X) = \lambda \mathcal{D}(X)$ *for all X and all $\lambda > 0$,*
(D3) $\mathcal{D}(X + X') \leq \mathcal{D}(X) + \mathcal{D}(X')$ *for all X and X',*
(D4) $\mathcal{D}(X) > 0$ *for all nonconstant X, whereas $\mathcal{D}(X) = 0$ for constant X.*

Under these axioms, $\mathcal{D}(X)$ depends only on $X - EX$ (from the case of D1 where $C = -EX$), and it vanishes only if $X - EX \equiv 0$ (as seen from D4 with $X - EX$ in place of X; note that the property in D4 for constant X already follows from D1 and D2). Thus, $\mathcal{D}$ measures the degree of *uncertainty* in X. It acts as a sort of norm on the "pure uncertainty" subspace $\mathcal{L}_0^2$ of $\mathcal{L}^2$, except that the symmetry required by the definition of a norm — the additional condition that $\mathcal{D}(-X) = \mathcal{D}(X)$ for all X — may be absent. In allowing asymmetry, Definition 1 builds on the wide-spread recognition that symmetry is not always desirable in dealing with uncertainty, since downside results may not be viewed in the same way as upside results in applications. In any case, if $\mathcal{D}$ is a deviation measure then so too is its *reflection* $\tilde{\mathcal{D}}$ and its *symmetrization* $\tilde{\tilde{\mathcal{D}}}$, given by

$$\tilde{\mathcal{D}}(X) = \mathcal{D}(-X), \qquad \tilde{\tilde{\mathcal{D}}}(X) = \tfrac{1}{2}[\mathcal{D}(X) + \tilde{\mathcal{D}}(X)].$$

Axiom D2 is *positive homogeneity*. The combination of D2 with D3 is the property known as *sublinearity*. It implies that $\mathcal{D}$ is a *convex* functional on $\mathcal{L}^2$.

Definition 1 allows a deviation measure $\mathcal{D}$ to have $\mathcal{D}(X) = \infty$ for some r.v.'s X. When this is excluded, we have a *finite* deviation measure. In D2 and D3, infinite values are to be handled in the obvious way: $\alpha + \infty = \infty$ for any $\alpha \in (-\infty, \infty]$, and $\lambda\infty = \infty$ for any $\lambda > 0$, whereas $0\infty = 0$. (These conventions are standard in convex analysis.)

Standard deviation furnishes an immediate example: for $\mathcal{D}(X) = \sigma(X)$, all the properties D1, D2, D3 and D4 hold. This deviation measure is symmetric. The standard *semi*deviations $\sigma_+(X)$ and $\sigma_-(X)$, where $\sigma_+(X) = (E[\max\{X - \mu(X), 0\}^2])^{1/2}$ and $\sigma_-(X) = (E[\max\{\mu(X) - X, 0\}^2])^{1/2}$, satisfy D1, D2, D3 and D4 as well, but are not symmetric. Other examples, symmetric and asymmetric, will be described after more background on "risk".

Although deviation measures are designed for use in risk analysis, they are not "risk measures" in the sense proposed by Artzner, Delbaen, Eber and Heath [3] in their landmark paper. The connection between deviation measures and risk measures is close, but a crucial distinction must be appreciated clearly. Instead of measuring the uncertainty in X, in the sense of nonconstancy, a risk measure evaluates the "overall seriousness of possible losses" associated with X, where *a loss is an outcome below* 0, in contrast to a gain, which is an outcome above 0. In applying a risk measure, this orientation is crucial; if the concern is over the extent to which a given r.v. X might have outcomes $X(\omega)$ that drop below a threshold C, one needs to replace X by $X - C$.

The risk measures we need to consider in relation to deviation measures are based very much on the ideas of Artzner et al. in [3] and follow their axioms, but with the difference that we do not insist on the property of monotonicity and emphasize instead a new property of "expectation-boundedness".

Definition 2 (expectation-bounded risk measures)**.** *By an expectation-bounded risk measure on* $\mathcal{L}^2$ *will be meant any functional* $\mathcal{R} : \mathcal{L}^2 \to (-\infty, \infty]$ *satisfying*

(R1) $\mathcal{R}(X + C) = \mathcal{R}(X) - C$ *for all* X *and constants* C,

(R2) $\mathcal{R}(0) = 0$, *and* $\mathcal{R}(\lambda X) = \lambda\mathcal{R}(X)$ *for all* X *and all* $\lambda > 0$,

(R3) $\mathcal{R}(X + X') \leq \mathcal{R}(X) + \mathcal{R}(X')$ *for all* X *and* X',

(R4) $\mathcal{R}(X) > E[-X]$ *for all nonconstant* X, *whereas* $\mathcal{R}(X) = E[-X]$ *for constant* X.

An expectation-bounded risk measure is said to be coherent if it satisfies the additional axiom

(R5) $\mathcal{R}(X) \leq \mathcal{R}(X')$ *when* $X \geq X'$.

Once more, R2 is *positive homogeneity*, R3 is *subadditivity*, and the combination of R2 and R3 is *sublinearity*, implying *convexity*. Again too, $\mathcal{R}(X) = \infty$ is allowed (under the arithmetic conventions already noted, which now affect R1 as well as R2 and R3). When $\mathcal{R}(X) < \infty$ for all X, we have a *finite* expectation-bounded risk measure on $\mathcal{L}^2$.

The property in R4 is what we specifically refer to as *expectation-boundedness*. From R1, it is already clear that $\mathcal{R}(X) = E[-X]$ for constant X, so added condition is really just the inequality $\mathcal{R}(X) > E[-X]$ for nonconstant X. Expectation-boundedness has not previously been identified as fundamentally important for a risk measure, and was not contemplated in Artzner et al. in [3], where in fact there was no base probability distribution for the states in Ω. We will return to this point later.

Property R5 is *monotonicity*. A functional $\mathcal{R}$ satisfying R1, R2, R3 and R5 (although not necessarily R4) is a *coherent risk measure* as conceived in [3]. Actually, only finite risk measures were considered in [3], and axiom R1 was slightly different, involving investment in a reference r.v., and moreover the state space Ω was assumed to be a finite set. Extensions to the present context and formulation were made, however, in follow-up work by Delbaen [5].

The rationale behind the concept of a coherent risk measure was very well argued in [3]. We will elaborate on the notion below, in explaining the meaning of our expectation-boundedness condition R4 in that context. For now, we proceed directly to the key mathematical consequence of expectation-boundedness.

Although the R2 and R3 for a risk measure agree with the D2 and D3 for a deviation measure, R1 and D1 are entirely different, in fact mutually incompatible — no functional on $\mathcal{L}^2$ can satisfy both R1 and D1. Despite this, there is a simple relationship between the two notions.

Theorem 1 (deviation versus risk)**.** *Deviation measures correspond one-to-one with expectation-bounded risk measures under the relations*

(a) $\mathcal{D}(X) = \mathcal{R}(X - EX)$,

(b) $\mathcal{R}(X) = \mathcal{D}(X) - EX$.

Specifically, if $\mathcal{R}$ is an expectation-bounded risk measure and $\mathcal{D}$ is defined by (a), then $\mathcal{D}$ is a deviation measure that yields back $\mathcal{R}$ through (b). On the other hand, if $\mathcal{D}$ is any deviation measure and $\mathcal{R}$ is defined by (b), then $\mathcal{R}$ is a risk measure that yields back $\mathcal{D}$ through (a). In this correspondence, $\mathcal{R}$ is coherent if and only if $\mathcal{D}$ has the further property that

(D5) $D(X) \leq EX - \inf X$ *for all* X.

Proof. In passing from $\mathcal{R}$ to $\mathcal{D}$ by way of (a), axiom D1 is immediate (since $\mathcal{R}([X + C] - E[X + C]) = \mathcal{R}(X - EX)$; axioms D2 and D3 follow from R2 and R3, while D4 comes out of R4. Because $\mathcal{R}(X - EX) = \mathcal{R}(X) + EX$ by R1, we also get (b). On the other hand, in passing from $\mathcal{D}$ to $\mathcal{R}$ by way of (b), the properties in R2 and R3 are immediate from D2 and D3. We have

$$\mathcal{R}(X + C) = \mathcal{D}(X + C) - E[X + C] = \mathcal{D}(X) - EX - C = \mathcal{R}(X) - C$$

via D1, so that $\mathcal{R}$ satisfies R1. That also shows that (a) will give back $\mathcal{D}$. Axiom D4 clearly corresponds to R4 through (b).

In the presence of R2 and R3, the monotonicity property R5 implies that $\mathcal{R}(X) \leq 0$ when $X \geq 0$ and in fact it is equivalent to that seemingly weaker property, since if $X \geq X'$ we have $X = X' + X''$ for $X'' \geq 0$ and consequently $\mathcal{R}(X) \leq \mathcal{R}(X') + \mathcal{R}(X'')$, hence $\mathcal{R}(X) \leq \mathcal{R}(X')$ when $\mathcal{R}(X'') \leq 0$. In the (a)(b) correspondence, the condition that $\mathcal{R}(X) \leq 0$ when $X \geq 0$ comes out as requiring $\mathcal{D}(X) \leq EX$ when $X \geq 0$. That is definitely true under D5, but in turn it actually guarantees D5. Indeed, to say that $\mathcal{D}(X) \leq EX - \inf X$ is to say that $\mathcal{D}(X) \leq EX - C$ whenever $X \geq C$, and through D1 that is the same as having $\mathcal{D}(X - C) \leq E[X - C]$ whenever $X - C \geq 0$. □

Under the (a)(b) correspondence in Theorem 1, $\mathcal{D}$ will be called the *deviation measure associated with* $\mathcal{R}$, whereas $\mathcal{R}$ will be the *risk measure associated with* $\mathcal{D}$. Note that $\mathcal{D}$ is finite if and only if $\mathcal{R}$ is finite. Because of the theorem's final assertion, it makes sense to speak of $\mathcal{D}$ as a *coherent* deviation measure when D5 is satisfied along with D1, D2, D3 and D4.

Next we run through some fundamental examples of functionals that satisfy the axioms in Definitions 1 or 2, fully or just in part.

Example 1 (risk measures from variance). *For any $\rho > 0$, a finite, expectation-bounded risk measure is furnished by*

$$\mathcal{R}(X) = \rho(E[X - EX]^2)^{1/2} - EX = \rho\sigma(X) - \mu(X).$$

It corresponds to the deviation measure $\mathcal{D}(X) = \rho\sigma(X)$. For $\rho = 0$, one gets $\mathcal{R}(X) = -EX$, which satisfies all the axioms except it lacks the strict inequality demanded by the expectation-boundedness condition in R4. Indeed the corresponding functional $\mathcal{D}$ then is simply $\mathcal{D}(X) \equiv 0$, which, although it satisfies D1, D2 and D3, totally fails D4.

Detail. The expectation-boundedness comes from Theorem 1: the standard deviation measure σ satisfies the axioms of Definition 1, and these properties carry over then to $\rho\sigma$ for any $\rho > 0$. The properties claimed for $\mathcal{R}(X) = -EX$ are evident. □

The case $\rho = 0$ in Example 1 is worth pinpointing because it is the basis of so much of the treatment of risk in stochastic programming and other approaches to optimization under uncertainty, even though it has long been felt inadequate in portfolio optimization.

Example 2 (deviation measures from range). *An example of a deviation measure $\mathcal{D}(X)$ that can take on ∞ is*

$$\mathcal{D}(X) = EX - \inf X = \sup[EX - X],$$

which gives the size of the lower range of X. It is coherent and corresponds to the worst-loss risk measure

$$\mathcal{R}(X) = -\inf X = \sup[-X],$$

which likewise can take on ∞ but is expectation-bounded and coherent.

Another example is furnished by the reflection $\tilde{\mathcal{D}}(X) = E[-X] - \inf[-X] = \sup[X - EX]$, which gives the size of the upper range of X and corresponds to $\tilde{\mathcal{R}}(X) = \sup X - 2EX$, and still another by the symmetrization $\tilde{\tilde{\mathcal{D}}}(X) = \frac{1}{2}[\sup X - \inf X]$, which corresponds to $\tilde{\tilde{\mathcal{R}}}(X) = \frac{1}{2}[\sup X - \inf X] - EX$. In contrast to $\mathcal{R}$, however, the risk measures $\tilde{\mathcal{R}}$ and $\tilde{\tilde{\mathcal{R}}}$, although expectation-bounded, are not coherent, and thus neither $\tilde{\mathcal{D}}$ nor $\tilde{\tilde{\mathcal{D}}}$ is coherent.

Detail. The verification of D1, D2, D3 and D4 for $\mathcal{D}(X) = EX - \inf X$ is elementary. These properties then hold automatically for the reflection $\tilde{\mathcal{D}}$ and symmetrization $\tilde{\tilde{\mathcal{D}}}$, and by Theorem 1 we obtain in turn that the corresponding risk measures $\mathcal{R}$, $\tilde{\mathcal{R}}$ and $\tilde{\tilde{\mathcal{R}}}$ satisfy R1, R2, R3 and R4. Note that the formulas provided for $\tilde{\mathcal{R}}$ and $\tilde{\tilde{\mathcal{R}}}$ do conform to having $\tilde{\mathcal{R}}(X) = \tilde{\mathcal{D}}(X) - EX$ and $\tilde{\tilde{\mathcal{R}}}(X) = \tilde{\tilde{\mathcal{D}}}(X) - EX$.

The coherency of $\mathcal{D}(X) = EX - \inf X$ is immediate from the criterion provided by D5. The D5 condition cannot hold for $\tilde{\mathcal{D}}(X) = \sup X - EX$, because it would imply that $\sup X - EX \leq EX - \inf X$ for all $X \in \mathcal{L}^2$, which is is false. Similarly, the symmetrized deviation measure lacks coherency. □

This example, of evident interest, indicates some of the motivation for not insisting on the finiteness of risk and deviation measures. When the state space Ω is finite, these measures are indeed finite and could well be utilized in some applications. If we made finiteness part of Definitions 1 and 2, we would be in the awkward position of not being able to speak of them without that restriction on Ω.

Example 3 (VaR, value at risk)**.** *For any level* $\alpha \in (0,1)$, *the quantile functional*

$$\mathcal{R}(X) = \mathrm{VaR}_\alpha(X) = -\inf\{\, z \,|\, P\{X \le z\} > \alpha\}$$

is finite and satisfies R1, R2, R5, but not the expectation-boundedness condition R4. Moreover it lacks the subadditivity in R3. The corresponding functional

$$\mathcal{D}(X) = \mathrm{VaR}_\alpha^\Delta(X) = \mathrm{VaR}_\alpha(X - EX)$$

satisfies D1 and D2, but not D3 or D5, and it can sometimes have $\mathcal{D}(X) = 0$ *for nonconstant* X, *thereby failing D4.*

Detail. The misbehavior of $\mathcal{R} = \mathrm{VaR}_\alpha$ with respect to R3, although it satisfies R1, R2 and R5, was noted in [3] and served as one of the main incentives in that work. The failure of $\mathcal{D} = \mathrm{VaR}_\alpha^\Delta$ to satisfy D4 corresponds to the failure of $\mathcal{R} = \mathrm{VaR}_\alpha$ to satisfy R4, which is seen as follows.

Since $EX = -\mathrm{VaR}_{0.5}(X)$ for symmetric distributions, the expectation-boundedness of $\mathcal{R} = \mathrm{VaR}_\alpha$ would require for such distributions, if nonconstant, that $-\mathrm{VaR}_\alpha(X) < -\mathrm{VaR}_{0.5}(X)$. The trouble is that $\mathrm{VaR}_\alpha(X)$, although nondecreasing with respect to α, can have intervals of constancy. Specifically, it is possible to have $\mathrm{VaR}_\alpha(X) = \mathrm{VaR}_{0.5}(X)$ for a range of α values around 0.5 if X has a probability atom at EX. Then one can get $-\mathrm{VaR}_\alpha(X) = EX$ even though X is not constant. □

It should be noted that when the distribution function Ψ for X has an interval of constancy at the α level, $-\mathrm{VaR}_\alpha(X)$ gives upper endpoint, marking the so-called *upper* α*-quantile*, in contrast to the lower endpoint, marking the *lower* α*-quantile*. (In our previous papers [11] and [12], we worked with "loss" r.v.'s X, for which positive outcomes represented losses. In adjusting here to r.v.'s X for which negative outcomes represent losses, we reflect the pictures in those papers from right to left.)

A robust substitute for VaR in assessing r.v.'s X has been developed by taking into account, along with $\mathrm{VaR}_\alpha(X)$, the extent of losses $-X$ that might be suffered when $X < -\mathrm{VaR}_\alpha(X)$. It relies on the notion of the *lower* α*-tail* of X, defined precisely as follows through distributions:

$$\begin{cases} \text{in terms of the distribution function } \Psi \text{ for } X, \text{ the } \textit{lower } \alpha\textit{-tail} \text{ of } X \\ \text{is the r.v. for which the distribution function is } \Psi_\alpha = \alpha^{-1}\min\{\alpha, \Psi\}. \end{cases}$$

Thus, the distribution function for the lower α-tail is obtained by truncating the distribution function from above at the α level and then rescaling so that the vertical range fits with $[0,1]$ instead of just $[0,\alpha]$. (This makes that function Ψ_α be identically 1 on the interval $[-\mathrm{VaR}_\alpha(X), \infty)$ and even on the interval from the lower α-quantile to ∞ when the upper and lower α-quantiles for X are different.)

This definition of the lower α-tail of X, developed in our paper [12], properly addresses with the circumstance that the distribution of X might have a probability atom at the point $-\mathrm{VaR}_\alpha(X)$, corresponding to a jump discontinuity in the cumulative distribution function. (The definition effectively "splits" that atom when present, adjoining just enough of it to erase any gap between α and the possibly lower value of $P\{\, X \,|\, X < -\mathrm{VaR}_\alpha(X)\}$. Further details and interpretations can be found in [12]. When there is no atom at $-\mathrm{VaR}_\alpha(X)$ (as is typical for applications in which X is not a discrete r.v.), the lower α-tail corresponds to the *conditional probability distribution* for $X \le -\mathrm{VaR}_\alpha(X)$.

Example 4 (CVaR, conditional value-at-risk)**.** *For any* $\alpha \in (0,1)$, *the functional*

$$\mathcal{R}(X) = \mathrm{CVaR}_\alpha(X) = -[\text{expectation of the lower } \alpha\text{-tail of } X]$$

(that expectation being same as the conditional expectation of X *subject to* $X \le -\mathrm{VaR}_\alpha(X)$ *when there is zero probability of having* $X = -\mathrm{VaR}_\alpha(X)$*) is a finite, expectation-bounded risk measure, which moreover is coherent and satisfies* $\mathrm{CVaR}_\alpha(X) \ge \mathrm{VaR}_\alpha(X)$. *Accordingly, the functional*

$$\mathcal{D}(X) = \mathrm{CVaR}_\alpha^\Delta(X) = \mathrm{CVaR}_\alpha(X - EX)$$

is a finite, coherent deviation measure. It is not symmetric.

For fixed X, both $\mathrm{CVaR}_\alpha(X)$ *and* $\mathrm{CVaR}^\Delta_\alpha(X)$ *depend continuously on α as a parameter and are nondecreasing with respect to α, moreover with*

$$\lim_{\alpha\to 0}\mathrm{CVaR}_\alpha(X) = -\inf X, \qquad \lim_{\alpha\to 1}\mathrm{CVaR}_\alpha(X) = -EX,$$

$$\lim_{\alpha\to 0}\mathrm{CVaR}^\Delta_\alpha(X) = EX - \inf X, \qquad \lim_{\alpha\to 1}\mathrm{CVaR}^\Delta_\alpha(X) = 0.$$

Detail. The fact that $\mathrm{CVaR}_\alpha(X)$, when based on the definition of lower α-tail articulated above, yields a functional $\mathcal{R}$ that satisfies R1, R2, R3 and R5, is known from our paper [12]. (It would not be true without the indicated "atom splitting" when probability atoms can be present.) To verify that R4 holds for this $\mathcal{R}$, we note that since $-\mathrm{CVaR}_\alpha(X)$ is by definition the expectation of the lower α-tail of X, it cannot exceed the expectation of X with respect to the entire distribution, namely EX, and must be less for nonconstant X. The claims about $\mathcal{D} = \mathrm{CVaR}^\Delta_\alpha$ follow then from Theorem 1.

The continuity of $\mathrm{CVaR}_\alpha(X)$ with respect to α was proven in [12]. The monotonicity with respect to α is evident from its definition, as are the limit statements. The corresponding properties of $\mathrm{CVaR}^\Delta_\alpha(X)$ are evident then from the fact that $\mathrm{CVaR}^\Delta_\alpha(X) = \mathrm{CVaR}_\alpha(X) + EX$. □

The limit relations in Example 4 leads us to define

$$\mathrm{CVaR}_0(X) = -\inf X, \qquad \mathrm{CVaR}_1(X) = -EX.$$

It must be remembered, however, that the second of these is *not* an expectation-bounded risk measure because it fails the *strict* inequalities in R4. The VaR definition beyond the case of $\alpha \in (0,1)$ in Example 3 that harmonizes with this is

$$\mathrm{VaR}_0(X) = -\inf X, \qquad \mathrm{VaR}_1(X) = -\sup X.$$

The distinction between $\mathrm{VaR}_\alpha(X)$ and $\mathrm{VaR}^\Delta_\alpha(X)$, or $\mathrm{CVaR}_\alpha(X)$ and $\mathrm{CVaR}^\Delta_\alpha(X)$, needs emphasis. Sometimes in finance, the "VaR" and "CVaR" of X are taken to refer to $\mathrm{VaR}^\Delta_\alpha(X)$ and $\mathrm{CVaR}^\Delta_\alpha(X)$ instead of $\mathrm{VaR}_\alpha(X)$ and $\mathrm{CVaR}_\alpha(X)$. This is inconsistent with the functionals VaR_α and CVaR_α being risk measures in the sense of Definition 2 — which corresponds instead to the defining formulas in Examples 3 and 4. Without caution being exercised in this direction, there could be mathematical ambiguity and confusion. That could be avoided by calling only $\mathrm{VaR}_\alpha(X)$ and $\mathrm{CVaR}_\alpha(X)$ the "VaR" and "CVaR" of X (for a specified α), and speaking of $\mathrm{VaR}^\Delta_\alpha(X)$ and $\mathrm{CVaR}^\Delta_\alpha(X)$ as the "VaR-deviation" and "CVaR-deviation" of X.

A major advantage of the CVaR-based measures in an optimization context comes from the *minimization formula*

$$\mathrm{CVaR}_\alpha(X) = \min_C \Big\{ C - \alpha^{-1} E[\min\{X + C, 0\}] \Big\},$$

which we developed in [11], concentrating on r.v.'s with continuous distributions, and subsequently extended in [12] to general r.v.'s by means of our α-tail concept. In the interim, Pflug [7] explored the right side of our minimization formula, showing that it could be used to define a coherent risk measure, but without providing a connection between that measure and the original notion of CVaR.

A side benefit of the minimization formula for $\mathrm{CVaR}_\alpha(X)$ is that it also calculates $\mathrm{VaR}_\alpha(X)$ as one of the values of C for which the minimum is attained:

$$\mathrm{VaR}_\alpha(X) \in \operatorname*{argmin}_C \Big\{ C - \alpha^{-1} E[\min\{X + C, 0\}] \Big\}.$$

If the minimum is attained at a unique C, that C must therefore equal $\mathrm{VaR}_\alpha(X)$; in the case of nonuniqueness, the attainment set is an interval having $\mathrm{VaR}_\alpha(X)$ as its left endpoint (cf. [11], [12]).

A different track has been followed by Acerbi [1], in observing that CVaR can equivalently be expressed as a VaR average through the integral formula

$$\mathrm{CVaR}_\alpha(X) = \frac{1}{\alpha}\int_0^\alpha \mathrm{VaR}_p(X)dp.$$

(Acerbi began by using the expression on the right as the definition of a functional he called "expected shortfall," but subsequently realized its connection with CVaR.) Since $\mathrm{VaR}_\alpha(X)$ is nonincreasing and left-continuous as a function of α for fixed X, this integral formula also implies, by the way, that $\alpha\mathrm{CVaR}_\alpha(X)$ is concave as a function of α and thus possesses right and left derivatives with respect to α, the left derivative being

$$\frac{d^-}{d\alpha}[\alpha\mathrm{CVaR}_\alpha(X)] = \mathrm{VaR}_\alpha(X).$$

This further reveals that $\mathrm{CVaR}_\alpha(X)$ itself has right and left derivatives with respect to α, with

$$\frac{d^-}{d\alpha}\mathrm{CVaR}_\alpha(X) = \alpha^{-1}[\mathrm{VaR}_\alpha(X) - \mathrm{CVaR}_\alpha(X)].$$

Acerbi further showed in [1] that his integral formula for CVaR extends to provide "spectral representations" for the risk measures in the much broader class that can be generated by introducing CVaR mixtures.

Example 5 (mixed CVaR and risk profile)**.** *For any weighting measure $d\lambda$ on $(0,1)$ (nonnegative with total measure 1) such that $\int_0^1 p^{-1}d\lambda(p) < \infty$, the mixed* CVaR *functional*

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

is a finite, coherent, expectation-bounded risk measure. It can equally be expressed by the spectral formula

$$\mathcal{R}(X) = \int_0^1 \mathrm{VaR}_\alpha(X)\varphi(\alpha)d\alpha$$

for a function φ on $(0,1)$ which is said to provide the corresponding risk profile, namely

$$\varphi(\alpha) = \int_{[\alpha,1)} p^{-1}d\lambda(p).$$

This function φ is left-continuous and nonincreasing with $\varphi(0^+) < \infty$, $\varphi(1^-) = 0$ and $\int_0^1 \varphi(\alpha)d\alpha = 1$; conversely, any function φ with those properties arises from a unique choice of $d\lambda$ as described. The associated coherent deviation measure, likewise finite, has the formula

$$\mathcal{D}(X) = \int_0^1 \mathrm{CVaR}^\Delta_\alpha(X)d\lambda(\alpha) = \int_0^1 \mathrm{VaR}^\Delta_\alpha(X)\varphi(\alpha)d\alpha.$$

In particular, in taking $d\lambda$ to be comprised solely of atoms having weights λ_i at points α_i for $i = 1, \ldots, m$, with $\lambda_i > 0$ and $\lambda_1 + \cdots + \lambda_m = 1$, one gets the weighted CVaR *sums*

$$\mathcal{R}(X) = \lambda_1\mathrm{CVaR}_{\alpha_1}(X) + \cdots + \lambda_m\mathrm{CVaR}_{\alpha_m}(X),$$

$$\mathcal{D}(X) = \lambda_1\mathrm{CVaR}^\Delta_{\alpha_1}(X) + \cdots + \lambda_m\mathrm{CVaR}^\Delta_{\alpha_m}(X).$$

In that case the risk profile function is $\varphi(\alpha) = \Sigma_{\alpha_i \geq \alpha} \lambda_i$.

Detail. The fact that R1, R2, R3 and R4 carry over from the individual risk measures CVaR_α to the mixture determined by the weighting measure is easily verified directly. The finiteness of $\mathcal{R}$ will emerge as a by-product of the analysis of the risk profile function φ.

The formula for φ yields immediately the fact that φ is left-continuous and nonincreasing on $(0,1)$ with boundary limits $\varphi(1^-) = 0$ and $\varphi(0^+) < \infty$; the finiteness of $\varphi(0^+)$ comes from the assumption that $\int_0^1 p^{-1} d\lambda(p) < \infty$. The Radon-Stieltjes measure $d\varphi$ derived from φ relates to $d\lambda$ by $d\varphi(\alpha) = -\alpha^{-1} d\lambda(\alpha)$.

Conversely, for any function φ that is left-continuous and nonincreasing on $(0,1)$ with $\varphi(1^-) = 0$ and $\varphi(0^+) < \infty$, consider the measure $d\lambda$ on (0,1) defined by $d\lambda(\alpha) = -\alpha d\varphi(\alpha)$. This is nonnegative with

$$\int_0^1 p^{-1} d\lambda(p) = -\int_0^1 p^{-1} p d\varphi(p) = -[\varphi(1^-) - \varphi(0^+)] = \varphi(0^+) < \infty.$$

To establish that $\int_0^1 d\lambda(p) = 1$ as well, we appeal to integration-by-parts (cf. [4, Prop. 8.5.5] for this rule in a Radon-Stieltjes framework):

$$\int_0^1 d\lambda(p) = -\int_0^1 p d\varphi(p) = -[p\varphi(p)]_{0^+}^{1^-} + \int_0^1 \varphi(p) dp,$$

where the boundary expressions both vanish (because $\varphi(0^+) < \infty$ and $\varphi(1^-) = 0$) and the final integral has been assumed to equal 1.

In preparation for confirming the spectral formula for $\mathcal{R}(X)$, we introduce $\psi(\alpha) = \alpha \mathrm{CVaR}_\alpha(X)$. On the basis of Acerbi's integral formula, which entails the integrability of $\mathrm{VaR}_\alpha(X)$ with respect to α, this function also has the expression $\psi(\alpha) = \int_0^\alpha \mathrm{VaR}_\alpha(X) d\alpha$. That implies that ψ is continuous and nonincreasing on $(0,1)$ with $\psi(0^+) = 0$, $\psi(1^-) = E[-X]$ and $d\psi(\alpha) = \mathrm{VaR}_\alpha(X) d\alpha$. Utilizing integration-by-parts once more, along with the relation $d\lambda(\alpha) = -\alpha d\varphi(\alpha)$, we see that

$$\begin{aligned} \textstyle\int_0^1 \mathrm{CVaR}_\alpha(X) d\lambda(\alpha) &= -\textstyle\int_0^1 [\alpha \mathrm{CVaR}_\alpha(X)] d\varphi(\alpha) = -\int_0^1 \psi(\alpha) d\varphi(\alpha) \\ &= -\psi(1^-)\varphi(1^-) + \psi(0^+)\varphi(0^+) + \textstyle\int_0^1 \varphi(\alpha) d\psi(\alpha) = \int_0^1 \varphi(\alpha) \mathrm{VaR}_\alpha(X) d\alpha. \end{aligned}$$

Thus, the spectral formula is correct. The integrability of $\mathrm{VaR}_\alpha(X)$ with respect to α, along with fact that φ is nonnegative with finite upper bound $\varphi(0^+)$, ensures moreover that $\int_0^1 \varphi(\alpha) \mathrm{VaR}_\alpha(X) d\alpha < \infty$, and hence that $\int_0^1 \mathrm{CVaR}_\alpha(X) d\lambda(\alpha) < \infty$. ◻

Functionals directly defined by the spectral formula in Example 5, utilizing a profile function φ, were studied in 1987 by Yaari [15] and Roell [14] in connection with a theory of "dual utility". (In those days, before VaR assumed such importance in finance, the expression being integrated along with φ was only regarded as a form of the inverse of the cumulative distribution function for X.) Acerbi [1] showed that the listed properties of φ are necessary for such a functional $\mathcal{R}$ to be a coherent risk measure, which he then termed a *spectral* risk measure. He identified spectral risk measures to some degree with CVaR mixtures, although not under the full generality here, where the mixture is given by a weighting measure. The risk profile for an "unmixed" risk measure CVaR_α itself is of course the function φ that has the value α^{-1} on $(0,\alpha]$ but 0 on $(\alpha,1)$. This corresponds to Acerbi's integral formula for $\mathrm{CVaR}_\alpha(X)$, as stated ahead of Example 5.

Mixed CVaR can be generalized to a weighting measure $d\lambda$ on $[0,1]$ instead of $(0,1)$. That amounts to admitting the limit cases $\mathrm{CVaR}_0(X) = -\inf X$ and $\mathrm{CVaR}_1(X) = -EX$ in some proportions. As long as the weight is not all placed on the endpoint 1, and weight is placed on 0 only in the finite-dimensional case of $\mathcal{L}^2$ (corresponding to a finite space Ω), one still gets a finite, coherent risk measure that is expectation-bounded.

To conclude this presentation of examples, we turn our sights on deviation measures that, like standard deviation and semideviations, can be derived merely by assigning penalties, possibly asymmetrically, to differences between outcomes $X(\omega)$ and their expected value EX.

Example 6 (deviation measures from simple penalties)**.** *Consider a function θ on $(-\infty,\infty)$ of type*

$$\theta(z) = a\max\{0,z\} + b\max\{0,-z\} = \begin{cases} az & \textit{when } z>0, \\ 0 & \textit{when } z=0, \\ b|z| & \textit{when } z<0, \end{cases}$$

for arbitrary coefficients $a \geq 0$ and $b \geq 0$, not both 0. Choose any $p \in [1,\infty)$ and let

$$\mathcal{D}(X) = (E[\theta(X-EX)]^p)^{1/p} = \left(\int_\Omega [\theta(X(\omega)-EX)]^p dP(\omega)\right)^{1/p}.$$

Then $\mathcal{D}$ is a deviation measure. Furthermore, $\mathcal{D}$ is coherent when $a=0$ and $b \leq 1$, or when merely $a+b \leq 1$ in the case of $p=1$, but not otherwise (for state spaces Ω with subsets of probability $\pi \in (0,1)$ for π arbitrarily small).

Detail. By its formula, $\mathcal{D}(X)$ depends only on $X - EX$, so $\mathcal{D}$ satisfies D1. To verify that $\mathcal{D}$ fits the remaining criteria for a deviation measure, it will be helpful to appeal to the representation $\mathcal{D}(X) = [\Theta(X-EX)]^{1/p}$ for $\Theta(X) = \int_\Omega[\theta(X(\omega))]^p dP(\omega)$. The "integrand" in this formula, namely the function θ^p on $I\!R$ (with $\theta^p(z) = [\theta(z)]^p$), is nonnegative and convex. It follows then that Θ is convex on $\mathcal{L}^2$, although perhaps having $\Theta(X) = \infty$ for some choices of X. Also, because $\theta^p(0) = 0$ and $\theta^p(\lambda z) = \lambda^p\theta^p(z)$ for $\lambda > 0$, we have $\Theta(0) = 0$ and $\Theta(\lambda X) = \lambda^p\Theta(X)$ for $\lambda > 0$, i.e., Θ is positively homogeneous of degree p.

It is understood in convex analysis that if a nonnegative functional Θ is convex and positively homogeneous of degree p, then $\Theta^{1/p}$ is convex and positively homogeneous of degree 1. Since $\mathcal{D}$ is obtained by composing $\Theta^{1/p}$ with the linear mapping $X \mapsto X - EX$, it too is convex and positively homogeneous of degree 1, thus satisfying D2 and D3.

We have $\theta(X(\omega)-EX) = a\max\{0, X(\omega)-EX\} + b\max\{0, EX - X(\omega)\} \geq 0$. Hence $\mathcal{D}(X) = 0$ when X is constant, but on the other hand $\mathcal{D}(X) > 0$ when X is nonconstant, because in that case each of the max expressions must be positive with positive probability, and therefore $\theta(X(\omega)-EX) > 0$ with positive probability by the conditions on a and b. Thus, D4 holds.

For $\mathcal{D}$ to be coherent, we would need D5 to be fulfilled: $\mathcal{D}(X) \leq EX - \inf X$ for all $X \in \mathcal{L}^2$. Because $\theta(X(\omega)-EX) \leq a[\sup X - EX] + b[EX - \inf X]$, we have $\mathcal{D}(X) \leq a[\sup X - EX] + b[EX - \inf X]$ in general, and are assured of coherency when $a = 0$ and $b \leq 1$. If $p = 1$, so that $\mathcal{D}(X) = aE[\max\{0, X - EX\}] + bE[\max\{0, EX - X\}]$, we can appeal the fact that $E[\max\{0, X - EX\}] = E[\max\{0, EX - X\}] \leq EX - \inf X$ to see that $\mathcal{D}(X) \leq EX - \inf X$ as long as $a + b \leq 1$.

To see how coherency fails otherwise, consider a subset Ω_0 of Ω having probability $\pi \in (0,1)$, and define $X \in \mathcal{L}^2$ by setting $X(\omega) = 1$ for $\omega \in \Omega_0$ and $X(\omega) = 0$ for $\omega \notin \Omega_0$. This yields $\sup X = 1$, $\inf X = 0$ and $EX = \pi$, and makes $\theta(X(\omega) - EX)$ take the value $a(1-\pi)$ with probability π and the value $b\pi$ with probability $1-\pi$. Then $\mathcal{D}(X) = [a^p(1-\pi)^p\pi + b^p\pi^p(1-\pi)]^{1/p}$ while $EX - \inf X = \pi$, so coherency would require that $a^p(1-\pi)^p\pi + b^p\pi^p(1-\pi) \leq \pi^p$, or on dividing both sides by π^p, that $a^p(\pi^{-1}-1)^p\pi + b^p(1-\pi) \leq 1$. By choosing values of π nearer and nearer to 0 (which is possible under the indicated assumption on Ω), we can produce a violation of this inequality in the case of $p = 1$ unless $a + b \leq 1$. In the case of $p > 1$, we can similarly produce a violation unless $a = 0$ and $b \leq 1$, due to the fact that $(\pi^{-1}-1)^p\pi$ tends to ∞ as π tends to 0 (as seen through the change of variables $s = \pi^{-1} - 1$, $\pi = (1+s)^{-1}$, by taking the limit as s tends to ∞). Thus, without the specified restrictions on a and b, coherency is impossible (in such state spaces Ω). □

The class of deviation measures in Example 6 includes standard deviation σ (for $a = b = 1$, $p = 2$) along with the semideviations σ_+ (for $a = 1$, $b = 0$, $p = 2$) and σ_- (for $a = 0$, $b = 1$, $p = 2$). When $a = b = 1$, one has $\theta(X(\omega) - EX) = |X(\omega) - EX|$, and $\mathcal{D}$ reduces to the $\mathcal{L}^p$ norm of $X - EX$. Of these particular possibilities, only σ_- is always coherent, according to the argument just given.

A final topic to be addressed in this section is the continuity, or semicontinuity, of deviation measures and risk measures. Recall that a functional $\mathcal{F} : \mathcal{L}^2 \to (-\infty, \infty]$ is called *lower semicontinuous* if its lower level sets $\{ X \,|\, \mathcal{F}(X) \leq c\}$ are all closed — in the norm topology of $\mathcal{L}^2$ (so that if a sequence of r.v.'s X_k satisfies $\mathcal{F}(X_k) \leq c$, and X_k converges to X in the sense that $\mu(X_k - X) \to 0$ and $\sigma(X_k - X) \to 0$, then $\mathcal{F}(X) \leq c$). Upper semicontinuity corresponds in the same way with upper level sets, given by the opposite inequality; the upper semicontinuity of $\mathcal{F}$ is the lower semicontinuity of $-\mathcal{F}$. Continuity is equivalent to the combination of lower semicontinuity with upper semicontinuity. For the fundamental role of semicontinuity in the theory of optimization, see [13].

Proposition 1 (continuity properties). *A deviation measure $\mathcal{D}$ is lower semicontinuous if and only if the corresponding expectation-bounded risk measure $\mathcal{R}$ is lower semicontinuous. These measures are continuous if and only if they are lower semicontinuous and finite everywhere on $\mathcal{L}^2$.*

Proof. Lower semicontinuity is preserved when adding two functionals that might take on ∞ (but not $-\infty$), so the initial claim is justified by the correspondence between $\mathcal{D}$ and the continuity of the linear functional $X \mapsto EX$. Because of the positive homogeneity in D2 and R2, if either functional has the value ∞ at some X it also has that value at λX for all $\lambda > 0$, in which event continuity at the origin is impossible. On the other hand, it is known that any lower semicontinuous *convex* functional on $\mathcal{L}^2$ is continuous everywhere when it is finite everywhere. (This is true on any Banach space; see Rockafellar [9, Corollary 8B].) ◻

Proposition 2 (cases of continuity and semicontinuity). *The risk and deviation measures based on variance in Example 1, and the ones based on* CVaR *in Examples 4 and 5, are continuous. All the deviation measures belonging to the class in Example 6 are lower semicontinuous, indeed continuous when finite everywhere.*

Proof. For the measures based on standard deviation, this is obvious from the definition of the topology on $\mathcal{L}^2$. For pure CVaR, we can use our minimization formula to see that, by taking $C = 0$, we have

$$\mathrm{CVaR}_\alpha(X) \leq \alpha^{-1} E \max\{-X, 0\} \leq \alpha^{-1} E|X|.$$

The convex functional $\mathcal{R} = \mathrm{CVaR}_\alpha$ is thus dominated by a finite, continuous functional and hence must itself be continuous. (For this well known property of convexity, see e.g. Rockafellar [9, Theorem 8].) Turning to the task of establishing lower semicontinuity for mixed CVaR, we consider a convergent sequence of r.v.'s, $X_k \to X$. From the fact that $\mathrm{CVaR}^\Delta_\alpha(X_k) \to \mathrm{CVaR}^\Delta_\alpha(X)$, through the continuity just verified, and the fact that $\mathrm{CVaR}^\Delta_\alpha(X)$ is nonnegative, we get (by Fatou's Lemma)

$$\begin{aligned} \liminf_{k\to\infty} \int_{(0,1)} \mathrm{CVaR}^\Delta_\alpha(X_k) d\lambda(\alpha) \;&\geq\; \int_{(0,1)} [\liminf_{k\to\infty} \mathrm{CVaR}^\Delta_\alpha(X_k)] d\lambda(\alpha) \\ &=\; \int_{(0,1)} \mathrm{CVaR}^\Delta_\alpha(X) d\lambda(\alpha), \end{aligned}$$

which is the meaning of lower semicontinuity. Actual continuity follows by Proposition 1 when ∞ is excluded as a value, as guaranteed by the extra condition in Example 5 that $\int_0^1 \alpha^{-1} d\lambda(\alpha) < \infty$.

For the deviation measures in Example 6, the key fact is the lower semicontinuity of the integral functional Θ introduced in the details. Its integrand θ^p is a finite convex function on $\mathbb{R}$, thus falling into the category of a "normal convex integrand" in the theory of integral functionals, and it is bounded below on $\mathbb{R}$ by 0; these properties guarantee the lower semicontinuity of Θ, cf. [10]. ◻

3 Acceptable Risks and Sureness

A fundamental feature of the approach to risk taken by Artzner et al. in [3] is its connection with a notion of acceptability of the risk associated with a random outcome $X(\omega)$ from X. Acceptability in their sense means essentially that "the loss aspect of X is insignificant and not worth worrying about" — from the vantage point of a particular investor or risk evaluator. This is not an absolute concept but one with a subjective component, dependent on attitudes toward risk and specifically on a choice of risk measure $\mathcal{R}$.

The set of all X that are deemed to be acceptable is some subset $\mathcal{A}$ of $\mathcal{L}^2$. In [3], axioms are presented for $\mathcal{A}$ and shown to correspond to axioms for a risk measure $\mathcal{R}$. The next theorem captures the result in that paper for our $\mathcal{L}^2$ setting, with a small addition to take care of expectation-boundedness. It will be the starting point for an interpretation which we hope will promote a richer understanding of what underlies the minimization of deviation or risk. Recall once more our convention in this paper that C denotes a constant r.v. (and also the value of that constant in $I\!R$).

Theorem 2 (risk measures versus acceptance sets)**.** *The relations*

(a) $\mathcal{A} = \{\, X \,|\, \mathcal{R}(X) \leq 0\}$,

(b) $\mathcal{R}(X) = \inf\{\, C \,|\, X + C \in \mathcal{A}\}$

(where the infimum is taken to be ∞ when the set of such constants C is empty), set up a one-to-one correspondence between expectation-bounded risk measures $\mathcal{R}$ that are lower semicontinuous and acceptance sets $\mathcal{A}$ that satisfy

(A1) *$\mathcal{A}$ is closed and contains the positive constants C,*

(A2) *$0 \in \mathcal{A}$, and $\lambda X \in \mathcal{A}$ whenever $X \in \mathcal{A}$ and $\lambda > 0$,*

(A3) *$X + X' \in \mathcal{A}$ for any $X \in \mathcal{A}$ and $X' \in \mathcal{A}$,*

(A4) *$EX > 0$ for every $X \not\equiv 0$ in $\mathcal{A}$.*

Moreover, $\mathcal{R}$ has the monotonicity property R5 if and only if

(A5) *$\mathcal{A}$ contains every $X \geq 0$.*

Thus, $\mathcal{R}$ is coherent, as well as lower semicontinuous and expectation-bounded, if and only if $\mathcal{A}$ satisfies A1, A2, A3, A4 and A5.

Proof. The correspondence between R1, R2, R3 and R5 on the one hand, and A1, A2, A3 and A5 on the other, along with the exclusion of negative constants C from $\mathcal{A}$, was demonstrated by Artzner et al. in their more limited context in [3] (with $\mathcal{R}$ finite and Ω a finite set), but their arguments readily carry over. (Here we rely on A4 to prevent $\mathcal{A}$ from containing negative constants C.) In view of the rescaling property in A2, it would be enough in A1, of course, to assume $\mathcal{A}$ contains the constant 1.

The only thing really new here is the identification of A4 as corresponding to R4. To establish that, it is convenient to work with two alternative conditions which will be shown to be equivalent to R4, namely

(R4′) $\mathcal{R}(X) > 0$ for every $X \not\equiv 0$ having $EX \leq 0$,

(R4″) $\mathcal{R}(X) > 0$ for every $X \not\equiv 0$ having $EX = 0$. Clearly A4 $\Leftrightarrow$ R4′. We have R4 $\Rightarrow$ R4′ because if $EX \leq 0$ then $E[-X] \geq 0$, hence $\mathcal{R}(X) \geq 0$ by R4 for such X, moreover with the case of $\mathcal{R}(X) = 0$ excluded unless X is constant with $EX = 0$, i.e., $X \equiv 0$. It is obvious that R4′ $\Rightarrow$ R4″, so we only need to confirm now that R4″ $\Rightarrow$ R4. If R4″ holds, we can apply it to $X' = X - EX$ in place of X to see that $\mathcal{R}(X - EX) > 0$ whenever $X - EX \not\equiv 0$ (i.e., whenever X is nonconstant). Since $\mathcal{R}(X - EX) = \mathcal{R}(X) + EX$ by R1 (and on the other hand $\mathcal{R}(X - EX) = 0$ by R2 whenever $X - EX \equiv 0$), this yields R4.

When $\mathcal{R}$ is lower semicontinuous, its lower level sets are closed, and therefore $\mathcal{A}$ must be closed when it is defined by (a). On the other hand, suppose that $\mathcal{A}$ is closed and $\mathcal{R}$ is defined by (b). For any

X, the set of $c \in I\!R$ such that $X + c\,1 \in \mathcal{A}$ is closed; moreover, unless it is empty, this set must be an interval without upper bound, since whenever $X+c\,1 \in \mathcal{A}$ and $c' > c$ we have $X+c'1 = X+c\,1+[c'-c]1$ and consequently $X+c'1 \in \mathcal{A}$ by A3 and A2, along with the fact in A1 that $1 \in \mathcal{A}$. Hence the epigraph E of $\mathcal{R}$, which by definition consists of all the pairs $(X, c) \in \mathcal{L}^2 \times I\!R$ such that $\mathcal{R}(X) \le c$, is the same as the set of all $(X, c) \in \mathcal{L}^2 \times I\!R$ such that $X + c\,1 \in \mathcal{A}$. That in turn can be construed as the inverse image $T^{-1}(\mathcal{A})$ of $\mathcal{A}$ under continuous linear transformation $T : (X, c) \mapsto X + c\,1$ from $\mathcal{L}^2 \times I\!R$ to $\mathcal{L}^2$. Since $\mathcal{A}$ is closed, that inverse image, which is the epigraph E, must be closed as well. In particular, then, for any fixed $c \in I\!R$ the set of X such that $(X, c) \in E$ is a closed subset of $\mathcal{L}^2$. But that is the lower level set $\{\, X \mid \mathcal{R}(X) \le c\}$, so we conclude that $\mathcal{R}$ is lower semicontinuous. □

On the basis of A2 and A3, the acceptance set $\mathcal{A}$ is a *convex cone*. Any such cone defines a partial ordering. We formalize that now with a terminology aimed at helping to create the right picture of the meaning of risk acceptance.

Definition 3 (acceptably sure inequalities)**.** *For X and X' in $\mathcal{L}^2$ and an acceptance set $\mathcal{A}$, the condition that $X - X' \in \mathcal{A}$ will be expressed in the form $X \ge_{\mathcal{A}} X'$ and referred to as meaning that $X(\omega) \ge X'(\omega)$ acceptably surely for $\omega \in \Omega$, in contrast to the ordinary inequality $X \ge X'$, meaning that $X(\omega) \ge X'(\omega)$ almost surely for $\omega \in \Omega$, i.e., with probability 1.*

In particular, the elements $X \in \mathcal{A}$ are then the r.v.'s such that X is nonnegative acceptably surely, or in other words, X is loss-free acceptably surely.

It is well known that when an ordering is defined in this manner for a convex cone $\mathcal{A}$, it has the welcome properties that if $X \ge_{\mathcal{A}} X'$ then $\lambda X \ge_{\mathcal{A}} \lambda X'$ for any $\lambda > 0$, and $X + X'' \ge_{\mathcal{A}} X' + X''$ for any X''. Having $\mathcal{A}$ closed yields the further property that whenever a sequence of r.v.'s X_k converges to some r.v. X in $\mathcal{L}^2$ (which, as noted earlier, corresponds to $\mu(X_k) \to \mu(X)$ and $\sigma(X_k - X) \to 0$), and if $X_k \ge_{\mathcal{A}} X_0$ for all $k = 1, 2, \ldots$, then $X \ge_{\mathcal{A}} X_0$. Obviously $\mathcal{A} = \{\, X \mid X \ge_{\mathcal{A}} 0\}$.

The meaning of X being *loss-free acceptably surely* in the sense of Definition 3 will be elucidated now by looking at the acceptance sets $\mathcal{A}$ that correspond to the examples of risk measures (and deviation measures) already encountered.

Example 7 (acceptability for the worst-loss risk measure)**.** *The acceptance set for the risk measure $\mathcal{R}(X) = -\inf X = \sup[-X]$ is*

$$\mathcal{A} = \mathcal{L}^2_+ = \{\, X \mid X \ge 0\}.$$

The inequality $X \ge_{\mathcal{A}} X'$ then has the same meaning as $X \ge X'$, so that X is loss-free acceptably surely if and only if it is loss-free almost surely.

Example 8 (acceptability for risk measures from variance)**.** *The acceptance set for a risk measure of the form $\mathcal{R}(X) = \rho\sigma(X) - \mu(X)$ with $\rho > 0$ is*

$$\mathcal{A} = \{\, X \mid \mu(X) \ge \rho\sigma(X)\}.$$

On this basis, X is loss-free acceptably surely if and only if there are no losses aside from outcomes $X(\omega)$ that are more than ρ standard deviation units below the mean. Here, however, $\mathcal{A}$ fails to satisfy axiom A5, with the consequence that having $X \ge X'$ may not guarantee that $X \ge_{\mathcal{A}} X'$.

In general, of course, having $X \ge_{\mathcal{A}} X$ entails having $X \ge X'$ when axiom A5 is satisfied, but not necessarily otherwise. Thus, without the coherency of the risk measure $\mathcal{R}$ associated with $\mathcal{A}$, it is possible to have X loss-free almost surely, and yet not loss-free acceptably surely. This underscores the significance of coherency as a critical property to maintain in working with risk measures and deviation measures, even if it excludes the measures based on variance. Fortunately, the CVaR-based measures do enjoy coherency and thus do not run into such trouble.

Example 9 (acceptability for VaR)**.** *For $\mathcal{R}(X) = \mathrm{VaR}_\alpha(X)$, the associated acceptance set is*

$$\mathcal{A} = \{\, X \,|\, P\{X < 0\} \le \alpha\},$$

so that X is loss-free acceptably surely if and only if the probability of incurring a loss is no more than α. But this risk measure does not satisfy all the axioms of Definition 2, and accordingly this set $\mathcal{A}$ lacks some of properties in Theorem 2, in particular A3, and is therefore not convex.

Detail. For this $\mathcal{R}$, the condition $\mathcal{R}(X) \le 0$ used to define $\mathcal{A}$ means (by the formula for $\mathrm{VaR}_\alpha(X)$ in Example 3) that $\inf\{\, z \,|\, P\{X \le z\} > \alpha\} \ge 0$. That is equivalent to having $P\{X \le z\} > \alpha$ imply $z \ge 0$, or in reverse form, having $z < 0$ imply $P\{X \le z\} \le \alpha$. That means $P\{X < 0\} \le \alpha$. ☐

Example 10 (acceptability for CVaR)**.** *For $\mathcal{R}(X) = \mathrm{CVaR}_\alpha(X)$, the associated acceptance set is*

$$\mathcal{A} = \{\, X \,|\, [\text{expectation of the lower } \alpha\text{-tail of } X] \ge 0\}.$$

In this case, to say that X is loss-free acceptably surely is to say that, for the lower α-tail of X, the gains are enough to counterbalance any losses.

Equipped with the concept of acceptability and these examples, let us return to the earlier definitions of risk measure and deviation measure. In the terminology of Definition 3, every risk measure $\mathcal{R}$ satisfying R1, R2, R3 and R4 can be seen to have the character that

$$\mathcal{R}(X) \;=\; \inf\{\, C \,|\, X + C \ge_{\mathcal{A}} 0\}$$

with respect to a uniquely determined acceptance set $\mathcal{A}$ satisfying the axioms A1, A2, A3 and A4. Artzner et al. [3] were motivated in their treatment of risk measures by the interpretation that, through such a formula, $\mathcal{R}(X)$ is the smallest "cash reserve" that would need to be added to every outcome $X(\omega)$ of the r.v. X, representing an investment subject to uncertainty, in order to guarantee that possible losses are covered up to some acceptable standard of sureness — specifically (in our terminology) so that $X + C$ is loss-free acceptably surely for the given $\mathcal{A}$. (This description oversimplifies their idea, which handled "cash" more subtly and focussed on present input instead of future output, but it captures the essence of their innovation while conforming to our slightly different axiomatic pattern.)

We can carry this further by using risk acceptability to introduce alternatives to the essential infimum $\inf X$ and essential supremum $\sup X$ of any r.v. X.

Definition 4 (effective extremes of a random variable)**.** *With respect to an acceptance set $\mathcal{A}$, we define the $\mathcal{A}$-effective infimum and the $\mathcal{A}$-effective supremum of X to be the values*

$$\mathcal{A}\text{-}\inf X \;=\begin{cases} \textit{highest } C \textit{ such that } C \le_{\mathcal{A}} X & \textit{if a constant } C \le_{\mathcal{A}} X \textit{ exists,} \\ -\infty & \textit{otherwise,} \end{cases}$$

$$\mathcal{A}\text{-}\sup X =\begin{cases} \textit{lowest } C \textit{ such that } C \ge_{\mathcal{A}} X & \textit{if a constant } C \ge_{\mathcal{A}} X \textit{ exists,} \\ \infty & \textit{otherwise.} \end{cases}$$

When $\mathcal{A} = \mathcal{L}^2_+$ as in Example 7, we obviously get $\mathcal{A}\text{-}\inf X = \inf X$ and $\mathcal{A}\text{-}\sup X = \sup X$, but strict inequalities are otherwise possible. The advantage of the notion of $\mathcal{A}$-*effective* infimum and supremum, in general, is that it gives us a nice way to think about the risk measure $\mathcal{R}$ and deviation measure $\mathcal{D}$ associated with the standards of risk acceptance dictated by $\mathcal{A}$. Namely, we have

$$\mathcal{R}(X) \;=\; -\mathcal{A}\text{-}\inf X \;=\; \mathcal{A}\text{-}\sup[-X],$$

as an alternative expression of relation (b) in Theorem 2 and the description above. On the other hand,

$$\mathcal{D}(X) \,=\, \mathcal{A}\text{-}\sup[EX - X] \,=\, EX - \mathcal{A}\text{-}\inf X.$$

Thus, $\mathcal{D}(X)$ is the least upper bound for $EX - X$ in the sense of an *acceptably sure inequality* instead of an almost sure inequality. The r.v. $EX - X$ is the *downside* of X, which may well be more crucial to an application than the *upside* r.v. $X - EX$, which would be measured instead by $\tilde{\mathcal{D}}(X) = \mathcal{A}\text{-}\sup[X - EX]$.

To emphasize, *every* deviation measure $\mathcal{D}$ that satisfies the axioms D1, D2, D3 and D4, and is lower semicontinuous can be characterized as having this type of formula with respect to a *uniquely determined* acceptance set $\mathcal{A}$ satisfying A1, A2, A3 and A4. We have $\mathcal{D}$ and the associated $\mathcal{R}$ coherent if and only if $\mathcal{A}$ also satisfies A5.

Still more insights can be gained by looking at this picture, not from the angle of $\mathcal{R}$ and $\mathcal{D}$, but that of the functional $\mathcal{S} = -\mathcal{R}$, which has the description

$$\mathcal{S}(X) \,=\, \mathcal{A}\text{-}\inf X.$$

It extracts from X the highest amount of return that can be counted on, up to the standard of sureness fixed by $\mathcal{A}$, despite the uncertainty in the outcomes $X(\omega)$ from X. We can appropriately refer to $\mathcal{S}$ as the *sureness valuation* associated with $\mathcal{A}$.

Theorem 2 could be restated equivalently in terms of sureness valuations under the correspondence $\mathcal{S} \leftrightarrow \mathcal{R}$ in which $\mathcal{S} = -\mathcal{R}$, or for that matter, $\mathcal{R} = -\mathcal{S}$. They are the functionals $\mathcal{S} : \mathcal{L}^2 \to [-\infty, \infty)$ satisfying

(S1) $\mathcal{S}(X + C) = \mathcal{S}(X) + C$ for all X and constants C,
(S2) $\mathcal{S}(0) = 0$, and $\mathcal{S}(\lambda X) = \lambda \mathcal{S}(X)$ for all X and all $\lambda > 0$,
(S3) $\mathcal{S}(X + X') \geq \mathcal{S}(X) + \mathcal{S}(X')$ for all X and X',
(S4) $\mathcal{S}(X) < EX$ for all nonconstant X, whereas $\mathcal{S}(X) = EX$ for constant X.

The closedness of $\mathcal{A}$ translates to the upper semicontinuity of $\mathcal{S}$, whereas coherency comes out as entailing the additional property that

(S5) $\mathcal{S}(X) \geq \mathcal{S}(X')$ when $X \geq X'$,

Observe that, through S2 and S3, any sureness valuation $\mathcal{S}$ is in particular a concave functional which might be regarded as a sort of *utility* function for a decision-maker employing it. Minimizing the risk $\mathcal{R}(X)$ of X, under some system of constraints on X, is equivalent to maximizing the corresponding "acceptably sure" part $\mathcal{S}(X)$ of X.

The comparison between $\mathcal{S}(X) = \mathcal{A}\text{-}\inf X$ and $\mathcal{D}(X) = \mathcal{A}\text{-}\sup[X - EX]$ is illuminating as well. When the underlying probability measure P is a so-called "pricing measure", EX can be regarded as the price of X. Then S4 has the interpretation that the part of X that one can count on getting (under the standards provided by $\mathcal{A}$) can never exceed its market price, and must be less when there is uncertainty in X. This gives us yet another way to think of the deviation $\mathcal{D}(X)$, namely as the *risk premium* associated with X relative to $\mathcal{A}$, inasmuch as relation (b) of Theorem 1, when expressed through $\mathcal{S}$ instead of $\mathcal{R}$, says that

$$\mathcal{S}(X) = EX - \mathcal{D}(X).$$

4 Risk Envelopes and Duality

Other useful perspectives on $\mathcal{D}$, $\mathcal{R}$ and $\mathcal{A}$ come from dual characterizations of these objects furnished by convex analysis.

We begin with an extension of a result of Artzner et al. in [3] about risk measures. In their special context of a finite state space Ω, they observed that any finite, coherent risk measure $\mathcal{R}$ has a unique representation of the form

$$\mathcal{R}(X) = \max_{p \in \mathcal{R}} \left\{ -\sum\nolimits_{\omega \in \Omega} X(\omega) p(\omega) \right\}$$

for a nonempty, closed, convex set $\mathcal{R}$ of probability measures p (assigning to each future state $\omega \in \Omega$ a probability $p(\omega) \geq 0$, with $\sum_{\omega \in \Omega} p(\omega) = 1$).

To a similar kind of representation in our context, where Ω is a possibly infinite space supplied with a reference probability measure P, and the random variables X that we work with belong to the associated $\mathcal{L}^2$ space, we must proceed a bit differently. We need to concentrate on probability measures on Ω that are described by density functions Q with respect to P, so that the corresponding expectation of X comes out as $\int_\Omega X(\omega) Q(\omega) dP(\omega)$. Furthermore, we have to insist on having $Q \in \mathcal{L}^2$, in order for this integral to be well defined for every $X \in \mathcal{L}^2$. Then, of course, we have

$$E_Q X = \int_\Omega X(\omega) Q(\omega) dP(\omega) = E[XQ], \text{ expectation under the probability measure } QP.$$

The conditions on an element Q of $\mathcal{L}^2$ that identify it as a probability density function are: $Q(\omega) \geq 0$ (almost surely), $\int_\Omega Q(\omega) dP(\omega) = 1$, or in other words, $Q \geq 0$, $EQ = 1$. Note that the density function corresponding to the underlying measure P itself is the constant r.v. 1. When Ω is finite and P is specified by assigning a probability $P(\omega)$ to each state ω, a density function Q merely assigns to each ω a factor $Q(\omega) \geq 0$ in such a way that $\sum_{\omega \in \Omega)} Q(\omega) P(\omega) = 1$. The probability of ω under the probability measure QP is thus $Q(\omega) P(\omega)$.

Along with this notational maneuver, our extended representation result has to account for the effect of expectation-boundedness. On the other hand, it has to go somewhat beyond coherency, because we wish to include risk measures coming from standard deviation, for instance. That requires admitting, in some situations, measures QP having $Q \not\geq 0$ (which are not probability measures).

Theorem 3 (risk envelope characterizations)**.** *The relations*

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

(b) $\mathcal{D}(X) = EX - \inf\{ E[XQ] \mid Q \in \mathcal{Q} \}$,

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

(d) $\mathcal{A} = \{ X \mid E[XQ] \geq 0 \text{ for all } Q \in \mathcal{Q} \}$,

set up a one-to-one correspondence between, on the one hand, lower semicontinuous deviation measures $\mathcal{D}$ paired with expectation-bounded risk measures $\mathcal{R}$ and acceptance sets $\mathcal{A}$, and on the other hand, sets $\mathcal{Q} \subset \mathcal{L}^2$, to be called risk envelopes, that satisfy the following axioms:

(Q1) *$\mathcal{Q}$ is a closed, convex set containing* 1 *(constant),*

(Q2) *every $Q \in \mathcal{Q}$ has $EQ = 1$,*

(Q3) *there is no nonconstant $X \in \mathcal{L}^2$ such that $E[XQ] \leq EX$ for all $Q \in \mathcal{Q}$.*

Within this correspondence, $\mathcal{R}$ is coherent (along with $\mathcal{D}$) if and only if $\mathcal{Q}$ is such that

(Q4) *$Q \geq 0$ for all $Q \in \mathcal{Q}$,*

in which case $\mathcal{Q}$ is a set of densities Q with respect to P and describes a class of probability measures of form QP on Ω. In that setting, X is loss-free acceptably surely if and only if none of these probability measures QP yields an expectation $E_Q X < 0$.

Proof. We can focus on the correspondence between (a) and (c), inasmuch as the extension to (b) and (d) is then immediate from the relations already developed. In convex analysis, the lower semicontinuous convex functionals on $\mathcal{L}^2$ that are positively homogeneous are known to be the support functions of nonempty closed convex sets in $\mathcal{L}^2$; in the present notation, this translates to having a one-to-one correspondence between such functionals $\mathcal{R}$ and such sets $\mathcal{Q}$ in the pattern of (a) and (c). We clearly have the expectation-boundedness inequality $\mathcal{R}(X) \geq -EX$ if and only if $1 \in \mathcal{Q}$. Strictness in this inequality for nonconstant X means that no nonconstant X can furnish the maximum in (c). Thus, R2, R3 and R4 correspond to Q1 and Q3.

On the other hand, R1 corresponds in the (a)-(c) relationship to Q2. To see this, observe first that if $\mathcal{Q}$ is obtained from $\mathcal{R}$ by (a) and Q is any element of $\mathcal{Q}$, then in particular we have for arbitrary $X \in \mathcal{L}^2$ that $-E[(X+C)Q] \leq \mathcal{R}(X+C)$ for all constants C. Through the R1 property of $\mathcal{R}$ this translates into having $-E[(X+C)Q] \leq \mathcal{R}(X) - C$ for all constants C, hence $C(1-EQ) \leq \mathcal{R}(X)+EX$ for all C. That is impossible unless $1 - EQ = 0$, so we must have Q2. Conversely, if Q2 holds and $\mathcal{R}$ is given by (a), we have for any constant C that $\mathcal{R}(X+C) = \sup\{ -E[(X+C)Q] \,|\, Q \in \mathcal{Q}\}$, where $E[(X+C)Q] = E[XQ] + C$. Therefore, $\mathcal{R}(X+C) = \mathcal{R}(X) - C$, and we conclude that R1 holds.

The confirmation that the nonnegativity in Q4 corresponds to the monotonicity in R5 proceeds similarly. If $\mathcal{R}$ is obtained from $\mathcal{Q}$ by (c), and Q4 holds, we have for $X' \leq X$ that $-E[X'Q] \geq -E[XQ]$ for all $Q \in \mathcal{Q}$, so that R5 holds. Conversely, under the assumption that $\mathcal{Q}$ is obtained from $\mathcal{R}$ by (a) with $\mathcal{R}$ satisfying R5, consider any $Q \in \mathcal{Q}$. Since R5 entails having $\mathcal{R}(X) \leq 0$ when $X \geq 0$, we have through (a) that $-E[XQ] \leq 0$ when $X \geq 0$. Applying this to $X = Q_-$ under the definition that $Q_-(\omega) = \min\{0, Q(\omega)\}$, we get $0 \leq E[XQ] = -E[Q_-]^2$, hence $Q_- = 0$ (almost surely). Thus, $Q \geq 0$, and we have Q4. □

Geometrically, $\mathcal{Q}$ can be viewed as a closed, convex set (by Q1) that lies within the hyperplane $H = \{ Q \in \mathcal{L}^2 \,|\, EQ = 1\}$ (by Q2). It contains (also by Q1) the "point" 1 which, as a constant r.v. interpreted as a density, stands for P. Condition Q3 has the meaning that no (closed) hyperplane other than H passes through this point 1 without having some points of $\mathcal{Q}$ on both sides of it. When the state space Ω is finite and $\mathcal{L}^2$ is therefore finite-dimensional, this property (the lack of a "nontrivial supporting hyperplane" at a point of $\mathcal{Q}$) can be identified through convex analysis [8] with the property that the point belongs to the interior of $\mathcal{Q}$ relative to H.

Note from characterizations (b) and (c) in Theorem 3 that $\mathcal{R}$ and $\mathcal{D}$ are finite on $\mathcal{L}^2$ if and only if, for each $X \in \mathcal{L}^2$, the linear functional $Q \mapsto E[XQ]$ is bounded over the set $\mathcal{Q}$ (i.e., has finite infimum and supremum). Such "linear boundedness" is known in functional analysis to be equivalent to $\mathcal{Q}$ being "norm bounded" in $\mathcal{L}^2$. Moreover, because every $Q \in \mathcal{Q}$ has $\mu(Q) = 1$ (by Q2), that comes down to there being an upper bound to the variances $\sigma^2(Q)$ as Q ranges over $\mathcal{Q}$.

Although Theorem 3 covers noncoherent risk measures and deviation measures by allowing Q to take on negative values, the coherent case where $Q \geq 0$ (axiom Q4) provides the clearest motivation. We therefore discuss it now at some length.

Artzner et al., in their simpler framework [3], did not give a name to the convex sets of probability measures in their representation result for a coherent risk measure $\mathcal{R}$. Here, with that set taking the form $\{ QP \,|\, Q \in \mathcal{Q}\}$, we have dubbed $\mathcal{Q}$ the *risk envelope* associated with $\mathcal{R}$ in order to underscore the view that the elements $Q \in \mathcal{Q}$ describe alternatives to, or perturbations of, the reference probability measure P. The elements $Q \in \mathcal{Q}$ can themselves be called the *risk monitors* associated with $\mathcal{R}$. The "envelope" part of the name arises from the manner in which $\mathcal{Q}$ "envelopes" the constant density 1 (in a relative neighborhood, as explained above on the basis of axiom Q3, which corresponds to expectation boundedness).

We can think, for example, that P may have been derived from the current financial markets through no-arbitrage considerations as the probability measure giving the market consensus on the likelihood of the future states $\omega \in \Omega$. Then EX would be the *market expectation* of X. No one really knows a "true" distribution for the future states, however. To guard against overconfidence, a set of alternative probability measures "enveloping" P may be designated for purposes of risk assessment. That would correspond, in our terminology, to selecting a particular risk envelope $\mathcal{Q}$. It would dictate, through the relations in Theorem 3, not only the risk measure $\mathcal{R}$ but also the deviation measure $\mathcal{D}$ and the acceptance set $\mathcal{A}$.

The same goes for the corresponding sureness valuation $\mathcal{S}$, of course, and there the interpretation is immediate. Theorem 3 gives the characterization that

$$\mathcal{S}(X) = \inf_{Q \in \mathcal{Q}} E_Q X,$$

which tells us that the value $\mathcal{S}(X)$ comes from a type of *worst-case analysis* of expectations relative to the alternative views of the future portrayed by the various risk monitors $Q \in \mathcal{Q}$.

Another way to think about this "coherent" situation (enjoying axiom Q4) is that the states $\omega \in \Omega$ represent "scenarios", whereas probability distributions on those states represent "mixed scenarios" (much like mixed strategies in classical game theory). Perhaps, in the near future toward which we are headed, we will not yet know the exact state ω that eventually will be reached, but will have additional information allowing us to refine the current probability measure P. Each Q stands then for such a future state of (still partial) information.

For deviations, the "coherent" case allows us to pose the formula for $\mathcal{D}(X)$ in Theorem 3(b) as

$$\mathcal{D}(X) = \sup_{Q \in \mathcal{Q}} E_Q[EX - X],$$

where the r.v. $EX - X$ is the *shortfall* of X relative to its expectation. In other words, $\mathcal{D}(X)$ stands for the worst shortfall that may be incurred in the environment of the probability densities Q that comprise the the risk envelope $\mathcal{Q}$.

Example 11 (probability basis for worst-case risk)**.** *The coherent and expectation-bounded risk measure* $\mathcal{R}(X) = -\inf X = \sup[-X]$ *corresponds to the risk envelope consisting of all possible probability densities:*

$$\mathcal{Q} = \left\{ Q \in \mathcal{L}^2 \;\middle|\; Q(\omega) \geq 0,\ \int_\Omega Q(\omega) dP(\omega) = 1 \right\}.$$

Detail. Because $\mathcal{R}$ is coherent, the elements Q of $\mathcal{Q}$ must all have $Q \geq 0$ and $EQ = 1$. But any such Q has the property that $E[XQ] \geq \inf X$ for every $X \in \mathcal{L}^2$ and therefore belongs to $\mathcal{Q}$ by characterization (a) of Theorem 3. □

In the framework of $\mathcal{Q}$ offering alternative pictures of the future, the worst-case risk measure thus corresponds to admitting "all" alternatives. At the opposite extreme, we might contemplate just having $\mathcal{Q} = \{1\}$, which would mean complete reliance on the underlying probability measure P. That choice of $\mathcal{Q}$ would not fit with Q3 (the corresponding $\mathcal{R}$ is not expectation-bounded in the strict sense of R4), but the same idea could be mimicked by taking $\mathcal{Q}$ to be a tiny set around 1 that does satisfy Q3. This would amount to very strong, if not complete, faith in P.

Example 12 (probability basis for CVaR)**.** *The coherent and expectation-bounded risk measure* $\mathcal{R}(X) = \mathrm{CVaR}_\alpha(X)$ *for* $\alpha \in (0,1)$ *corresponds to the risk envelope*

$$\mathcal{Q} = \left\{ Q \in \mathcal{L}^2 \;\middle|\; 0 \leq Q(\omega) \leq \alpha^{-1},\ \int_\Omega Q(\omega) dP(\omega) = 1 \right\}.$$

Detail. The indicated set $\mathcal{Q}$ evidently satisfies Q1, Q2, Q3 and Q4, so by Theorem 3 it will be enough to show that it yields $\mathcal{R}(X) = \mathrm{CVaR}_\alpha(X)$ through the maximization in (c) of Theorem 3. In that maximization, a particular Q will fall short unless it concentrates on low values of $X(\omega)$ as much as possible. Fix any X and α, and partition Ω into the three (measurable) sets Ω^-, Ω^0 and Ω^+, according to whether $X(\omega) < -\mathrm{VaR}_\alpha(X)$, $X(\omega) = -\mathrm{VaR}_\alpha(X)$, or $X(\omega) > -\mathrm{VaR}_\alpha(X)$. By the definition of $\mathrm{VaR}_\alpha(X)$, we have $P(\Omega^-) \leq \alpha$ and $P(\Omega^+) \leq 1-\alpha$; in particular, the gap $\alpha - P(\Omega^-)$ cannot exceed $P(\Omega^0)$. For Q to achieve the maximum, we must certainly have $Q(\omega) = \alpha^{-1}$ for $\omega \in \Omega^-$ but $Q(\omega) = 0$ for $\omega \in \Omega^+$. On Ω^0, however, any (measurably chosen) values $Q(\omega) \in [0, \alpha^{-1}]$ are possible, consistent with having the integral of Q over Ω^0 come out as making up the gap in the inequality $P\{\omega \in \Omega^-\} \leq \alpha$, if any. For any such choice of Q, one gets $\mathrm{CVaR}_\alpha(X) = -E[XQ]$ by the definition of $\mathrm{CVaR}_\alpha(X)$ in Example 4. In particular, when there is a gap, one could take $Q(\omega) = \alpha^{-1}[\alpha - P(\Omega^-)]/P(\Omega^0)$ on Ω^0 to get the maximum. □

Delbaen, in unpublished lecture notes [6], demonstrated that the choice of $\mathcal{Q}$ in Example 12 yields $\mathcal{R}(X) = \mathrm{CVaR}_\alpha(X)$ when the distribution of X is atomless. He further observed that, if atoms were present, it would yield a more subtle expression which can now be recognized as $\mathrm{CVaR}_\alpha(X)$ defined in general in terms of the lower α-tail we introduced in [12]. In this sense, the facts in Example 12 can be credited to Delbaen, although at the time of his work there was only the "tail-VaR" predecessor to full CVaR, in terms of an atomless conditional expectation.

The risk envelope formula in Example 12 sheds additional light on the limit relations in Example 4 as $\alpha \to 0$ or $\alpha \to 1$. As $\alpha \to 0$, the set $\mathcal{Q}$ in Example 12 turns into the one in Example 11. On the other hand, if we take $\alpha = 1$, we get $\mathcal{Q} = \{1\}$.

Example 13 (worst conditional expectation risk measure)**.** *The worst conditional expectation risk measure for any $\alpha \in (0,1)$ is defined in terms of the conditional expectation $E[X|\mathcal{E}]$ of $X(\omega)$ subject to $\omega \in \mathcal{E}$ by*

$$\mathcal{R}(X) = \mathrm{WCE}_\alpha(X) = \sup\{\, -E[X|\mathcal{E}] \mid \mathcal{E} \subset \Omega,\ P(\mathcal{E}) > \alpha\}.$$

It is a coherent and expectation-bounded risk measure which corresponds to the set $\mathcal{Q}$ that is the closed convex hull of all functions of the form

$$Q_{\mathcal{E}}(\omega) = \begin{cases} 1/P(\mathcal{E}) & \text{if } \omega \in \mathcal{E}, \\ 0 & \text{if } \omega \notin \mathcal{E}, \end{cases}$$

for the various choices of a (measurable) set $\mathcal{E} \subset \Omega$ having $P(\mathcal{E}) > \alpha$.

Detail. The set $\mathcal{Q}$ so described does satisfy Q1, Q2, Q3 and Q4, and it therefore corresponds in the (a)(c) relationship of Theorem 3 to a coherent, expectation-bounded risk measure $\mathcal{R}$. Taking the supremum over the closed convex hull is the same as taking the supremum over the elements $Q_{\mathcal{E}}$ from which the hull is generated. □

The WCE_α risk measure in Example 13 was offered in [3] as a prime example of coherency in the days before CVaR came to the fore. Its close relationship to the CVaR_α risk measure is apparent from the characterization in Example 12. Each $Q_{\mathcal{E}}$ is a particular element of the set $\mathcal{Q}$ in Example 13, so that necessarily $\mathrm{WCE}_\alpha(X) \leq \mathrm{CVaR}_\alpha(X)$. The difference can be strict, however, for the reasons that emerged in justifying Example 12. It was seen there that when the set of ω for which $X(\omega) = -\mathrm{VaR}_\alpha(X)$ has positive probability, i.e., the distribution of X has an atom at $-\mathrm{VaR}_\alpha(X)$, it may be impossible to reach $\mathrm{CVaR}_\alpha(X)$ in the maximization without resorting to a Q that has a different value on that set from the one it has on the set of ω for which $X(\omega) < -\mathrm{VaR}_\alpha(X)$. Such a Q cannot lie in the convex hull in Example 13. In short, although $\mathrm{CVaR}_\alpha(X)$ and $\mathrm{WCE}_\alpha(X)$ are sure to coincide in an atomless environment, they can differ in general through the ability of $\mathrm{CVaR}_\alpha(X)$ to split atoms when $\mathrm{WCE}_\alpha(X)$ cannot.

Example 14 (risk envelopes from standard deviation)**.** *For a risk measure having the form* $\mathcal{R}(X) = \rho\sigma(X) - \mu(X)$ *with* $\rho > 0$*, the associated risk envelope is*

$$\mathcal{Q} = \{\, Q \in \mathcal{L}^2 \,|\, EQ = 1,\ \sigma(Q) \le \rho \}.$$

Detail. From Theorem 3(a), we have $Q \in \mathcal{Q}$ if and only if $-E[XQ] \le \rho\sigma(X) - \mu(X)$ for all X, which is the same as $E[X(1-Q)] \le \rho\sigma(X)$ for all X. That is equivalent to the cited conditions. □

Observe that because the risk measure in Example 14 is not coherent the elements of $\mathcal{Q}$ can fail to satisfy $Q \ge 0$ and thus do not stand for probability densities.

To tie these notions together and demonstrate that the characterizations in Theorem 3 have practical as well as theoretical significance, we elaborate on the application of CVaR risk envelope in Example 12 to the case of a discrete r.v.

Example 15 (discrete CVaR)**.** *Suppose that* X *takes on only finitely many values* $z_1, \ldots, z_n$*, doing so with probabilities* $p_1, \ldots, p_n$*. Then, for any* $\alpha \in (0,1)$*,*

$$\mathrm{CVaR}_\alpha(X) = -\min_{q_k}\Big\{ \textstyle\sum_{k=1}^n q_k p_k z_k \,\Big|\, 0 \le q_k \le \alpha^{-1},\ \sum_{k=1}^n q_k p_k = 1 \Big\}.$$

This formula also covers the limit cases of $\alpha = 1$ *and* $\alpha = 0$ *(in evident interpretation), and it can readily be extended to mixed* CVaR *as described in Example 5:*

$$\textstyle\sum_{i=1}^m \lambda_i \mathrm{CVaR}_{\alpha_i}(X) = -\min\limits_{q_{ik}}\Big\{ \sum_{i=1,\,k=1}^{m,n} \lambda_i q_{ik} p_k z_k \,\Big|\, 0 \le q_{ik} \le \alpha_i^{-1},\ \sum_{k=1}^n q_{ik} p_k = 1 \Big\}.$$

Detail. We can view X as belonging to the n-dimensional $\mathcal{L}^2$ space associated with a discrete state space $\Omega = \{\,\omega_1, \ldots, \omega_n\}$ where each ω_k has probability p_k. From Example 12, the risk envelope $\mathcal{Q}$ for the risk measure $\mathcal{R}(X) = \mathrm{CVaR}_\alpha(X)$ consists then of all Q having values q_k at the points ω_k that satisfy $0 \le q_k \le \alpha^{-1}$ and $\sum_{k=1}^n q_k p_k = 1$. The formula for $\mathrm{CVaR}_\alpha(X)$ emerges then from Theorem 3(c). The final formula is obtained by inserting such an expression for each $\mathrm{CVaR}_{\alpha_i}(X)$. □

5 Minimization of Risk or Deviation

In a wide range of financial applications, decisions revolve around trying to minimize "risk" within the available circumstances. We look now at what this entails in the framework developed above. Besides elucidating the behavior of general risk and deviation measures in problems of optimization, we develop optimality conditions through an appeal to basic tools of convex analysis. We demonstrate that risk envelopes have a crucial role in such conditions, and that aspects of risk acceptability can require attention as well.

An abstract approach is best overall, because there are different categories of constraints under which minimization might be desirable, in general. On that level, let us simply contemplate a subset $\mathcal{X}$ of $\mathcal{L}$, consisting of the r.v.'s X deemed to be *feasible* possibilities in an attempt at optimization, and ask only that $\mathcal{X}$ be *nonempty, closed and convex.* We are interested then in the problem

$$\mathcal{P}_{\mathcal{D}} \qquad \text{minimize } \mathcal{D}(X) \text{ over all } X \in \mathcal{X},$$

for a deviation measure $\mathcal{D}$, or alternatively in the problem

$$\mathcal{P}_{\mathcal{R}} \qquad \text{minimize } \mathcal{R}(X) \text{ over all } X \in \mathcal{X}$$

for the corresponding risk measure $\mathcal{R}$. Although $\mathcal{P}_\mathcal{D}$ fits a familiar pattern in finance, the motivation for $\mathcal{P}_\mathcal{R}$ may at first seem less clear. But $\mathcal{P}_\mathcal{R}$ can be equally be viewed in terms of the sureness valuation $\mathcal{S} = -\mathcal{R}$ as the problem

$$\mathcal{P}_\mathcal{S} \qquad\qquad \text{maximize } \mathcal{S}(X) \text{ over all } X \in \mathcal{X}.$$

That version makes immediate sense: it aims at choosing X from the feasible set $\mathcal{X}$ in such a way as to maximize the part of the return that we can count on getting *risk-free acceptably surely*, with all the ramifications of that notion that have been presented in various cases.

For the sake of concreteness, the reader may wish to keep in mind the following type of set $\mathcal{X}$ over which these problems may be of interest, even though our results will be applicable more broadly.

Example 16 (linear constraints on linear combinations). *Let $X_1, \ldots, X_n$ be a family of r.v.'s in $\mathcal{L}^2$, and let $\mathcal{X}$ consist of all X that can be expressed as a linear combination $x_1X_1 + \cdots + x_nX_n$ for a coefficient vector $(x_1, \ldots, x_n) \in I\!R^n$ satisfying a given constraint system of the form*

$$\sum_{j=1}^{n} a_{ij}x_j \begin{cases} \leq b_i & \text{for } i = 1, \ldots, s, \\ = b_i & \text{for } i = s+1, \ldots, m. \end{cases}$$

Then $\mathcal{X}$ is convex and closed, and of course is nonempty as long as some $(x_1, \ldots, x_n)$ exists that satisfies these equations and inequalities.

The r.v.'s X_j in Example 16 could, for instance, give the returns from a specified collection of financial instruments, and then X would be the return from a portfolio constructed from those instruments with weights x_j. The constraints could involve prices and expectations, or conditions like $x_j \geq 0$ (for instruments that cannot be shorted). They could also enforce diversification requirements. In the $\mathcal{P}_\mathcal{S}$ version of $\mathcal{P}_\mathcal{R}$ over such a set with $\mathcal{R}$ coherent, for example, we would be trying in the face of such restrictions to

$$\text{maximize the } \mathcal{Q}\text{-worst-case return } \inf_{Q \in \mathcal{Q}} E_Q\left[\sum_{j=1}^{m} x_jX_j\right] \text{ over the portfolio weights } x_j.$$

What is the relationship between $\mathcal{P}_\mathcal{D}$ and $\mathcal{P}_\mathcal{R}$? In finance, due to classical portfolio theory with its emphasis on minimizing standard deviation, it has become second nature to think of risk and deviation as virtually the same thing, and this has sometimes been reflected in ambiguities over $\mathrm{VaR}_\alpha(X)$ versus $\mathrm{VaR}_\alpha^\Delta(X)$, or $\mathrm{CVaR}_\alpha(X)$ versus $\mathrm{CVaR}_\alpha^\Delta(X)$, as mentioned earlier. Yet the axioms for risk measures and deviation measures are definitely different, and a corresponding difference between problems $\mathcal{P}_\mathcal{D}$ and $\mathcal{P}_\mathcal{R}$ cannot be ruled out. The precise state of affairs will be brought to light in Theorem 4 below.

In any optimization problem, it is essential to distinguish between the *optimal value* (the infimum of the expression being minimized) and the *optimal solutions* (the elements, if any, at which the minimization is achieved). For problem $\mathcal{P}_\mathcal{D}$, we will utilize the notation

$$\inf \mathcal{P}_\mathcal{D} = \text{ optimal value}, \qquad \operatorname{argmin} \mathcal{P}_\mathcal{D} = \text{ set of optimal solutions},$$

and similarly for the other problems.

To state our theorem with satisfying generality, we also need to make use of the concept of the *recession cone* $\operatorname{rc}\mathcal{X}$ of $\mathcal{X}$, which by its definition in convex analysis is given by

$$\operatorname{rc}\mathcal{X} = \{\, X' \in \mathcal{L}^2 \mid \text{whenever } X \in \mathcal{X}, \text{ one also has } X + X' \in \mathcal{X}\}.$$

Under our assumption that $\mathcal{X}$ is nonempty, closed and convex, $\operatorname{rc}\mathcal{X}$ is a closed convex cone which is characterized by the fact that, for any choice of $\bar{X} \in \mathcal{X}$, this cone consists of 0 and the nonzero elements $X' \in \mathcal{L}^2$ such that the "half-line" $\{\, \bar{X} + \lambda X' \mid \lambda \geq 0\}$ lies in $\mathcal{X}$ (see [8]).

Example 17 (recession for linear constraints)**.** *Suppose $\mathcal{X}$ has the structure in Example 16. The recession cone* $\mathrm{rc}\,\mathcal{X}$ *consists then of the linear combinations $X' = x'_1X_1 + \cdots + x'_nX_n$ obtained from coefficient vectors $(x'_1, \ldots, x'_n) \in I\!R^n$ that satisfy the corresponding homogeneous system of constraints,*

$$\sum\nolimits_{j=1}^{n} a_{ij}x'_j \begin{cases} \leq 0 & \text{for } i = 1, \ldots, s, \\ = 0 & \text{for } i = s+1, \ldots, m. \end{cases}$$

Detail. Consider any particular $\bar{X} = \bar{x}_1X_1 + \cdots + \bar{x}_mX_m$ in $\mathcal{X}$ and any $X' = x'_1X_1 + \cdots + x'_mX_m$. We will have $\bar{X} + \lambda X' \in \mathcal{X}$ for all $\lambda > 0$ if and only if the relations

$$\sum\nolimits_{j=1}^{n} a_{ij}(\bar{x}_j + \lambda x'_j) \begin{cases} \leq b_i & \text{for } i = 1, \ldots, s, \\ = b_i & \text{for } i = s+1, \ldots, m, \end{cases}$$

are satisfied for all $\lambda > 0$. That can readily be seen to hold if and only if the coefficients x'_j satisfy the homogeneous system in question. □

The key issue for understanding the relationship between $\mathcal{P}_{\mathcal{D}}$ and $\mathcal{P}_{\mathcal{R}}$ is the extent to which the description of $\mathcal{X}$ imposes a condition on the expectation EX of the r.v.'s $X \in \mathcal{X}$. Financial models commonly require, for instance, that $EX = \xi$ for some specified value ξ. People working in the broader optimization community know from experience, however, that in setting up a model it is good to be cautious about posing a constraint in equality form when it might just as well be posed in inequality form. Would there be anything wrong with the weaker requirement that $EX \geq \xi$? That question cannot be answered outside the context of a particular application, but it deserves attention because this is exactly where a difference between $\mathcal{P}_{\mathcal{D}}$ and $\mathcal{P}_{\mathcal{R}}$ can arise.

Theorem 4 (minimum risk versus minimum deviation)**.**

(a) *If the elements $X \in \mathcal{X}$ have to satisfy a constraint of type $EX = \xi$ for some given ξ, problems $\mathcal{P}_{\mathcal{D}}$ and $\mathcal{P}_{\mathcal{R}}$ are equivalent in the sense that*

$$\operatorname{argmin} \mathcal{P}_{\mathcal{D}} = \operatorname{argmin} \mathcal{P}_{\mathcal{R}}, \qquad \inf \mathcal{P}_{\mathcal{D}} = \inf \mathcal{P}_{\mathcal{R}} + \xi.$$

(b) *If the elements $X \in \mathcal{X}$ merely have to satisfy a constraint of type $EX \geq \xi$ for some given ξ, problems $\mathcal{P}_{\mathcal{D}}$ and $\mathcal{P}_{\mathcal{R}}$ are equivalent in the same sense as long as it can be established that their optimal solutions must actually satisfy this inequality constraint with equality (i.e., this constraint must be active in optimality). Otherwise, these problems can differ. In particular, if there exists*

$$X' \in \mathrm{rc}\,\mathcal{X} \text{ having } \mathcal{R}(X') < 0,$$

then $\inf \mathcal{P}_{\mathcal{R}} = -\infty$ *with* $\operatorname{argmin} \mathcal{P}_{\mathcal{R}} = \emptyset$, *even though* $\inf \mathcal{P}_{\mathcal{D}} \geq 0$, *possibly with* $\operatorname{argmin} \mathcal{P}_{\mathcal{D}} \neq \emptyset$.

Proof. Most of the claims are evident from the basic relation $\mathcal{R}(X) = \mathcal{D}(X) - EX$ in Theorem 1. Consider, however, a $X' \in \mathrm{rc}\,\mathcal{X}$ having $\mathcal{R}(X') < 0$. For any $X \in \mathcal{X}$ and $\lambda > 0$ we have $\mathcal{R}(X + \lambda X') \leq \mathcal{R}(X) + \lambda\mathcal{R}(X')$ by R2 and R3, where the right side tends to $-\infty$ as λ tends to ∞. Then certainly $\inf \mathcal{P}_{\mathcal{R}} = -\infty$ and $\operatorname{argmin} \mathcal{P}_{\mathcal{R}} = \emptyset$. □

The practical meaning of the discrepancy at the end of Theorem 4 can perhaps best be understood when interpreted through the equivalent problem $\mathcal{P}_{\mathcal{S}}$, where one is trying to maximize the "acceptably risk-free" return from among the available r.v.'s in $\mathcal{X}$. An element $X' \in \mathrm{rc}\,\mathcal{X}$ stands for some r.v. that can be added to any available X, in any positive multiple, without danger of causing a constraint violation (i.e., forcing an exit from $\mathcal{X}$). The condition $\mathcal{R}(X') < 0$ becomes $\mathcal{S}(X') > 0$: the r.v. X' yields a positive return, risk-free acceptably surely. It is an "acceptably free lunch", causing the supremum in $\mathcal{P}_{\mathcal{S}}$ to be ∞.

The existence of such a "free lunch" might not be detected if there were too narrow a focus on $\mathcal{P}_{\mathcal{D}}$ and an insensitivity to whether the expectation constraint should be written as an equation or as an inequality. In standard finance, such as arbitrage theory, "free lunches" are admitted only with respect to the sureness standard corresponding to the acceptance set $\mathcal{A} = \mathcal{L}^2_+$. We see here, however, the potential need to consider them under less stringent standards, namely "$\mathcal{A}$-free lunches".

We now take up the subject of optimality conditions in $\mathcal{P}_{\mathcal{D}}$ and $\mathcal{P}_{\mathcal{R}}$. For a foundation encompassing a variety of applications, we rely on two basic notions in convex analysis, as specialized to our space $\mathcal{L}^2$. First, for any convex functional $\mathcal{F} : \mathcal{L}^2 \to (-\infty, \infty)$ (we have in mind $\mathcal{D}$ and $\mathcal{R}$), the set of *subgradients* Y of $\mathcal{F}$ at a point X^* is

$$\partial\mathcal{F}(X^*) = \{\, Y \in \mathcal{L}^2 \,|\, \mathcal{F}(X) \geq \mathcal{F}(X^*) + E[(X - X^*)Y] \text{ for all } X \in \mathcal{L}^2 \}.$$

This is always a closed, convex set (empty when $\mathcal{F}(X^*) = \infty$), and it reduces to the singleton consisting of the gradient $\nabla\mathcal{F}(X^*)$ when $\mathcal{F}$ is differentiable at X^* (in the standard Fréchet sense). The second notion is that of the set of *normals* Y to a convex set $\mathcal{X}$ at a point $X^* \in \mathcal{X}$, namely

$$N_{\mathcal{X}}(X^*) = \{\, Y \in \mathcal{L}^2 \,|\, E[(X - X^*)Y] \leq 0 \text{ for all } X \in \mathcal{X} \}.$$

This is always a closed, convex cone (containing 0). Examples of subgradients of $\mathcal{F} = \mathcal{D}$ or $\mathcal{F} = \mathcal{R}$ will be developed out of Theorem 5 below, and therefore need to be postponed temporarily, but a valuable example of normals can be furnished right away.

Proposition 3 (normals for linear constraints)**.** *Suppose $\mathcal{X}$ has the structure in Example 16. Then, for any element X^* of $\mathcal{X}$, corresponding to coefficients x_j^*, one has the following characterization of normals: $Y \in N_{\mathcal{X}}(X^*)$ if and only if there exist multipliers $\lambda_1, \ldots, \lambda_m$ such that*

$$\begin{aligned} &E[X_jY] = \textstyle\sum_{i=1}^m \lambda_i a_{ij} \ \text{ for } j = 1, \ldots, n, \text{ with} \\ &\lambda_i \begin{cases} \geq 0 & \text{for } i \in \{1, \ldots, s\} \text{ with } \sum_{j=1}^n a_{ij} x_j^* = b_i, \\ = 0 & \text{for } i \in \{1, \ldots, s\} \text{ with } \sum_{j=1}^n a_{ij} x_j^* < b_i. \end{cases} \end{aligned}$$

Proof. The set $\mathcal{X}$ lies, by its definition, in the finite-dimensional subspace of $\mathcal{L}^2$ that is generated by $X_1, \ldots, X_n$. We are able for that reason to rely on the theory of linear constraints in finite dimensions. To say that $Y \in N_{\mathcal{X}}(X^*)$ is to say that $E[XY] \leq E[X^*Y]$ for all $X \in \mathcal{X}$. This condition can be reduced to an assertion of linear programming optimality. Having $Y \in N_{\mathcal{X}}(X^*)$ amounts to having

$$x_1 E[X_1Y] + \cdots + x_n E[X_nY] \ \leq \ x_1^* E[X_1Y] + \cdots + x_n^* E[X_nY]$$

for all choices of x_j satisfying the given linear constraints. That corresponds by linear programming duality to the existence of multipliers λ_i as described. □

Subgradients and normals share in characterizing the minimum of a functional $\mathcal{F}$ over $\mathcal{X}$. In the case where $\mathcal{F}$ is $\mathcal{D}$ or $\mathcal{R}$, the identification of the subgradients turns out to be closely tied to the associated risk envelope $\mathcal{Q}$ and thus, for top choices of $\mathcal{D}$ and $\mathcal{R}$, to the examples of $\mathcal{Q}$ that have already been seen. Our main result on optimality, which follows, brings this out this along with the extent to which the conditions involving subgradients and normals are necessary or sufficient.

Theorem 5 (optimality conditions for risk or deviation)**.** *Consider a deviation measure $\mathcal{D}$, the corresponding expectation-bounded risk measure $\mathcal{R}$, assumed to be lower semicontinuous, and the associated risk envelope $\mathcal{Q}$. For each $X \in \mathcal{L}^2$, let*

$$\mathcal{Q}_X \ = \ \operatorname*{argmin}_{Q \in \mathcal{Q}} E[XQ] \ = \ \Big\{\, Q \in \mathcal{Q} \,\Big|\, E[XQ'] \geq E[XQ] \text{ for all } Q' \in \mathcal{Q} \Big\}.$$

The subgradient sets for $\mathcal{D}$ and $\mathcal{R}$ then take the form

$$\begin{aligned}\partial\mathcal{R}(X) &= \{\, -Q \,|\, Q \in \mathcal{Q}_X \} = \{\, -Q \,|\, Q \in \mathcal{Q},\ \mathcal{R}(X) = -E[XQ] \,\},\\ \partial\mathcal{D}(X) &= \{\, Y \,|\, 1 - Y \in \mathcal{Q}_X \} = \{\, Y \,|\, 1 - Y \in \mathcal{Q},\ \mathcal{D}(X) = E[XY] \,\},\end{aligned}$$

satisfying in particular $Y \in \mathcal{D}(X)$ if and only if $Y - 1 \in \mathcal{R}(X)$. They have the following role in determining optimality in $\mathcal{P}_{\mathcal{R}}$ and $\mathcal{P}_{\mathcal{D}}$ (as long as $\mathcal{X}$ is nonempty, closed, and convex).

(a) *A sufficient condition for X^* to be an optimal solution to $\mathcal{P}_{\mathcal{R}}$, which moreover is necessary when $\mathcal{R}$ is finite everywhere on $\mathcal{L}^2$, is the existence of*

$$-Q^* \in \partial\mathcal{R}(X^*) \text{ such that } Q^* \in N_{\mathcal{X}}(X^*).$$

(b) *A sufficient condition for X^* to be an optimal solution to $\mathcal{P}_{\mathcal{D}}$, which moreover is necessary when $\mathcal{D}$ is finite everywhere on $\mathcal{L}^2$, is the existence of*

$$Y^* \in \partial\mathcal{D}(X^*) \text{ such that } -Y^* \in N_{\mathcal{X}}(X^*).$$

Proof. By definition, having $Y \in \partial\mathcal{R}(X^*)$ means having $\mathcal{R}(X) \geq \mathcal{R}(X^*) + E[(X - X^*)Y]$ for all $X \in \mathcal{L}^2$, but the positive homogeneity axiom R2 allows us to draw a special consequence by considering the cases where $X = \lambda X^*$ when $\lambda > 0$. Specifically, we must have for all $\lambda > 0$ that $\mathcal{R}(X^*)+E[(\lambda-1)X^*Y] \leq \mathcal{R}(\lambda X^*) = \lambda\mathcal{R}(X^*)$, hence $(\lambda-1)\mathcal{R}(X^*) \geq (\lambda-1)E[X^*Y]$; and because $\lambda-1$ can be either positive or negative in this context, we must therefore actually have $\mathcal{R}(X^*) = E[X^*Y]$.

The general subgradient inequality $\mathcal{R}(X) \geq \mathcal{R}(X^*) + E[(X - X^*)Y]$ reduces then to $\mathcal{R}(X) \geq E[XY]$, and with this holding for all $X \in \mathcal{L}^2$ we can conclude that $-Y \in \mathcal{Q}$ by relation (a) of Theorem 3. Conversely, if $-Y \in \mathcal{Q}$ and $\mathcal{R}(X^*) = E[X^*Y]$ we get from relation (a) that $\mathcal{R}(X) \geq \mathcal{R}(X^*) + E[(X - X^*)Y]$ for all $X \in \mathcal{L}^2$.

The claims about $\partial\mathcal{D}(X^*)$ then follow from the fact that, since $\mathcal{D}(X) = \mathcal{R}(X) + EX$, we have $Y \in \partial\mathcal{D}(X^*)$ if and only if $Y - 1 \in \partial\mathcal{R}(X^*)$.

A sufficient condition for the minimum of a convex functional $\mathcal{F}$ with respect to a convex set $\mathcal{X}$ to be attained at X^* is the existence of some $Y \in N_{\mathcal{X}}(X^*)$ such that $-Y \in \partial\mathcal{F}(X^*)$. Moreover, this condition is also necessary when $\mathcal{X}$ is closed, $\mathcal{F}$ is lower semicontinuous, and there is at least one point of $\mathcal{X}$ at which $\mathcal{F}$ is actually continuous. (The sufficiency is an elementary consequence of the definitions of subgradients and normals; for the well known necessity under the additional assumption, see e.g. the Example 1 on p. 57 of [9].) All we need to do is to apply this fact to $\mathcal{F} = \mathcal{R}$ and $\mathcal{F} = \mathcal{D}$, utilizing our knowledge of the subgradients of these functionals. □

It is worth recording that, in the case of a coherent deviation measure $\mathcal{D}$, the second version of the general description of $\partial\mathcal{D}(X)$ in Theorem 5 can be cast in the form

$$\partial\mathcal{D}(X) = \{\, 1 - Q \mid Q \in \mathcal{Q},\ \mathcal{D}(X) = E_Q[EX - X] \,\}.$$

This may be compared with the description of $\mathcal{D}(X)$ just ahead of Example 11.

The conditions in Theorem 5 can be applied for instance to a set $\mathcal{X}$ of the kind in Example 16 by utilizing the normal cone formula in Proposition 3. All that remains for such an application is the task of fleshing out the subgradients that come up, and this can be accomplished in referring back to the formulas already uncovered for the risk envelopes $\mathcal{Q}$ associated with various choices of $\mathcal{D}$ and $\mathcal{R}$.

Example 18 (subgradients for standard deviation)**.** *For $\mathcal{D}(X) = \sigma(X)$ and nonconstant X^*, the subgradient set $\partial\mathcal{D}(X^*)$ consists of a unique element Y, namely $Y = \sigma(X^*)^{-1}[X^* - EX^*]$. Correspondingly, for $\mathcal{R}(X) = \rho\sigma(X) - \mu(X)$ with $\rho > 0$, the subgradient set $\partial\mathcal{R}(X^*)$ consists uniquely of the element $-Q = \rho\sigma(X^*)^{-1}[X^* - EX^*] - 1$.*

Detail. This comes out of Example 14 and the subgradient rules in the first part of Theorem 5. □

Example 19 (subgradients for worst-case risk). *For $\mathcal{R}(X) = -\inf X$, the subgradient set $\partial\mathcal{R}(X^*)$ consists of all $-Q$ obtained from elements $Q \in \mathcal{L}^2$ (if any) having the following description as functions on Ω (up to almost sure equivalence): $EQ = 1$ with $Q(\omega) \geq 0$ where $X^*(\omega) = \inf X^*$, but $Q(\omega) = 0$ where $X^*(\omega) > \inf X^*$.*

Detail. The functions Q having this description clearly are the ones belonging to the set $\mathcal{Q}_{X^*}$ of Theorem 5 for the risk envelope $\mathcal{Q}$ identified in Example 11. □

In Example 19, $\mathcal{R}(X^*)$ might be the empty set for some choices of X^*, even ones such that $\mathcal{R}(X^*)$ is finite (unless the state space Ω is itself finite). In general, the set $\mathcal{Q}_X$ in the subgradient formulas in Theorem 5 is nonempty for *every* X if and only if the risk envelope $\mathcal{Q}$ is bounded (and therefore compact in the weak topology of $\mathcal{L}^2$). That corresponds to the functionals $\mathcal{R}$ and $\mathcal{D}$ being finite everywhere on $\mathcal{L}^2$.

Example 20 (subgradients for CVaR). *For $\mathcal{R}(X) = \mathrm{CVaR}_\alpha(X)$ with $\alpha \in (0,1)$, and any X^*, the subgradient set $\partial\mathcal{R}(X^*)$ consists of all $-Q$ obtained from elements $Q \in \mathcal{L}^2$ having the following description as functions on Ω (up to almost sure equivalence):*

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

For $\mathcal{D}(X) = \mathrm{CVaR}^\Delta_\alpha(X)$, the subgradient set $\partial\mathcal{D}(X^*)$ consists of all $1 - Q$ derived from such Q.

Detail. Such elements Q emerged in the justification of Example 12 as furnishing the maximizing set $\mathrm{argmax}\{-E[X^*Q] \mid Q \in \mathcal{Q}\}$ in the case of the risk envelope corresponding to $\mathcal{R}(X) = \mathrm{CVaR}_\alpha(X)$. That observation then feeds into the subgradient formulas in Theorem 5. □

6 Conclusions

We have endeavored to complement the ground-breaking work of Artzner, Delbaen, Eber and Heath by demonstrating that when risk measures in their sense are applied to $X - EX$ instead of to a return r.v. X itself — in the presence of a new property of expectation-boundedness — a parallel class of functionals, appropriately termed deviation measures, arises. This should help bridge the gap between the theory of risk measures and the way that risk is typically viewed by practitioners in the finance industry. To further this aim, we have explored the distinction between risk measures and deviation measures in a range of key examples, providing insights also in terms of risk acceptance notions and their dualization by risk envelopes.

An interpretation with wide-ranging possibilities has been offered through the idea of acceptably sure inequalities, which might in some circumstances be violated but only to a degree that an assessor of risk would deem negligible to the point of not being worth the worry. Out of this notion comes that of an "acceptably free lunch", which we have shown to influence whether the minimization of deviation is equivalent to the minimization of risk in financial optimization.

Utilizing tools of convex analysis, we have furthermore developed a scheme for determining optimality in problems of minimizing deviation or risk. We have shown that risk envelopes have major significance in solving such problems and have illustrated the particulars in a number of settings. The results open the way for many applications and advances in their numerical methodology.

References

[1] C. Acerbi, Spectral Measures of Risk: a Coherent Representation of Subjective Risk Aversion, *Journal of Banking and Finance* **26** (2002), 1505–1518.

[2] C. Acerbi, D. Tasche, On the Coherence of Expected Shortfall, *Journal of Banking and Finance* **26** (2002), 1487–1503.

[3] P. Artzner, F. Delbaen, J.-M. Eber, D. Heath, Coherent Measures of Risk, *Mathematical Finance* **9** (1999), 203–227.

[4] E. Asplund, L. Bungart, *A First Course in Integration*, Holt, Rinehart and Winston, 1966.

[5] F. Delbaen, Coherent Risk Measures on General Probability Spaces, preprint 2000.

[6] F. Delbaen, *Draft: Coherent Risk Measures*, lecture notes, Pisa, 2000.

[7] G.C. Pflug, Some Remarks on the Value-at-Risk and the Conditional Value-at-Risk, in *Probabilistic Constrainted Optimization: Methodology and Applications* (S.P. Uryasev, ed.), Kluwer, 2000, 278–287.

[8] R.T. Rockafellar, *Convex Analysis*, Princeton University Press, 1970.

[9] R.T. Rockafellar, *Conjugate Duality and Optimization*, monograph No. 16 in Conference Board of Math. Sciences Series, SIAM Publications, 1974.

[10] R.T. Rockafellar, *Integrals which are convex functionals*, Pacific Journal of Math. **24** (1968), 525–540.

[11] R.T. Rockafellar, S.P. Uryasev, Optimization of Conditional Value-at-Risk, *Journal of Risk* **2** (2000), 21–42.

[12] R.T. Rockafellar, S.P. Uryasev, Conditional Value-at-Risk for General Loss Distributions, *Journal of Banking and Finance* **26** (2002), 1443–1471.

[13] R. T. Rockafellar, R. J.-B. Wets, *Variational Analysis*, Springer-Verlag, Berlin, 1998.

[14] A. Roell, Risk Aversion in Quiggin and Yaari's Rank-Order Model of Choice Under Uncertainty, *The Economic Journal* **97** (1987) [Issue Supplement: Conference papers], 143–159.

[15] M.E. Yaari, The Dual Theory of Choice Under Risk, *Econometrica* **55** (1987), 95–115.