J Optim Theory Appl (2012) 155:1105–1123
DOI 10.1007/s10957-011-9968-2

Entropic Value-at-Risk: A New Coherent Risk Measure

A. Ahmadi-Javid

Received: 24 February 2011 / Accepted: 12 November 2011 / Published online: 17 December 2011

Abstract This paper introduces the concept of *entropic value-at-risk* (EVaR), a new coherent risk measure that corresponds to the tightest possible upper bound obtained from the Chernoff inequality for the *value-at-risk* (VaR) as well as the *conditional value-at-risk* (CVaR). We show that a broad class of stochastic optimization problems that are computationally intractable with the CVaR is efficiently solvable when the EVaR is incorporated. We also prove that if two distributions have the same EVaR at all confidence levels, then they are identical at all points. The dual representation of the EVaR is closely related to the Kullback-Leibler divergence, also known as the relative entropy. Inspired by this dual representation, we define a large class of coherent risk measures, called *g-entropic* risk measures. The new class includes both the CVaR and the EVaR.

Keywords Chernoff inequality · Coherent risk measure · Conditional value-at-risk (CVaR) · Convex optimization · Cumulant-generating function · Duality · Entropic value-at-risk (EVaR) · g-entropic risk measure · Moment-generating function · Relative entropy · Stochastic optimization · Stochastic programming · Value-at-risk (VaR)

1 Introduction

In the last twenty years, a great deal of effort has gone into achieving suitable methods of measuring risk. A risk measure assigns to a random outcome or risk position a real number that scalarizes the degree of risk involved in that random outcome. This concept has found many applications in different areas such as finance, actuary science, operations research and management science.

A. Ahmadi-Javid (✉)
Department of Industrial Engineering, Amirkabir University of Technology (Tehran Polytechnic), Tehran, Iran
e-mail: ahmadi_javid@aut.ac.ir

Most theoretical research papers studying risk measures take two main approaches. In the first approach, after specifying a set of desirable properties for a risk measure, the class of risk measures satisfying those properties is studied. In the second approach, a particular risk measure is investigated from different points of view, including interpretation, axiomatic properties, alternative representations, computational issues, and optimization of models incorporating that type of risk measure. This paper introduces a new risk measure and thus fits in the category of the second approach. Preliminary versions of this paper appeared in conference papers [1, 2].

In the mathematical finance literature of last fourteen years, several papers have proposed general regularity conditions for a suitable risk measure. To regularize the measurement of market or nonmarket risks, Artzner et al. [3, 4] proposed in their seminal work, the concept of a *coherent* risk measure that has four basic properties: translation invariance (cash-additivity), monotonicity, subadditivity and positive homogeneity. Using classic convex duality, they showed that, for a finite dimensional probability space, each coherent risk measure has a dual representation. Delbaen [5] extended this result to a general probability space. He provided a similar representation result for a lower semi-continuous risk measure defined on the space of bounded measurable functions. Föllmer and Schied [6] and Frittelli and Rosazza Gianin [7] generalized coherent risk measures to the convex case by replacing the two properties of subadditivity and positive homogeneity with the property of convexity. They showed that, as with coherent risk measures, each convex lower semi-continuous risk measure admits a dual representation. Ruszczynski and Shapiro [8] and Kaina and Rüschendorf [9] studied convex risk measures over more flexible sets of random variables. Pflug [10] looked at dual representations of several known convex risk measures. See [11] for further technical details and other interesting results on convex risk measures until 2006. Also, see [12] for a survey of the new paradigm used for risk measures. Recently, other axiomatic characterizations have been considered. El Karoui and Ravanelli [13] and Cerreia-Vioglio et al. [14] studied convex and quasiconvex cash-subadditive risk measures, respectively. Balbas and Balbas [15] recently investigated the notion of compatibility between pricing rules and risk measures. Cont et al. [16] examined risk measures from a statistical point of view. For a review on conditional and dynamic risk measures and various time-consistency properties, see [17].

Similar approaches are considered in the actuary science literature. In this field, premium principles are common risk measures; see, for example [18] and Chapter 5 in [19]. Wang et al. [20] proposed specific desirable properties for risk measures used in pricing. Other risk measures are used to determine the provisions and capital requirements of an insurer, in order to prevent insolvency. Panjer [21] defined coherent risk measures in a similar way to [4]. Goovaerts et al. [22] presented an axiomatic characterization of risk measures that are additive for independent random variables. See [23] for additional issues.

The operations research and management science literature uses risk measures to designate stochastic objectives. This also arises in the actuary science and finance literatures when one wants to design a suitable portfolio with the minimum total loss or the maximum total return, both of which are random variables. Ruszczynski and Shapiro [8] studied convex and coherent risk measures from an optimization point of view. Rockafellar et al. [24] studied generalized deviations that are counterparts to risk measures. Other papers, such as [25–28], have tried to study stochastic programming models involving a general risk measure. See also [29] for a survey.

Unfortunately almost all popular risk measures introduced in the literature before the year of 2000 lack some of the axiomatic properties required for coherency. For example, the celebrated *mean-variance* risk measure introduced by Markowitz [30] fails to be monotone and is therefore not economically meaningful. This risk measure has dominated the asset-allocation process for over fifty years. Since it is now known that many return and loss distributions are not normally distributed, investor preferences go beyond mean and variance, and consequently, the popularity of this risk measure has decreased over the last decade. Another well-known risk measure is the *value-at-risk* (VaR), which has been popular for over a decade and has even been written into industry regulations ([31], see also [32] or pp. 13–19 [33], for more on the history of the VaR). However, the VaR is not coherent, since it lacks subadditivity. Other shortcomings of the VaR are that, for non-normal returns or losses, working numerically with the VaR is unstable, and optimizing models involving the VaR are intractable in high dimensions. See [34, 35] for more details and references on the drawbacks of the VaR. Note that throughout this paper, we refer to a problem—whether it is an optimization problem or a computing procedure—as *efficiently computable*, *efficiently solvable*, *computationally tractable* or *tractable* if it can be solved in a reasonable amount of time, and as *intractable* or *computationally intractable* if it cannot.

The *conditional value-at-risk* (CVaR), studied by Rockafellar and Uryasev [34, 35] has gained popularity in recent years. It measures the expected loss in the left tail given that a particular threshold has been crossed. This overcomes a weakness of the VaR, that is, lack of control on the extent of the losses that might be suffered beyond the threshold amount indicated by the VaR. An important shortcoming of the CVaR is that it cannot be computed efficiently, even for the sum of arbitrary independent random variables. In fact, there are only a few cases for which the CVaR can be computed efficiently, while for other cases, we need to approximate the CVaR through sampling methods. There are also many other interesting coherent risk measures available, such as spectral risk measures [36]; however, they cannot be efficiently computed even for simple cases. The need for an efficiently computable coherent risk measure becomes more serious when the risk measure is incorporated into a stochastic optimization problem where it needs to be computed frequently.

To the best of our knowledge, no flexible coherent risk measure that can be computed efficiently at least for the sum of arbitrary independent bounded random variables has been discovered in the literature. In this paper, we introduce a new coherent risk measure called the *entropic value-at-risk* (EVaR) which will be shown to be the tightest upper bound one can find using the Chernoff inequality for the VaR. The new risk measure is also an upper bound for the CVaR and its dual representation is related to the relative entropy. Finally, we shall see that the EVaR is efficiently computable in several cases where the CVaR is not.

The plan of the paper is as follows. In Sect. 2, we provide a mathematical introduction to risk measures and review a few of the most important risk measures. In Sect. 3, we introduce the new risk measure and investigate its properties. In Sect. 4, we study optimization problems incorporating the proposed risk measure. Lastly, in

Sect. 5, we present a generalization of the new risk measure. Our concluding remarks are in Sect. 6.

2 Mathematical Preliminaries of Risk Measures

A *risk measure* is a function ρ assigning a real value to a random variable X. This notion enables one to pick a suitable member from the set $\mathbf{X}$ of allowable random variables. If we prefer smaller values of $X \in \mathbf{X}$, then we use the following optimization problem to make the selection:

$$\min_{X \in \mathbf{X}} \rho(X). \tag{1}$$

To define the concept of a risk measure more precisely, let $(\Omega, \mathbf{F}, P)$ be a probability space where Ω is a set of all simple events, $\mathbf{F}$ is a σ-algebra of subsets of Ω and P is a probability measure on $\mathbf{F}$. Moreover, suppose that $\mathbf{L}$ is the set of all Borel measurable functions (random variables) $X : \Omega \to \mathbb{R}$, and $\mathbf{X} \subseteq \mathbf{L}$ is a subspace including all real numbers. Now we define the risk measure $\rho : \mathbf{X} \to \overline{\mathbb{R}}$, where $\overline{\mathbb{R}} = \mathbb{R} \cup \{-\infty, +\infty\}$ is the extended real line. For $p \geq 1$ let $\mathbf{L}_p$ be the set of all Borel measurable functions $X : \Omega \to \mathbb{R}$ for which $\mathrm{E}(|X|^p) = \int |X|^p \, dP < +\infty$, $\mathbf{L}_\infty$ be the set of all bounded Borel measurable functions, $\mathbf{L}_M$ be the set of all Borel measurable functions $X : \Omega \to \mathbb{R}$ whose moment-generating function $M_X(z) = \mathrm{E}(e^{zX})$ exists for all $z \in \mathbb{R}$, and $\mathbf{L}_{M^+}$ be the set of all Borel measurable functions $X : \Omega \to \mathbb{R}$ whose moment-generating function $M_X(z)$ exists for all $z \geq 0$. Note that $\mathbf{L}_\infty \subseteq \mathbf{L}_M \subseteq \mathbf{L}_p$ for all $p \geq 1$.

The literature introduces a number of properties that are used to determine a suitable risk measure. The following outlines the most important properties for the risk measure $\rho : \mathbf{X} \to \overline{\mathbb{R}}$:

(P1) Translation Invariance: $\rho(X + c) = \rho(X) + c$ for any $X \in \mathbf{X}$ and $c \in \mathbb{R}$;
(P2) Subadditivity: $\rho(X_1 + X_2) \leq \rho(X_1) + \rho(X_2)$ for all $X_1, X_2 \in \mathbf{X}$;
(P3) Monotonicity: If $X_1, X_2 \in \mathbf{X}$ and $X_1 \leq X_2$, then $\rho(X_1) \leq \rho(X_2)$;
(P4) Positive Homogeneity: $\rho(\lambda X) = \lambda \rho(X)$ for all $X \in \mathbf{X}$ and $\lambda \geq 0$.

Artzner et al. [4] define a coherent risk measure as one that satisfies P1–P4. Note that our definition of a coherent risk measure is taken from the operations research and actuarial science literature, which differs slightly from the one used in the mathematical finance literature, where the risk measure ρ is coherent if $\psi(X) = \rho(-X)$ satisfies the above four properties.

Previous studies have proposed a number of risk measures that have been used by researchers. We briefly review some of the most significant examples.

The *expectation* and *worst-case* risk measures are the two elementary coherent risk measures:

$$\rho_{\mathrm{E}}(X) := \mathrm{E}(X), \quad X \in \mathbf{L}_1, \qquad \rho_{\mathrm{W}}(X) := \operatorname{ess\,sup}(X), \quad X \in \mathbf{L}.$$

Unfortunately, most other risk measures that are frequently used in the literature are not coherent. For example, the *value-at-risk (VaR) with confidence level* $1-\alpha$, that is,

$$\mathrm{VaR}_{1-\alpha}(X) := \inf_{t \in \mathbb{R}} \{t : \Pr(X \leq t) \geq 1 - \alpha\}, \quad X \in \mathbf{L}, \; \alpha \in \,]0, 1],$$

lacks the subadditivity property. As for the *mean-standard-deviation* risk measure,

$$\text{M-SD}_\lambda(X) := \text{E}(X) + \lambda\sqrt{\text{var}(X)}, \quad X \in \mathbf{L}_2,\ \lambda > 0,$$

it does not have the property of monotonicity.

The *conditional value-at-risk* (CVaR) is an important coherent risk measure that was introduced and studied recently in [34, 35]. The CVaR with confidence level $1-\alpha$ is defined as follows:

$$\text{CVaR}_{1-\alpha}(X) := \inf_{t\in\mathbb{R}}\left\{t + \frac{1}{\alpha}\text{E}[X-t]_+\right\}, \quad X \in \mathbf{L}_1, \tag{2}$$

where $[s]_+ = \max\{0, s\}$ and $\alpha \in]0, 1]$. This measure can be interpreted by the VaR in this way:

$$\text{CVaR}_{1-\alpha}(X) = \frac{1}{\alpha}\int_0^\alpha \text{VaR}_{1-t}(X)\,dt.$$

This means that $\text{CVaR}_{1-\alpha}(X)$ is the mean of the worst $\alpha\%$ of values of X, and that for small values of α, it focuses on the worst losses of the random outcome. $\text{CVaR}_{1-\alpha}(X)$ is therefore more sensitive than $\text{VaR}_{1-\alpha}(X)$ to the shape of the distribution of X in the right tail.

3 Entropic Value-at-Risk

In this section, we propose a new coherent risk measure that is the tightest possible upper bound obtained from the Chernoff inequality for the VaR. The Chernoff inequality [37] for any constant a and $X \in \mathbf{L}_{M^+}$ is as follows:

$$\Pr(X \geq a) \leq e^{-za}M_X(z), \quad \forall z > 0.$$

By solving the equation $e^{-za}M_X(z) = \alpha$ with respect to a for $\alpha \in]0, 1]$, we obtain

$$a_X(\alpha, z) := z^{-1}\ln(M_X(z)/\alpha),$$

for which we have $\Pr(X \geq a_X(\alpha, z)) \leq \alpha$. In fact, for each $z > 0$, $a_X(\alpha, z)$ is an upper bound for $\text{VaR}_{1-\alpha}(X)$. We now consider the best upper bound of this type as a new risk measure that bounds $\text{VaR}_{1-\alpha}(X)$ by using exponential moments. Note that the special case $a_X(1, z)$ has long been known in the actuarial literature as the *exponential premium* [18, 38]. In the finance literature [6, 47], it has recently been considered as a convex risk measure, which is called the *entropic risk measure* in [39, Example 4.33].

Definition 3.1 The *entropic value-at-risk* (EVaR) of $X \in \mathbf{L}_{M^+}$ *with confidence level* $1-\alpha$ is defined as follows:

$$\text{EVaR}_{1-\alpha}(X) := \inf_{z>0}\left\{a_X(\alpha, z)\right\} = \inf_{z>0}\left\{z^{-1}\ln(M_X(z)/\alpha)\right\}.$$

In fact, the EVaR is the tightest possible upper bound that can be obtained from the Chernoff inequality for the VaR. We now prove that this risk measure is coherent.

Lemma 3.1 *The function* $\kappa_\alpha(X,t) := a_X(\alpha, t^{-1})$, *with* $X \in \mathbf{L}_{M^+}$ *and* $t > 0$, *is convex in* (X,t) *for all* $\alpha \in]0,1]$.

Proof We must show that, for all $\lambda \in [0,1]$, $X, Y \in \mathbf{L}_{M^+}$ and $t_1, t_2 > 0$:

$$\lambda\kappa_\alpha(X,t_1) + (1-\lambda)\kappa_\alpha(Y,t_2) \ge \kappa_\alpha\big(\lambda X + (1-\lambda)Y, \lambda t_1 + (1-\lambda)t_2\big),$$

which is equivalent to

$$\begin{aligned}&\lambda t_1 \ln M_X\big(t_1^{-1}\big) + (1-\lambda)t_2 \ln M_Y\big(t_2^{-1}\big)\\ &\quad\ge \big(\lambda t_1 + (1-\lambda)t_2\big) \ln M_{\lambda X+(1-\lambda)Y}\big(\big(\lambda t_1 + (1-\lambda)t_2\big)^{-1}\big).\end{aligned}$$

Denoting $t = \lambda t_1 + (1-\lambda)t_2$ and $w = \lambda t_1/t$, the left-hand side of the above inequality can be expressed as

$$t\big(w \ln M_X\big(t_1^{-1}\big) + (1-w)\ln M_Y\big(t_2^{-1}\big)\big).$$

From Jensen's inequality (note $0 \le w,\ 1-w \le 1$)) one has

$$w \ln M_X\big(t_1^{-1}\big) \ge \ln E\big[\big(e^{Xt_1^{-1}}\big)^w\big], \qquad (1-w)\ln M_Y\big(t_2^{-1}\big) \ge \ln E\big[\big(e^{Yt_2^{-1}}\big)^{1-w}\big],$$

and then by using the known fact that the cumulant-generating function is convex, it yields

$$\begin{aligned}&t\big(w \ln M_X\big(t_1^{-1}\big) + (1-w)\ln M_Y\big(t_2^{-1}\big)\big)\\ &\quad\ge t\big(\ln E\big(e^{\lambda X t^{-1}}\big) + \ln E\big(e^{(1-\lambda)Yt^{-1}}\big)\big)\\ &\quad\ge t \ln E\big(e^{\lambda X t^{-1} + (1-\lambda)Yt^{-1}}\big)\\ &\quad= \big(\lambda t_1 + (1-\lambda)t_2\big) \ln M_{\lambda X+(1-\lambda)Y}\big(\big(\lambda t_1 + (1-\lambda)t_2\big)^{-1}\big).\end{aligned}$$

This completes the proof. □

Theorem 3.1 *The* $\mathrm{EVaR}_{1-\alpha}$ *is coherent for every* $\alpha \in]0,1]$.

Proof Verifying properties P1, P3 and P4 is straightforward. By proving that the EVaR is convex and by taking into account property P4, we see that property P2 is also implied.

The EVaR can be written as $\mathrm{EVaR}_{1-\alpha}(X) = \inf_{t>0}\{\kappa_\alpha(X,t)\}$, where the right-hand side is a convex function because, by Lemma 3.1, $\kappa_\alpha(X,t)$ is jointly convex in (X,t). □

The following theorem gives an important property of coherent risk measures, which was presented in Artzner et al. [4] for the discrete case, and then extended for a general case by Delbaen [5].

Theorem 3.2 *For every coherent risk measure $\rho : \mathbf{L}_\infty \to \mathbb{R}$ with the Fatou property, there exists a set of probability measures $\Im$ on $(\Omega, \mathbf{F})$ such that*

$$\rho(X) = \sup_{Q\in\Im} \mathrm{E}_Q(X).$$

This is known as the dual representation or robust representation of ρ. Moreover, having the above expression means that the risk measure is a coherent one.

Note that the Fatou property is much weaker than lower semicontinuity; however, when the argument space is $\mathbf{L}_\infty$, they are equivalent for coherent risk measures [18, Theorem 4.31]. Moreover, the above representation is related to the conjugate of ρ and holds by the Fenchel-Moreau theorem. See [8] for more technical details.

The dual representation of the EVaR reveals its relationship with the relative entropy. The next theorem provides this representation by applying the Donsker-Varadhan variational formula (see [40], pp. 405–408).

Lemma 3.1 (Donsker-Varadhan Variational Formula) *For any $X \in \mathbf{L}_\infty$,*

$$\ln \mathrm{E}_P\big(e^X\big) = \sup_{Q\ll P} \{\mathrm{E}_Q(X) - D_{KL}(Q \parallel P)\},$$

where $D_{KL}(Q \parallel P) := \int \frac{dQ}{dP}(\ln \frac{dQ}{dP})\, dP$ is the relative entropy of Q with respect to P, or the Kullback–Leibler divergence from Q to P.

Theorem 3.3 *The dual representation of* $\mathrm{EVaR}_{1-\alpha}(X)$ *for $X \in \mathbf{L}_\infty$ has the form*

$$\mathrm{EVaR}_{1-\alpha}(X) = \sup_{Q\in\Im} \mathrm{E}_Q(X),$$

where $\Im = \{Q \ll P : D_{KL}(Q \parallel P) \le -\ln\alpha\}$.

Proof By Lemma 3.1 we have

$$\begin{aligned}
\mathrm{EVaR}_{1-\alpha}(X) &= \inf_{z>0}\big\{z^{-1}\ln\big(M_X(z)/\alpha\big)\big\} \\
&= \inf_{z>0}\big\{z^{-1}\ln \mathrm{E}_P\big(e^{zX}\big) - z^{-1}\ln\alpha\big\} \\
&= \inf_{z>0}\Big\{z^{-1}\sup_{Q\ll P}\{\mathrm{E}_Q(zX) - D_{KL}(Q \parallel P)\} - z^{-1}\ln\alpha\Big\} \\
&= \inf_{z>0}\Big\{\sup_{Q\ll P}\big\{\mathrm{E}_Q(X) - z^{-1}\big(\ln\alpha + D_{KL}(Q \parallel P)\big)\big\}\Big\} \\
&= \sup_{Q\ll P, D_{KL}(Q\parallel P)\le -\ln\alpha} \{\mathrm{E}_Q(X)\}.
\end{aligned}$$

Thus, the proof is completed. □

From Definition 3.1 we have

$$\mathrm{EVaR}_{1-\alpha}(X) = \inf_{z>0}\bigl\{a_X(1,z) - z^{-1}\ln\alpha\bigr\} = \inf_{z>0}\bigl\{z^{-1}\ln M_X(z) - z^{-1}\ln\alpha\bigr\}$$

which shows that the EVaR only depends on $a_X(1,z)$, the moment-generating function or the cumulant-generating function. The next proposition demonstrates how these functions can be represented by means of the EVaR. Note that cumulant-generating functions play a key role in the large deviation theory [40] and the quantity $a_X(1,z)$ is of interest for the actuary and finance literature, as discussed in the paragraph preceding Definition 3.1.

Proposition 3.1 *For $X \in \mathbf{L}_{M^+}$ and $z > 0$,*

$$\ln M_X(z) = \sup_{0<\alpha\le 1}\bigl\{z\mathrm{EVaR}_{1-\alpha}(X) + \ln\alpha\bigr\},$$

$$M_X(z) = \sup_{0<\alpha\le 1}\bigl\{\alpha\exp(z\mathrm{EVaR}_{1-\alpha}(X))\bigr\},$$

$$a_X(1,z) = \sup_{0<\alpha\le 1}\bigl\{\mathrm{EVaR}_{1-\alpha}(X) + z^{-1}\ln\alpha\bigr\}.$$

Proof Based on the equation

$$-\mathrm{EVaR}_{1-\alpha}(X) = \sup_{x\ge 0}\{x\ln\alpha - g(x)\}$$

with

$$g(x) = \begin{cases} x\ln M_X(x^{-1}) & x > 0,\\ \operatorname{ess\,sup}(X) & x = 0,\end{cases}$$

one can observe that the function $-\mathrm{EVaR}_{1-e^y}(X)$ with domain $]-\infty, 0]$ is the conjugate of the function $g(x)$ with domain $[0, +\infty[$. Being convex and closed, the function $g(x)$ is the conjugate of its own conjugate. This completes the proof. □

The fact that the moment-generating function can be deduced from the EVaR unveils an important property of the EVaR, which is given in the following corollary.

Corollary 3.1 *For $X, Y \in \mathbf{L}_M$, $\mathrm{EVaR}_{1-\alpha}(X) = \mathrm{EVaR}_{1-\alpha}(Y)$ for all $\alpha \in]0,1]$ if and only if $F_X(b) = F_Y(b)$ for all $b \in \mathbb{R}$.*

Proof The proof follows from Proposition 3.1 and from the well-known property of moment-generating functions stating that two distributions are identical if they have the same moment-generating functions. Because of the assumption $X, Y \in \mathbf{L}_M$, the moment-generating functions $M_X(b)$ and $M_Y(b)$ are smooth and analytic; thus, the identity of $M_X(b)$ and $M_Y(b)$ for all $b \ge 0$ implies their identity for all $b \in \mathbb{R}$. □

This property indicates that the $\mathrm{EVaR}_{1-\alpha}(X)$ as a function of its parameter α characterizes the distribution of $X \in \mathbf{L}_M$. The condition $X, Y \in \mathbf{L}_M$ in the above corollary can be weakened to the existence of $M_X(b)$ and $M_Y(b)$ over the interval $b \in]-\varepsilon, +\infty]$ for some positive constant $\varepsilon > 0$.

The next proposition investigates another property for the EVaR. From the discussion preceding Definition 3.1, it is clear that the EVaR is an upper bound for the VaR. Here, we compare the EVaR and the CVaR.

Proposition 3.2 *The* EVaR *is an upper bound for both the* VaR *and the* CVaR *with the same confidence levels*, *i.e.*, *for* $X \in \mathbf{L}_{M^+}$ *and every* $\alpha \in]0, 1]$

$$\mathrm{CVaR}_{1-\alpha}(X) \le \mathrm{EVaR}_{1-\alpha}(X).$$

Moreover,

$$\mathrm{E}(X) \le \mathrm{EVaR}_{1-\alpha}(X) \le \operatorname{ess\,sup}(X),$$

where $\mathrm{EVaR}_0(X) = \mathrm{E}(X)$ *and* $\lim_{\alpha\to 0} \mathrm{EVaR}_{1-\alpha}(X) = \operatorname{ess\,sup}(X)$.

Proof If $\mathrm{EVaR}_{1-\alpha}(X) \le l$ holds for a real number l, then there exists $z > 0$ such that $E(e^{(X-l)z}) \le \alpha$, and consequently $\frac{1}{\alpha}E[X - l + \frac{1}{z}]_+ - \frac{1}{z} \le 0$ (note that $[s + 1]_+ \le e^s$ for all $s \in \mathbb{R}$). Therefore from the translation-invariance property of the CVaR, we obtain the implication:

$$\mathrm{EVaR}_{1-\alpha}(X) \le l \quad \Rightarrow \quad \mathrm{CVaR}_{1-\alpha}(X) \le l.$$

This means that if $\mathrm{EVaR}_{1-\alpha}(X) \le l$ for some $l \in \mathbb{R}$, then $\mathrm{CVaR}_{1-\alpha}(X) \le l$. By setting $l = \mathrm{EVaR}_{1-\alpha}(X)$ within this, we obtain the inequality $\mathrm{CVaR}_{1-\alpha}(X) \le \mathrm{EVaR}_{1-\alpha}(X)$. Given $X \in \mathbf{L}_\infty$, another proof for this part can be provided using the dual representations of the EVaR and the CVaR, given in Theorem 3.3 and (11), respectively. The other parts are straightforward to prove. □

This proposition shows that the EVaR is more risk-averse compared to the CVaR at the same confidence level. Hence, the EVaR suggests a financial or insurance company allocating more resources to avoid risk. However, this solution may not be favored by companies that want to allocate the lowest possible amount of resources. This issue makes the EVaR less attractive for such companies, but its most important feature is that it is computationally tractable for several cases where the CVaR is not. In fact, when we need to incorporate a risk measure in a stochastic optimization problem, it is very important that it be computationally tractable. In the next section, we discuss the computational tractability aspect of the EVaR.

We end this section by comparing the VaR, CVaR and EVaR for two examples. It is easy to show that for $X \sim N(\mu, \sigma^2)$

$$\mathrm{VaR}_{1-\alpha}(X) = \mu + z_\alpha \sigma,$$

$$\mathrm{CVaR}_{1-\alpha}(X) = \mu + \frac{\phi(z_\alpha)}{\alpha}\sigma,$$

$$\mathrm{EVaR}_{1-\alpha}(X) = \mu + \sqrt{-2\ln\alpha}\,\sigma,$$

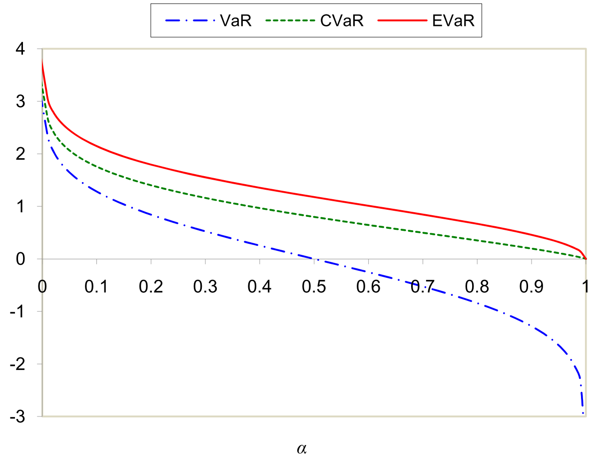

Fig. 1 Comparing the VaR, CVaR and EVaR for the standard normal distribution

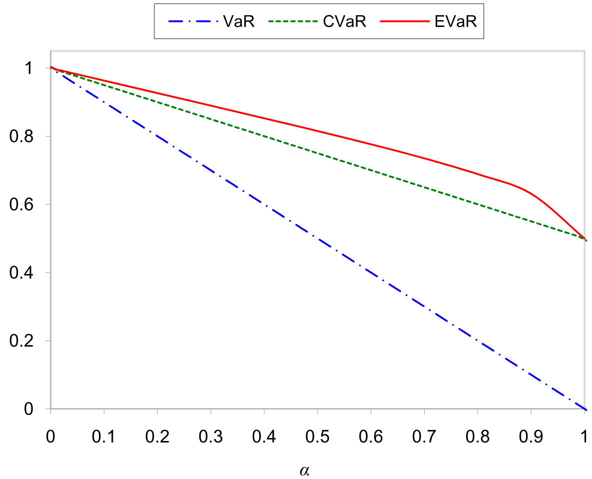

Fig. 2 Comparing the VaR, CVaR and EVaR for the uniform distribution over interval]0, 1[

where $\phi(.)$ and z_α are the density function and the upper α-percentile of the standard normal, respectively. Moreover, for $X \sim U(a,b)$, we have

$$\mathrm{VaR}_{1-\alpha}(X) = a + (1-\alpha)(b-a),$$

$$\mathrm{CVaR}_{1-\alpha}(X) = a + \frac{1}{2}(2-\alpha)(b-a),$$

$$\mathrm{EVaR}_{1-\alpha}(X) = \inf_{t>0}\left\{ t\ln\left(t\frac{e^{t^{-1}b} - e^{t^{-1}a}}{b-a}\right) - t\ln\alpha \right\}.$$

Figures 1 and 2 compare the VaR, CVaR and EVaR for $X \sim N(0,1)$ and $X \sim U(0,1)$, respectively. As is indicated by Proposition 3.2, the EVaR is an upper bound for both the VaR and the CVaR.

4 Optimization

In practice, we use composite random variables such as $X = G(\boldsymbol{w}, \boldsymbol{\Psi})$ where $\boldsymbol{w}$ is an n-dimensional real decision vector in $\mathbf{W} \subseteq \mathbb{R}^n$, $\boldsymbol{\Psi}$ is an m-dimensional real random vector with a known probability distribution, and the function $G(\boldsymbol{w}, .) : \mathbb{R}^m \to \mathbb{R}$ is a Borel measurable function for all values $\boldsymbol{w} \in \mathbf{W}$. Then, problem (1) becomes

$$\min_{\boldsymbol{w}\in\mathbf{W}} \rho\big(G(\boldsymbol{w}, \boldsymbol{\Psi})\big). \tag{3}$$

It can be easily shown that (3) is a convex optimization problem if the risk measure ρ is coherent and $G(., \boldsymbol{s})$ is convex for all $\boldsymbol{s} \in S_{\boldsymbol{\Psi}}$ where $S_{\boldsymbol{\Psi}}$ is the support of the random vector $\boldsymbol{\Psi}$.

In this section, we consider problem (3), when ρ is the EVaR, focusing on the computational tractability. In this case, problem (3) can be rewritten as the following optimization problem:

$$\min_{\boldsymbol{w}\in\mathbf{W}, t>0} \big\{t \ln M_{G(\boldsymbol{w},\boldsymbol{\Psi})}\big(t^{-1}\big) - t \ln \alpha\big\}. \tag{4}$$

If the function $G(., \boldsymbol{s})$ is convex for all $\boldsymbol{s} \in S_{\boldsymbol{\Psi}}$, then the objective function of the above problem is also convex. Moreover, this problem (4) is computationally tractable if the objective function can be computed efficiently. This requires that the term $M_{G(\boldsymbol{w},\boldsymbol{\Psi})}(t^{-1})$ be evaluated in polynomial time for any given values of $\boldsymbol{w}$ and t. Fortunately, it is entirely tractable for the following important and useful form:

$$G(\boldsymbol{w}, \boldsymbol{\Psi}) = g_0(\boldsymbol{w}) + \sum_{i=1}^{m} g_i(\boldsymbol{w})\Psi_i, \quad g_i : \mathbb{R}^n \to \mathbb{R},\ i = 0, 1, \ldots, m,$$

under one of the following mild assumptions about the distribution of $\boldsymbol{\Psi} = (\Psi_1, \ldots, \Psi_m)^t$:

1. $\Psi_1, \ldots, \Psi_m$ are independent random variables in $\mathbf{L}_M$;
2. $\Psi_i = \sum_{j=1}^{q} \alpha_{ij}\Gamma_j$, $i = 1, \ldots, m$, where $\Gamma_1, \ldots, \Gamma_q$ are independent random variables in $\mathbf{L}_M$, and $\alpha_{ij} \in \mathbb{R}$;
3. $\boldsymbol{\Psi} = \boldsymbol{\zeta} + \boldsymbol{\Xi}$, where the elements of $\boldsymbol{\zeta}$ are independent random variables in $\mathbf{L}_M$, $\boldsymbol{\zeta}$ is independent of $\boldsymbol{\Xi}$, and $\boldsymbol{\Xi}$ is a random vector with a distribution whose moment-generating function exists and is efficiently computable (e.g. multivariate normal distributions);
4. $\boldsymbol{\Psi}$ has a discrete distribution with a finite support set whose cardinality grows polynomially with the dimension of $\boldsymbol{\Psi}$.

If we apply the CVaR instead of the EVaR in problem (3), it reduces to the following optimization problem:

$$\min_{\boldsymbol{w}\in\mathbf{W}, t\in\mathbb{R}} \left\{t + \frac{1}{\alpha}\mathrm{E}\left[g_0(\boldsymbol{w}) + \sum_{i=1}^{m} g_i(\boldsymbol{w})\Psi_i - t\right]_+\right\}, \tag{5}$$

which is computationally intractable when the dimension of $\boldsymbol{\Psi}$ increases, for almost all instances of the three first cases, even the first case, where all random variables Ψ_i are independent discrete random variables that take a finite number of distinct values. This case is analyzed in the following proposition.

Proposition 4.1 *Let $\Psi_1, \ldots, \Psi_m$ be independent discrete random variables that take k distinct values and $G(\boldsymbol{w}, \boldsymbol{\Psi})$ be affine in $\boldsymbol{\Psi}$ for all $\boldsymbol{w} \in \mathbf{W}$. For fixed values of $\boldsymbol{w}$ and t, the complexity of computing the objective function given in problem (4) is a bilinear function of m and k, i.e., mk, whereas the computing time for the objective function given in problem (5) is of order k^m, that is, it grows exponentially with m and polynomially with k.*

Proof Suppose that $\Psi_1, \ldots, \Psi_m$ are independent and that each Ψ_i takes only k distinct values, namely, for each i there exist k distinct constants $a_{i,j}$, $j = 1, \ldots, k$ such that $\sum_{j=1}^{k} \Pr(\Psi_i = a_{i,j}) = 1$, $i = 1, \ldots, m$. The joint distribution of $\Psi_1, \ldots, \Psi_m$ takes k^m different scenarios, i.e., $\boldsymbol{s}_1, \ldots, \boldsymbol{s}_{k^m}$ with probabilities $p_1, \ldots, p_{k^m}$.

Problem (4) can be written as

$$\min_{\boldsymbol{w} \in \mathbf{W}, t>0} \left\{ g_0(\boldsymbol{w}) + t \sum_{i=1}^{m} \ln M_{g_i(\boldsymbol{w})\Psi_i}\left(t^{-1}\right) - t \ln \alpha \right\}.$$

The inner term involving moment-generating functions can then be presented as follows:

$$\sum_{i=1}^{m} \ln M_{g_i(\boldsymbol{w})\Psi_i}\left(t^{-1}\right) = \sum_{i=1}^{m} \ln\left(\sum_{j=1}^{k} \Pr(\Psi_i = a_{i,j}) \times \exp\left(t^{-1} g_i(\boldsymbol{w}) a_{i,j}\right)\right).$$

The complexity of computing this quantity is of order mk.

The inner expectation in the objective of problem (5) is computed as

$$\mathrm{E}\left[g_0(\boldsymbol{w}) + \sum_{i=1}^{m} g_i(\boldsymbol{w})\Psi_i - t \right]_+ = \sum_{l=1}^{k^m} p_l \left[g_0(\boldsymbol{w}) + \sum_{i=1}^{m} g_i(\boldsymbol{w}) s_{l,i} - t \right]_+,$$

where $s_{l,i}$ is the ith component of $\boldsymbol{s}_l$. Clearly, the complexity of computing this quantity is of order k^m. □

For example, if the summation of two numbers takes 10^{-12} seconds and $k = 2$, $m = 100$, one needs about 4×10^{10} years to compute the objective of problem (5) for fixed values of $\boldsymbol{w}$ and t, but the evaluation of the objective of problem (4) only needs about 10^{-10} seconds, which is dramatically less than 4×10^{10} years. This simple analysis demonstrates how important the EVaR can be from a numerical optimization perspective.

One might think that problem (5) can be solved approximately through the *sample average approximation* (SAA) method, as suggested by Rockafellar and Uryasev [35]. The SAA method may be the only known practical way of handling problem (5) that approximates the expectation in the objective by a Monte Carlo sampling method. However, this method is not accurate unless the number of scenarios is sufficiently large. Unfortunately, as we increase the number of scenarios, solving the resulting optimization problem becomes more difficult. In the following, we show that even for simple, practical cases with a reasonable desired accuracy level, the optimization problem involves billions of variables and constraints and therefore cannot be solved practically. To proceed, we first need to present a preliminary introduction to the SAA method.

Let $\boldsymbol{x}$ be an n-dimensional real decision vector in $\mathbf{X} \subseteq \mathbb{R}^n$, $\boldsymbol{\xi}$ an m-dimensional real random vector with a known probability distribution, and $F(\boldsymbol{x},.) : \mathbb{R}^m \to \mathbb{R}$ a Borel measurable function for all values $\boldsymbol{x} \in \mathbf{X}$. Consider the following stochastic optimization problem:

$$\min_{\boldsymbol{x}\in\mathbf{X}} \mathrm{E}\bigl(F(\boldsymbol{x},\boldsymbol{\xi})\bigr). \tag{6}$$

This problem can be approximated using the SAA method as follows:

$$\min_{\boldsymbol{x}\in\mathbf{X}} \frac{1}{N}\sum_{j=1}^{N} F(\boldsymbol{x},\boldsymbol{\xi}_j), \tag{7}$$

where $\boldsymbol{\xi}_j$, $j = 1,\dots,N$ is a random sample from the distribution of $\boldsymbol{\xi}$. The next theorem analyzes the relation between optimal solutions of problems (6) and (7). For more details on the SAA method, the reader may refer to [41].

Theorem 4.1 *Suppose that* $\mathbf{X} \subseteq \mathbb{R}^n$ *has the finite diameter* $D := \sup_{\boldsymbol{x},\boldsymbol{x}'\in\mathbf{X}} \|\boldsymbol{x}-\boldsymbol{x}'\|$ *and that there exists a positive constant* L *such that*

$$\bigl|F(\boldsymbol{x},\mathbf{z}) - F\bigl(\boldsymbol{x}',\mathbf{z}\bigr)\bigr| \le L\bigl\|\boldsymbol{x}-\boldsymbol{x}'\bigr\| \quad \forall \boldsymbol{x},\ \boldsymbol{x}' \in \mathbf{X},\ \mathbf{z} \in S_{\boldsymbol{\xi}},$$

where $S_{\boldsymbol{\xi}}$ *is the support of the bounded random vector* $\boldsymbol{\xi}$. *If the sample size* N *is greater than*

$$O(1)\left(\frac{DL}{\varepsilon}\right)^2\left(n\ln\left(\frac{DL}{\varepsilon}\right)+\ln\left(\frac{O(1)}{\delta}\right)\right)$$

($O(1)$ *denotes a generic positive constant*), *then an* $(\varepsilon/2)$*—optimal solution of problem* (7) *is an* ε*—optimal solution of problem* (6) *with probability at least* $1-\delta$($\bar{\boldsymbol{y}}$ *is an* ε*-optimal solution of problem* $\min_{\boldsymbol{y}\in\mathbf{Y}} f(\boldsymbol{y})$ *if* $f(\bar{\boldsymbol{y}}) \le \min_{\boldsymbol{y}\in\mathbf{Y}} f(\boldsymbol{y}) + \varepsilon$).

Now consider the following portfolio optimization problem, which is a simple and practical special case of problem (5):

$$\min_{w\in\mathbf{W}} \mathrm{CVaR}_{1-\alpha}\left(-\sum_{i=1}^{n} w_i R_i\right), \tag{8}$$

with $\mathbf{W} = \{w_i \ge 0, i = 1,\dots,n : \sum_{i=1}^{n} w_i = C\}$. $R_1,\dots,R_n$ are the random return rates of assets $1,\dots,n$. The aim here is to invest a certain capital C in these assets where $w_1,\dots,w_n$ denote the investment values. Using the SAA method, this problem can be approximated by the following problem:

$$\min_{\boldsymbol{w}\in\mathbf{W},t\in\mathbb{R}} \left\{t + \frac{1}{\alpha}\sum_{j=1}^{N}\frac{1}{N}\left[-\sum_{i=1}^{n} w_i r_{j,i} - t\right]_+\right\}, \tag{9}$$

where $r_{j,i}$ is the ith component of $\boldsymbol{r}_j$, and $\boldsymbol{r}_j$, $j = 1, \dots, N$ is a random sample from the distribution of $\boldsymbol{R} = (R_1, \dots, R_n)^t$. To solve this problem efficiently, we need to linearize it as follows:

$$\begin{aligned}
\min \quad & t + \frac{1}{\alpha}\sum_{j=1}^{N}\frac{1}{N}y_j \\
\text{s.t.} \quad & \\
& w_i \geq 0, \quad i = 1, \dots, n, \quad \sum_{i=1}^{n} w_i = C, \ t \in \mathbb{R} \\
& -\sum_{i=1}^{n} w_i r_{j,i} - t \leq y_j, \quad j = 1, \dots, N \\
& y_j \geq 0, \quad j = 1, \dots, N.
\end{aligned} \tag{10}$$

This model can be solved by commercial linear-programming solvers, as long as the number of variables $n + N$ and constraints $1 + n + 2N$ is not large. However, if one wants the solution obtained from problem (9), or equivalently (10), to be a good approximation for problem (8), then the sample size needs to be very large and therefore, the above linearized model cannot be solved practically. Consider the following result obtained from Theorem 4.1.

Proposition 4.2 *Suppose that $R_1, \dots, R_n$ are bounded random variables, i.e., that there exist constants l and u such that $l \leq R_1, \dots, R_n \leq u$, and define $B = u - l$. If*

$$\begin{aligned}
N \geq O(1)&\left(\frac{C\sqrt{2+B^2}(\frac{1}{\alpha}\sqrt{n\max\{u^2, l^2\}+1}+1)}{\varepsilon}\right)^2 \\
&\times \left(n\ln\left(\frac{C\sqrt{2+B^2}(\frac{1}{\alpha}\sqrt{n\max\{u^2, l^2\}+1}+1)}{\varepsilon}\right) + \ln\left(\frac{O(1)}{\delta}\right)\right),
\end{aligned}$$

then an $(\varepsilon/2)$-optimal solution of problem (9) (or equivalently (10)) with additional constraint $lC \leq -t \leq uC$ is an ε-optimal solution of problem (8) with probability at least $1 - \delta$.

Proof The proof follows from the result presented in Theorem 4.1. □

Note that adding the additional constraint $lC \leq -t \leq uC$ is necessary to make the feasible set of problem (8) bounded, which is a required condition for the validity of the result given in Theorem 4.1. One can add this constraint to problems (9) and (10) because the optimal value of t is $\mathrm{VaR}_{1-\alpha}(-\sum_{i=1}^{n} w_i R_i)$.

Consider the practical case where the capital available for investment is 100,000 units of money, the desired accuracy level for the approximation through the SAA method is 1000 units of money, the level of confidence of the SAA method is 0.99, the level of confidence of the CVaR is 0.95, the number of assets is 100, each return rate is a positive random variable, and the maximum variation of each return rate is 0.1, that is, $C = 10^5$, $\varepsilon = 10^3$, $\delta = 10^{-2}$, $\alpha = 0.05$, $n = 100$, $l = 0$, $u = 0.1$ and

$B = 0.1$. Then according to Proposition 4.2, the sample size N needs to satisfy

$$N \geq 14360747695 \times O(1) + 79379907 \times O(1) + 17237128 \times O(1).$$

This shows that the SAA method fails to practically approximate problem (8) as well as problem (5) even with low desired accuracy level ε. The reason is that increasing the sample size N significantly increases the size of problem (10), which is the smooth version of problem (9).

Finally, it is worth pointing out that if we use the EVaR in problem (8), instead of the CVaR, and then apply the SAA method, the resulting approximate problem is basically smooth and can be solved by optimization methods developed for smooth convex optimization problems.

5 Generalization

In this section, we generalize the EVaR based on its dual representation. To this end, we incorporate the class of generalized relative entropies, rather than the relative entropy, into the dual representation given in Theorem 3.3.

The generalized relative entropy of Q with respect to P, denoted by $H_g(P, Q)$, is an information-type *pseudo-distance* or *divergence measure* from Q to P:

$$H_g(P, Q) := \int g\left(\frac{dQ}{dP}\right) dP,$$

where g is a convex function with $g(1) = 0$. This quantity is an important non-symmetric (or directed) divergence measure, as initially discussed in [42, 43] (see [44, 45] for more details). Note that $H_g(P, Q) \geq 0$, and $H_g(P, Q) = 0$ if $Q = P$. For $g(x) = x \ln x$, one obtains the Kullback–Leibler generalization of Shannon's entropy [46], known as the relative entropy of Q with respect to P, or the Kullback–Leibler divergence from Q to P.

We now introduce the following class of risk measures.

Definition 5.1 Let g be a convex function with $g(1) = 0$, and β be a nonnegative number. The *g-entropic risk measure with divergence level β* is defined as

$$\mathrm{ER}_{g,\beta}(X) := \sup_{Q \in \Im} \mathrm{E}_Q(X),$$

where $\Im = \{Q \ll P : H_g(P, Q) \leq \beta\}$.

The next theorem provides a primal representation of the class of g-entropic risk measures by using the generalized Donsker-Varadhan variational formula considered in Theorem 4.2 of [47].

Lemma 5.1 (Generalized Donsker-Varadhan Variational Formula) *Let* $g : \mathbb{R} \to]-\infty, +\infty]$ *be a closed convex function with* $g(1) = 0$. *Then, for* $X \in \mathbf{L}_\infty$,

$$\inf_{\mu \in \mathbb{R}} \left\{\mu + \mathrm{E}_P\left(g^*(X - \mu)\right)\right\} = \sup_{Q \ll P} \left\{\mathrm{E}_Q(X) - H_g(P, Q)\right\},$$

where g^* is the conjugate (*the Legendre–Fenchel transform*) *of g. In particular, for*

$$g(x) = \begin{cases} x \ln x & x > 0, \\ 0 & x = 0, \\ +\infty & x < 0 \end{cases}$$

one retrieves the Donsker-Varadhan variational formula presented in Lemma 3.1.

Theorem 5.1 *Let g be a closed convex function with $g(1) = 0$ and β be a nonnegative number. Then, for $X \in \mathbf{L}_\infty$ the following holds*:

$$\mathrm{ER}_{g,\beta}(X) = \inf_{t>0, \mu \in \mathbb{R}} \left\{ t \left[\mu + \mathrm{E}_P \left(g^* \left(\frac{X}{t} - \mu + \beta \right) \right) \right] \right\}.$$

Proof By Lemma 5.1 one can write

$$\begin{aligned}
&\inf_{t>0, \mu \in \mathbb{R}} \left\{ t \left[\mu + \mathrm{E}_P \left(g^* \left(\frac{X}{t} - \mu + \beta \right) \right) \right] \right\} \\
&= \inf_{t>0} \left\{ t \left[\beta + \inf_{\mu \in \mathbb{R}} \left\{ (\mu - \beta) + \mathrm{E}_P \left(g^* \left(\frac{X}{t} - (\mu - \beta) \right) \right) \right\} \right] \right\} \\
&= \inf_{t>0} \left\{ t \left[\beta + \inf_{b \in \mathbb{R}} \left\{ b + \mathrm{E}_P \left(g^* \left(\frac{X}{t} - b \right) \right) \right\} \right] \right\} \\
&= \inf_{t>0} \left\{ t \left[\beta + \sup_{Q \ll P} \left\{ \mathrm{E}_Q \left(\frac{X}{t} \right) - H_g(P, Q) \right\} \right] \right\} \\
&= \inf_{t>0} \left\{ \sup_{Q \ll P} \{ \mathrm{E}_Q(X) + t(\beta - H_g(P, Q)) \} \right\} \\
&= \sup_{Q \ll P, H_g(P,Q) \le \beta} \{ \mathrm{E}_Q(X) \}.
\end{aligned}$$

This ends the proof. □

As an interesting example, we now show that the CVaR is a g-entropic risk measure. The dual representation of the CVaR is as follows:

$$\mathrm{CVaR}_{1-\alpha}(X) = \sup_{Q \in \Im} \mathrm{E}_Q(X), \tag{11}$$

where $\Im = \{Q : Q \le \frac{1}{\alpha} P\}$. The set $\Im$ can be expressed as $\{Q \ll P : \int g(\frac{dQ}{dP})\, dP \le 0\}$, where

$$g(x) = \begin{cases} 0 & 0 \le x \le \frac{1}{\alpha}, \\ +\infty & \text{otherwise}. \end{cases}$$

Moreover, the CVaR formula given in (2) can be obtained from the representation given in Theorem 5.1 by considering that $g^*(x) = \frac{1}{\alpha} \max\{0, x\}$ and $\beta = 0$. Note that the EVaR formula given in Definition 3.1 can also be obtained from Theorem 5.1 by setting

$$g(x) = \begin{cases} x \ln x & x > 0, \\ 0 & x = 0, \\ +\infty & x < 0 \end{cases}$$

with $g^*(x) = e^{x-1}$, and $\beta = -\ln \alpha$.

6 Concluding Remarks

In this paper, we introduce a new coherent risk measure called the *entropic value-at-risk* (EVaR). It is the tightest upper bound one can find from the Chernoff inequality for the *value-at-risk* (VaR) and the *conditional value-at-risk* (CVaR). The dual representation of the EVaR is related to the relative entropy. It is also shown that every random variable with a moment-generating function existing everywhere can be characterized using the associated EVaR values at all confidence levels.

Although recently, coherent risk measures have found applications in finance, actuary science and operations research, almost all known coherent risk measures like the CVaR cannot be computed efficiently, even for the sum of arbitrary independent random variables; however, a broad class of stochastic optimization problems involving the EVaR has the important property of being efficiently solvable.

We develop a new class of coherent risk measures that is inspired from the dual representation of the EVaR. A primal representation for this class is provided by applying a generalization of the Donsker-Varadhan Variational formula. This class includes the EVaR and the CVaR as particular cases.

Significant future research steps would involve introducing new risk measures that enable one to efficiently solve more complicated stochastic optimization problems. Moreover, the approach used to extend the EVaR can be implemented to develop other classes of coherent risk measures, for example, by applying other types of divergence measure, such as the Bergman or separable divergences.

Acknowledgements I first wish to thank editor-in-chief Professor Franco Giannessi for his kind consideration and valuable comments. I also appreciate the input of Professor Stan Uryasev, Professor Michael Zabarankin and Dr. Vahab Mirrokni on an earlier version of this paper. Parts of this paper were first presented in a seminar given in the summer of 2010 at the GERAD research center. I am grateful to former GERAD director Professor Roland Malhamé for arranging that seminar and for his helpful comments. I also extend my thanks to Dr. Hirbod Assa for his discussion following the seminar and for making points that significantly improved this paper. Last but not least, I wish to thank the three reviewers for their comments and time. Please note that, in earlier versions of this paper, which have circulated among my students and colleagues since 2004, I called the proposed risk measure *moment risk measure or moment value-at-risk*, but I recently changed its name to *entropic value-at-risk* (EVaR) to stress its connection with the VaR and the relative entropy.

References

1. Ahmadi-Javid, A.: Stochastic optimization via entropic value-at-risk: A new coherent risk measure. In: International Conference on Operations Research and Optimization, January 2011, Tehran, Iran (2011)

2. Ahmadi-Javid, A.: An information–theoretic approach to constructing coherent risk measures. In: Proceedings of IEEE International Symposium on Information Theory, August 2011, St. Petersburg, Russia, pp. 2125–2127 (2011)
3. Artzner, Ph., Delbaen, F., Eber, J.M., Heath, D.: Thinking coherently. Risk **10**, 68–71 (1997)
4. Artzner, Ph., Delbaen, F., Eber, J.M., Heath, D.: Coherent risk measures. Math. Finance **9**, 203–228 (1999)
5. Delbaen, F.: Coherent risk measures on general probability spaces. In: Sandmann, K., Schonbucher, P.J. (eds.) Advances in Finance and Stochastic, Essays in Honor of Dieter Sondermann, pp. 1–38. Springer, Berlin (2002)
6. Föllmer, H., Schied, A.: Convex risk measures and trading constraints. Finance Stoch. **6**, 429–447 (2002)
7. Frittelli, M., Rosazza Gianin, E.: Putting order in risk measures. J. Bank. Finance **26**, 1473–1486 (2002)
8. Ruszczynski, A., Shapiro, A.: Optimization of convex risk functions. Math. Oper. Res. **31**, 433–452 (2006)
9. Kaina, M., Ruschendorf, L.: On convex risk measures on L^p-spaces. Math. Methods Oper. Res. **69**, 475–495 (2009)
10. Pflug, G.Ch.: Subdifferential representations of risk measures. Math. Program. **108**, 339–354 (2006)
11. Schied, A.: Risk measures and robust optimization problems. Stoch. Models **22**, 753–831 (2006)
12. Balbas, A.: Mathematical methods in modern risk measurement: A survey. RACSAM Ser. Appl. Math. **101**, 205–219 (2007)
13. El Karoui, N., Ravanelli, C.: Cash sub-additive risk measures under interest rate ambiguity. Math. Finance **19**, 561–590 (2009)
14. Cerreia-Vioglio, S., Maccheroni, F., Marinacci, M., Montrucchio, L.: Risk measures: rationality and diversification. Math. Finance **21**, 743–774 (2011)
15. Balbas, A., Balbas, R.: Compatibility between pricing rules and risk measures: The CCVaR. RACSAM Ser. Appl. Math. **103**, 251–264 (2009)
16. Cont, R., Deguest, R., Scandolo, G.: Robustness and sensitivity analysis of risk measurement procedures. Quant. Finance **10**, 593–606 (2010)
17. Acciaio, B., Penner, I.: Dynamic risk measures. In: Di Nunno, G., Oksendal, B. (eds.) Advanced Mathematical Methods for Finance, pp. 1–34. Springer, Berlin (2011)
18. Goovaerts, M.J., De Vijlder, F., Haezendonck, J.: Insurance Premiums. North-Holland, Amsterdam (1984)
19. Kaas, R., Goovaerts, M.J., Dhaene, J., Denuit, M.: Modern Actuarial Risk Theory. Kluwer Academic, Dordrecht (2001)
20. Wang, S.S., Young, V.R., Panjer, H.H.: Axiomatic characterization of insurance prices. Insur. Math. Econ. **21**, 173–183 (1997)
21. Panjer, H.H.: Measurement of risk, solvency requirements and allocation of capital within financial conglomerates. Institute of Insurance and Pension Research, Research Report 1-15, University of Waterloo (2002)
22. Goovaerts, M.J., Kaas, R., Laeven, R.J.A., Tang, Q.: A comonotonic image of independence for additive risk measures. Insur. Math. Econ. **35**, 581–594 (2004)
23. Denuit, M., Dhaene, J., Goovaerts, M.J., Kaas, R., Laeven, R.J.A.: Risk measurement with equivalent utility principles. In: Ruschendorf, L. (ed.) Risk Measures: General Aspects and Applications (special issue). Statistics and Decisions, vol. 24, pp. 1–26 (2006)
24. Rockafellar, R.T., Uryasev, S., Zabarankin, M.: Generalized deviations in risk analysis. Finance Stoch. **10**, 51–74 (2006)
25. Rockafellar, R.T., Uryasev, S., Zabarankin, M.: Optimality conditions in portfolio analysis with generalized deviation measures. Math. Program. **108**, 515–540 (2006)
26. Ahmed, S.: Convexity and decomposition of mean-risk stochastic programs. Math. Program. **106**, 447–452 (2006)
27. Eichhorn, A., Romisch, W.: Polyhedral risk measures in stochastic programming. SIAM J. Optim. **16**, 69–95 (2005)
28. Miller, N., Ruszczyński, A.: Risk-averse two-stage stochastic linear programming: Modeling and decomposition. Oper. Res. **59**, 125–132 (2011)
29. Krokhmal, P., Zabarankin, M., Uryasev, S.: Modeling and optimization of risk. Surv. Oper. Res. Manag. Sci. **16**, 49–66 (2011)
30. Markowitz, H.: Portfolio Selection. J. Finance **7**, 77–91 (1952)

31. Pritsker, M.: Evaluating value at risk methodologies. J. Financ. Serv. Res. **12**, 201–242 (1997)
32. Guldimann, T.: The story of risk metrics. Risk **13**, 56–58 (2000)
33. Holton, G.: Value-at-Risk: Theory and Practice. Academic Press, San Diego (2002)
34. Rockafellar, R.T., Uryasev, S.: Optimization of conditional value-at-risk. J. Risk **2**, 21–41 (2000)
35. Rockafellar, R.T., Uryasev, S.: Conditional value-at-risk for general loss distributions. J. Bank. Finance **26**, 1443–1471 (2002)
36. Acerbi, C.: Spectral risk measures: A coherent representation of subjective risk aversion. J. Bank. Finance **26**, 1505–1518 (2002)
37. Chernoff, H.: A measure of asymptotic efficiency for tests of a hypothesis based on a sum of observations. Ann. Stat. **23**, 493–507 (1952)
38. Gerber, H.U.: On additive premium calculation principles. ASTIN Bull. **7**, 215–222 (1974)
39. Follmer, H., Schied, A.: Stochastic Finance: An Introduction in Discrete Time. de Gruyter, Berlin (2004)
40. Dupuis, P., Ellis, R.S.: A Weak Convergence Approach to the Theory of Large Deviations. Wiley, New York (1997)
41. Shapiro, A., Nemirovski, A.: On complexity of stochastic programming problems. In: Jeyakumar, V., Rubinov, A.M. (eds.) Continuous Optimization: Current Trends and Applications, pp. 111–144. Springer, New York (2005)
42. Ali, S.M., Silvey, S.D.: A general class of coefficients of divergence of one distribution from another. J. R. Stat. Soc. B **28**, 131–142 (1966)
43. Csiszar, I.: Information-type measures of difference of probability distributions and indirect observations. Studia Sci. Math. Hung. **2**, 299–318 (1967)
44. Liese, F., Vajda, I.: Convex Statistical Distances. Teubner, Leipzig (1987)
45. Ullah, A.: Entropy, divergence and distance measures with econometric applications. J. Stat. Plan. Inference **49**, 137–162 (1996)
46. Kullback, S., Leibler, R.A.: On information and sufficiency. Ann. Math. Stat. **22**, 79–86 (1951)
47. Ben-Tal, A., Teboulle, M.: An old-new concept of convex risk measures: the optimized certainty equivalent. Math. Finance **17**, 449–476 (2007)