Conditional Value-at-Risk for General Loss Distributions

R. Tyrrell Rockafellar[1] **and Stanislav Uryasev**[2]

Abstract. Fundamental properties of conditional value-at-risk, as a measure of risk with significant advantages over value-at-risk, are derived for loss distributions in finance that can involve discreetness. Such distributions are of particular importance in applications because of the prevalence of models based on scenarios and finite sampling. Conditional value-at-risk is able to quantify dangers beyond value-at-risk, and moreover it is coherent. It provides optimization short-cuts which, through linear programming techniques, make practical many large-scale calculations that could otherwise be out of reach. The numerical efficiency and stability of such calculations, shown in several case studies, are illustrated further with an example of index tracking.

Key Words: Value-at-risk, conditional value-at-risk, mean shortfall, coherent risk measures, risk sampling, scenarios, hedging, index tracking, portfolio optimization, risk management

Version of November 28, 2001

Correspondence should be addressed to Stanislav Uryasev

[1]University of Washington, Department of Mathematics, Box 354350, Seattle, WA 98195-4350; e-mail: rtr@math.washington.edu

[2]University of Florida, Risk Management and Financial Engineering Lab, Department of Industrial and Systems Engineering, PO Box 116595, Gainesville, FL 32611-6595; e-mail: uryasev@ufl.edu; URL: http://www.ise.ufl.edu/uryasev

1 Introduction

Measures of risk have a crucial role in optimization under uncertainty, especially in coping with the losses that might be incurred in finance or the insurance industry. Loss can be envisioned as a function $z = f(x, y)$ of a decision vector $x \in X \subset I\!R^n$ representing what we may generally call a portfolio, with X expressing decision constraints, and a vector $y \in Y \subset I\!R^m$ representing the future values of a number of variables like interest rates or weather data. When y is taken to be random with known probability distribution, z comes out as a random variable having its distribution dependent on the choice of x. Any optimization problem involving z in terms of the choice of x should then take into account not just expectations, but also the "riskiness" of x.

Value-at-risk, or VaR for short, is a popular measure of risk which has achieved the high status of being written into industry regulations. It suffers, however, from being unstable and difficult to work with numerically when losses are not "normally" distributed—which in fact is often the case, because loss distributions tend to exhibit "fat tails" or empirical discreteness. Moreover, VaR fails to be coherent in the sense of Artzner, Delbaen, Eber and Heath [6].

A very serious shortcoming of VaR, in addition, is that it provides no handle on the extent of the losses that might be suffered beyond the threshold amount indicated by this measure. It is incapable of distinguishing between situations where losses that are worse may be deemed only a little bit worse, and those where they could well be overwhelming. Indeed, it merely provides a lowest bound for losses in the tail of the loss distribution and has a bias toward optimism instead of the conservatism that ought to prevail in risk management.

An alternative measure that does quantify the losses that might be encountered in the tail is *conditional* value-at-risk, or CVaR. As a tool in optimization modeling, CVaR has superior properties in many respects. It maintains consistency with VaR by yielding the same results in the limited settings where VaR computations are tractable, i.e., for normal distributions (or perhaps "elliptical" distributions as in [17]); for portfolios blessed with such simple distributions, working with CVaR, VaR, or minimum variance [29] are equivalent (cf. [39]). Most importantly for applications, however, CVaR can be expressed by a remarkable minimization formula. This formula can readily be incorporated into problems of optimization with respect to $x \in X$ that are designed to minimize risk or shape it within bounds. Significant shortcuts are thereby achieved while preserving crucial problem features like convexity.

Such computational advantages of CVaR over VaR are turning into a major stimulus for the development of CVaR methodology, in view of the fact that efficient algorithms for optimization of VaR in high-dimensional settings are still not available, despite the substantial efforts that have gone into research in that direction [3, 7, 20, 21, 22, 26, 37, 43].

CVaR and its minimization formula were first developed in our paper [39]. There, we demonstrated numerical effectiveness through several case studies, including portfolio optimization and options hedging. In follow-up work in [34], investigations were carried out with the minimization of CVaR subject to a constraint on expected return, the maximization of return subject to a constraint on the CVaR, and the maximization of a utility function that balances CVaR against return. Strategies for investigating the efficient frontier between CVaR and return were considered as well. In [4], the approach was applied to credit risk management of a portfolio of bonds. Extensions in [12] have centered on a closely related notion of CDaR, conditional *drawdown*-at-risk, in the optimization of portfolios with drawdown constraints.

In these works, with their focus on demonstrating the potential of the new approach, discussion of CVaR in its full generality was postponed. Only continuous loss distributions were treated, and in fact, for the sake of an elementary initial justification of the minimization formula so as to get started with using it, distributions were assumed to have smooth density. In the present paper we drop those limitations and complete the foundations for our methodology. This step is needed of course not just for theory, but because many problems of optimization under uncertainty involve discontinuous loss distributions in which the discrete probabilities come out of scenario models or the finite sampling of random variables. While some consequences of our minimization formula itself have since been explored by Pflug [32] in territory outside of the assumptions we made in [39], an understanding of what the quantity given by the formula then represents in the usual framework of risk measures in finance has been missing.

For continuous loss distributions, the CVaR at a given confidence level is the expected loss given that the loss is greater than the VaR at that level, or for that matter, the expected loss given that the loss is greater than or equal to the VaR. For distributions with possible discontinuities, however, it has a more subtle definition and can differ from either of those quantities, which for convenience in comparision can be designated by CVaR^+ and CVaR^-, respectively. CVaR^+ has sometimes been called "mean shortfall" (cf. [31], although the seemingly identical term "expected shortfall" has been interpreted in other ways in [1] and [2], with the latter paper taking it as a synonym for CVaR itself), while "tail VaR" is a term that has been suggested for CVaR^- (cf. [6]). Here, in order to consolidate ideas and reduce the potential for confusion, we speak of CVaR^+ and CVaR^- simply as "upper" and "lower" CVaR. Generally $\text{CVaR}^- \leq \text{CVaR} \leq \text{CVaR}^+$, with equality holding when the loss distribution function does not have a jump at the VaR threshold;

but when a jump does occur, which for scenario models is *always* the situation, both inequalities can be strict.

On the basis of the general definition of CVaR elucidated below, and with the help of arguments in [32], CVaR is seen to be a *coherent* measure of risk in the sense of [6], whereas CVaR$^+$ and CVaR$^-$ are not. (A direct alternative proof of this fact has very recently been furnished by Acerbi, Nordio and Sirtori [1].) The lack of coherence of CVaR$^+$ and CVaR$^-$ in the presence of discreteness does not seem to be widely appreciated, although this shortcoming was already noted for CVaR$^-$ by the authors of [6]. They suggested, as a remedy, still another measure of risk which they called "worst conditional expectation" and proved to be coherent. That measure is impractical for applications, however, because it can only be calculated in very narrow circumstances. In contrast, CVaR is not only coherent but eminently practical by virtue of our minimization formula for it. That formula opens the door to computational techniques for dealing with risk far more effectively than before.

Interestingly, CVaR can be viewed as a weighted average of VaR and CVaR$^+$ (with the weights depending, like these values themselves, on the decision x). This seems surprising, in the face of neither VaR nor CVaR$^+$ being coherent. The weights arise from the particular way that CVaR "splits the atom" of probability at the VaR value, when one exists.

Besides laying out such implications of the general definition of CVaR and its associated minimization formula, we put effort here into bringing out properties of CVaR that enhance the usefulness of this approach when dealing with fully discrete distributions. For such distributions, we furnish an elementary way of calculating CVaR directly. We show how a suitable specification of the confidence level, depending on the finite, discrete distribution of y, can ensure that CVaR=CVaR$^+$ regardless of the choice of x. For confidence levels close enough to 1, we prove that CVaR, CVaR$^-$ and VaR coincide with maximum loss, and again this can be ensured independently of x.

We go over the optimization shortcuts offered by CVaR and extend them to models where risk is shaped at several confidence levels. As part of this, CVaR is proved to be stable with respect to the choice of the confidence level, although other proposed measures of risk are not.

Finally, we illustrate the main facts and ideas with a numerical example of portfolio replication with CVaR constraints. This example demonstrates how the incorporation of such constraints in a financial model may improve both the in-sample and the out-of-sample risk characteristics. The calculations confirm that CVaR methodology offers a management tool for efficiently controlling risks in practice.

Broadly speaking, problems of risk management with VaR and CVaR can be classified as falling under the heading of stochastic optimization. Various other concepts of risk in optimization have earlier been studied in the stochastic programming literature, but not in a context of finance; see [10, 19, 24, 25, 33, 41, 36]. The reader interested in applications of stochastic optimization techniques in the finance area can find relevant papers in [46, 47].

For elucidation of the many statements in this paper that rely on background in convex optimization, we refer the reader to the book [38] (or [40]).

Additional properties of CVaR, including a powerful result on estimation, are available in the new paper of Acerbi and Tasche [2].

2 General Concept of CVaR

In everything that follows, we suppose the random vector y is governed by a probability measure P on Y (a Borel measure) that is independent of x. (The independence could be relaxed for some purposes, but it is essential for key results about convexity that underly the use of linear programming reductions in computation.) For each x, we denote by $\Psi(x, \cdot)$ on $I\!R$ the resulting distribution function for the loss $z = f(x, y)$, i.e.,

$$\Psi(x, \zeta) = P\{y \,|\, f(x, y) \leq \zeta\}, \tag{1}$$

making the technical assumptions that $f(x, y)$ is continuous in x and measurable in y, and that $E\{|f(x, y)|\} < \infty$ for each $x \in X$. We denote by $\Psi(x, \zeta^-)$ the left limit of $\Psi(x, \cdot)$ at ζ; thus

$$\Psi(x, \zeta^-) = P\{y \,|\, f(x, y) < \zeta\}. \tag{2}$$

When the difference

$$\Psi(x, \zeta) - \Psi(x, \zeta^-) = P\{y \,|\, f(x, y) = \zeta\} \tag{3}$$

is positive, so that $\Psi(x, \cdot)$ has a jump at ζ, a probability "atom" is said to be present at ζ.

We consider a confidence level $\alpha \in (0, 1)$, which in applications would be something like $\alpha = .95$ or $\alpha = .99$. At this confidence level, there is a corresponding *value-at-risk*, defined in the following way.

Definition 1 (VaR). *The α-VaR of the loss associated with a decision x is the value*

$$\zeta_\alpha(x) = \min\{\zeta \,|\, \Psi(x, \zeta) \geq \alpha\}. \tag{4}$$

The minimum in (4) is attained because $\Psi(x,\zeta)$ is nondecreasing and right-continuous in ζ. When $\Psi(x,\cdot)$ is continuous and strictly increasing, $\zeta_\alpha(x)$ is simply the unique ζ satisfying $\Psi(x,\zeta)=\alpha$. Otherwise, this equation can have no solution or a whole range of solutions.

The case of no solution corresponds to a vertical gap in the graph of $\Psi(x,\cdot)$ as in Figure 1, with α lying in an interval of confidence levels that all yield the same VaR. The lower and upper endpoints of that interval are

$$\alpha^-(x) = \Psi(x,\zeta_\alpha(x)^-), \qquad \alpha^+(x) = \Psi(x,\zeta_\alpha(x)). \tag{5}$$

The case of a whole range of solutions corresponds instead to a constant segment of the graph, as shown in Figure 2. The solutions form an interval having $\zeta_\alpha(x)$ as its lower endpoint. The upper endpoint of the interval is the value $\zeta_\alpha^+(x)$ introduced next.

Definition 2 (VaR$^+$)**.** *The α-VaR$^+$ ("upper" α-VaR) of the loss associated with a decision x is the value*

$$\zeta_\alpha^+(x) = \inf\{\zeta \mid \Psi(x,\zeta) > \alpha\}. \tag{6}$$

Obviously $\zeta_\alpha(x) \le \zeta_\alpha^+(x)$ always, and these values are the same except when $\Psi(x,\zeta)$ is constant at level α over a certain ζ-interval. That interval is either $[\zeta_\alpha(x),\zeta_\alpha^+(x))$ or $[\zeta_\alpha(x),\zeta_\alpha^+(x)]$, depending on whether or not $\Psi(x,\cdot)$ has a jump at $\zeta_\alpha^+(x)$.

Both Figure 1 and Figure 2 illustrate phenomena that raise challenges in the treatment of general loss distributions. This is especially true for discrete distributions associated with finite sampling or scenario modeling, since then $\Psi(x,\cdot)$ is a *step* function (constant between jumps), and there is no getting around these circumstances.

Observe, for instance, that the situation in Figure 2 entails a discontinuity in the behavior of VaR: a jump is sure to occur if a slightly higher confidence level is demanded. This degree of instability is distressing for a measure of risk on which enormous sums of money might be riding. Furthermore, although x is fixed in this picture, examples easily show that the misbehavior in the dependence of VaR on α can effect its dependence on x as well. That makes it hard to cope successfully with VaR-centered problems of optimization in x.

These troubles, and many others, motivate the search for a better measure of risk than value-at-risk for practical applications. Such a measure is *conditional value-at-risk.*

Definition 3 (CVaR)**.** *The α-CVaR of the loss associated with a decision x is the value*

$$\phi_\alpha(x) = \text{ mean of the } \alpha\text{-tail distribution of } z = f(x,y), \tag{7}$$

Figure 1: Equation $\Psi(x, \zeta) = \alpha$ has no solution in ζ.

Figure 2: Equation $\Psi(x, \zeta) = \alpha$ has many solutions in ζ.

Figure 3: Distribution function $\Psi_\alpha(x,\zeta)$ is obtained by rescaling the function $\Psi(x,\zeta)$ in the interval $[\alpha,1]$.

where the distribution in question is the one with distribution function $\Psi_\alpha(x,\cdot)$ defined by

$$\Psi_\alpha(x,\zeta) = \begin{cases} 0 & \textit{for } \zeta < \zeta_\alpha(x), \\ [\Psi(x,\zeta)-\alpha]/[1-\alpha] & \textit{for } \zeta \geq \zeta_\alpha(x). \end{cases} \tag{8}$$

Note that $\Psi_\alpha(x,\cdot)$ truly is another *distribution* function, like $\Psi(x,\cdot)$: it is nondecreasing and right-continuous, with $\Psi_\alpha(x,\zeta) \to 1$ as $\zeta \to \infty$. The α-tail distribution referred to in (7) is thus well defined through (8).

The subtlety of Definition 3 resides in the case where the loss with distribution function $\Psi(x,\cdot)$ has a probability atom at $\zeta_\alpha(x)$, as illustrated in Figure 1. In that case the interval $[\zeta_\alpha(x),\infty)$ has probability greater than $1-\alpha$, inasmuch as

$$\Psi(x,\zeta_\alpha(x)^-) < \alpha \leq \Psi(x,\zeta_\alpha(x)) \quad \text{when} \quad \Psi(x,\zeta_\alpha(x)^-) < \Psi(x,\zeta_\alpha(x)), \tag{9}$$

and the issue comes up of what really should be meant by the α-tail distribution, since that term presumably ought to refer to the "upper $1-\alpha$ part" of the full distribution. This is resolved by specifying the α-tail distribution through the distribution function in (8), which is obtained by rescaling the portion of the graph of the original distribution between the horizontal lines at levels $1-\alpha$ and 1 so that it spans instead between 0 and 1. For the case shown in Figure 1, the result is depicted in Figure 3.

The consequences of this maneuver will be examined in relation to the following variants in which the whole interval $[\zeta_\alpha(x), \infty)$ or its interior $(\zeta_\alpha(x), \infty)$ are the focus.

Definition 4 (CVaR$^+$ and CVaR$^-$)**.** *The α-CVaR$^+$ ("upper" α-CVaR) of the loss associated with a decision x is the value*

$$\phi_\alpha^+(x) = E\{f(x,y) \,|\, f(x,y) > \zeta_\alpha(x)\}, \tag{10}$$

whereas the α-CVaR$^-$ ("lower" α-CVaR) of the loss is the value

$$\phi_\alpha^-(x) = E\{f(x,y) \,|\, f(x,y) \geq \zeta_\alpha(x)\}. \tag{11}$$

The conditional expectation in (11) is well defined because $P\{f(x,y) \,|\, f(x,y) \geq \zeta_\alpha(x)\} \geq 1 - \alpha > 0$, but the one in (10) only makes sense as long as $P\{f(x,y) \,|\, f(x,y) > \zeta_\alpha(x)\} > 0$, i.e., $\Psi(x, \zeta_\alpha(x)) < 1$, which is not assured merely through our assumption that $\alpha \in (0,1)$, since there might be a probability atom at $\zeta_\alpha(x)$ large enough to cover the interval $1 - \alpha^-(x)$.

As indicated in the introduction, (10) is sometimes called "mean shortfall". The closely related expression

$$E\{f(x,y) - \zeta_\alpha(x) \,|\, f(x,y) > \zeta_\alpha(x)\} \;=\; \phi_\alpha^+(x) - \zeta_\alpha(x) \tag{12}$$

goes however by the name of "mean excess loss"; cf. [8], [18]. In ordinary language, a shortfall might be thought the same as an excess loss, so "mean shortfall" for (10) potentially poses a conflict. The conditional expectation in (11) has been dubbed in [6] the "tail VaR" at level α, but as revealed in the proof of the next proposition, it is really the mean of the tail distribution for the confidence level $\alpha^-(x)$ in (5) rather than the one appropriate to α itself. The "upper" and "lower" terminology in Definition 4 avoids such difficulties while emphasizing the basic relationships among these values that are described next.

Proposition 5 (basic CVaR relations)**.** *If there is no probability atom at $\zeta_\alpha(x)$, one simply has*

$$\phi_\alpha^-(x) \;=\; \phi_\alpha(x) \;=\; \phi_\alpha^+(x). \tag{13}$$

If a probability atom does exist at $\zeta_\alpha(x)$, one has

$$\phi_\alpha^-(x) \;<\; \phi_\alpha(x) \;=\; \phi_\alpha^+(x) \quad \textit{when } \alpha = \Psi(x, \zeta_\alpha(x)), \tag{14}$$

or on the other hand,

$$\phi_\alpha^-(x) \;=\; \phi_\alpha(x) \quad \textit{when } \Psi(x, \zeta_\alpha(x)) = 1 \tag{15}$$

(with $\phi_\alpha^+(x)$ then being ill defined). But in all the remaining cases, characterized by

$$\Psi(x, \zeta_\alpha(x)^-) \;<\; \alpha \;<\; \Psi(x, \zeta_\alpha(x)) \;<\; 1, \tag{16}$$

one has the strict inequality

$$\phi_\alpha^-(x) \;<\; \phi_\alpha(x) \;<\; \phi_\alpha^+(x). \tag{17}$$

Proof. In comparison with the definition of $\phi_\alpha(x)$ in (7), the $\phi_\alpha^+(x)$ value in (10) is the mean of the loss distribution associated with

$$\Psi_\alpha^+(x, \zeta) = \begin{cases} 0 & \text{for } \zeta < \zeta_\alpha(x), \\ [\Psi(x, \zeta) - \alpha^+(x)]/[1 - \alpha^+(x)] & \text{for } \zeta \geq \zeta_\alpha(x), \end{cases} \tag{18}$$

whereas the $\phi_\alpha^-(x)$ value in (11) is the mean of the loss distribution associated with

$$\Psi_\alpha^-(x, \zeta) = \begin{cases} 0 & \text{for } \zeta < \zeta_\alpha(x), \\ [\Psi(x, \zeta) - \alpha^-(x)]/[1 - \alpha^-(x)] & \text{for } \zeta \geq \zeta_\alpha(x), \end{cases} \tag{19}$$

Recall that $\alpha^+(x)$ and $\alpha^-(x)$, defined in (5), mark the top and bottom of the vertical gap at $\zeta_\alpha(x)$ for the original distribution function $\Psi(x, \cdot)$ (if a jump occurs there).

The case of there being no probability atom at $\zeta_\alpha(x)$ corresponds to having $\alpha^-(x) = \alpha^+(x) = \alpha \in (0, 1)$. Then (13) obviously holds, because the distribution functions in (8), (18) and (19) are identical. When a probability atom exists, but $\alpha = \alpha^+(x)$, we get $\alpha^-(x) < \alpha^+(x) < 1$ and thus the relations in (14), while if $\alpha^+(x) = 1$ we nevertheless get (15) through (9). Under the alternative of (16), however, it is clear from the definitions of the distribution functions in (8), (18) and (19) that the strict inequalities in (17) prevail. □

For the situation in Figure 1, the distribution functions in (18) and (19) that have $\phi_\alpha^+(x)$ and $\phi_\alpha^-(x)$ as their means are illustrated in Figures 4 and 5. They are the tail distributions for the confidence levels $\alpha^+(x)$ and $\alpha^-(x)$.

Proposition 5 confirms, in the case in (13), that α-CVaR throughly reduces for *continuous* loss distributions (i.e., ones without any probability atoms induced by discreteness) to the more elementary expressions for conditional value-at-risk that we worked with in [39]. An important task ahead will be to demonstrate that the minimization formula we developed in [39], which is vital to the feasibility of practical applications of CVaR in risk management, carries over from that special context to the present one.

The α-CVaR and the α-CVaR$^+$ of the loss coincide often, but not always, according to Proposition 5. Another perspective on the connection between these two values is developed next.

Figure 4: Distribution function $\Psi_{\alpha}^{+}(x,\zeta)$ is obtained by rescaling the function $\Psi(x,\zeta)$ in the interval $[\alpha^{+}(x)$,1].

Figure 5: Distribution function $\Psi_{\alpha}^{-}(x,\zeta)$ is obtained by rescaling the function $\Psi(x,\zeta)$ in the interval $[\alpha^{-}(x)$,1].

Proposition 6 (CVaR as a weighted average)**.** *Let $\lambda_\alpha(x)$ be the probability assigned to the loss amount $z = \zeta_\alpha(x)$ by the α-tail distribution in Definition 3, namely*

$$\lambda_\alpha(x) \;=\; [\Psi(x, \zeta_\alpha(x)) - \alpha]/[1 - \alpha] \;\in\; [0, 1]. \tag{20}$$

If $\Psi(x, \zeta_\alpha(x)) < 1$, so there is a chance of a loss greater than $\zeta_\alpha(x)$, then

$$\phi_\alpha(x) \;=\; \lambda_\alpha(x)\zeta_\alpha(x) + [1 - \lambda_\alpha(x)]\,\phi_\alpha^+(x) \tag{21}$$

with $\lambda_\alpha(x) < 1$, whereas if $\Psi(x, \zeta_\alpha(x)) = 1$, so $\zeta_\alpha(x)$ is the highest loss that can occur (and thus $\lambda_\alpha(x) = 1$ but $\phi_\alpha^+(x)$ is ill defined), then

$$\phi_\alpha(x) \;=\; \zeta_\alpha(x). \tag{22}$$

Proof. These relations are evident from formulas (7) and (8), together with the observation that $\alpha \leq \Psi(x, \zeta_\alpha(x))$ always by Definition 1. □

Corollary 7 (CVaR over VaR)**.** *From its definition, α-CVaR dominates α-VaR: $\phi_\alpha(x) \geq \zeta_\alpha(x)$. Indeed, $\phi_\alpha(x) > \zeta_\alpha(x)$ unless there is no chance of a loss greater than $\zeta_\alpha(x)$.*

Proof. This was more or less clear from the beginning, but now it emerges explicitly from Proposition 6 and the fact, seen through (12), that $\phi_\alpha^+(x) > \zeta_\alpha(x)$. □

In representing CVaR as a certain weighted average of VaR and CVaR$^+$, formula (21) seems surprising. Neither VaR nor CVaR$^+$ behaves well as a measure of risk for general loss distributions, and yet CVaR has many advantageous properties, to be seen in what follows.

The unusual feature in the definition of CVaR that leads to its power is the way that probability atoms, if present, can be "split". Such splitting is highlighted in formulas (20) and (21) of Proposition 6. In the notation of $\alpha^+(x)$ and $\alpha^-(x)$ in (5) and the circumstances in (16), where $\alpha^-(x) < \alpha < \alpha^+(x)$, an atom at $\zeta_\alpha(x)$ having total probability $\alpha^+(x) - \alpha^-(x)$ is effectively split into two pieces with probabilities $\alpha^+(x) - \alpha$ and $\alpha - \alpha^-(x)$, respectively. In concept, only the first of these pieces is adjoined to the interval $(\zeta_\alpha(x), \infty)$, which itself has probability $1 - \alpha^+(x)$, so as to achieve a probability of $[1 - \alpha^+(x)] + [\alpha^+(x) - \alpha] = 1 - \alpha$, whereas, if the atom could not be split, we would have to choose between the intervals $[\zeta_\alpha(x), \infty)$ and $(\zeta_\alpha(x), \infty)$, neither of which actually has probability $1 - \alpha$.

The splitting of probability atoms in this manner also stabilizes the response of α-CVaR to shifts in α. This will be shown later in Proposition 13.

Our next result addresses the extreme case where discreteness of the loss distribution rules entirely, as in scenario-based optimization under uncertainty. In scenario models, finitely many elements $y \in Y$ are singled out in some way as representative "scenarios," and all the probability is concentrated in them.

Proposition 8 (CVaR for scenario models)**.** *Suppose the probability measure P is concentrated in finitely many points y_k of Y, so that for each $x \in X$ the distribution of the loss $z = f(x, y)$ is likewise concentrated in finitely many points, and $\Psi(x, \cdot)$ is a step function with jumps at those points. Fixing x, let those corresponding loss points be ordered as $z_1 < z_2 < \cdots < z_N$, with the probability of z_k being $p_k > 0$. Let k_α be the unique index such that*

$$\sum_{k=1}^{k_\alpha} p_k \;\geq\; \alpha \;>\; \sum_{k=1}^{k_\alpha - 1} p_k. \tag{23}$$

The α-VaR of the loss is given then by

$$\zeta_\alpha(x) = z_{k_\alpha}, \tag{24}$$

whereas the α-CVaR is given by

$$\phi_\alpha(x) = \frac{1}{1-\alpha}\Big[\Big(\sum_{k=1}^{k_\alpha} p_k - \alpha\Big) z_{k_\alpha} + \sum_{k=k_\alpha+1}^{N} p_k z_k\Big]. \tag{25}$$

Furthermore, in this situation

$$\lambda_\alpha(x) \;=\; \frac{1}{1-\alpha}\Big(\sum_{k=1}^{k_\alpha} p_k - \alpha\Big) \;\leq\; \frac{p_{k_\alpha}}{p_{k_\alpha} + \cdots + p_N}. \tag{26}$$

Proof. According to (23), we have

$$\Psi(x, \zeta_\alpha(x)) = \sum_{k=1}^{k_\alpha} p_k, \qquad \Psi(x, \zeta_\alpha(x)^-) = \sum_{k=1}^{k_\alpha - 1} p_k, \qquad \Psi(x, \zeta_\alpha(x)) - \Psi(x, \zeta_\alpha(x)^-) = p_{k_\alpha}.$$

The assertions then follow from (8) and Proposition 6, except for the upper bound claimed for $\lambda_\alpha(x)$. To understand that, observe that the expression for $\lambda_\alpha(x)$ in (26) decreases with respect to α, which belongs to the interval in (23). The upper bound is obtained by substituting the lower endpoint of that interval for α in this expression. □

Corollary 9 (highest losses)**.** *In the notation of Proposition 8, if the highest point z_N has probability $p_N > 1 - \alpha$, then actually $\phi_\alpha(x) = \zeta_\alpha(x) = z_N$.*

Proof. This amounts to having $k_\alpha = N$, and the result then comes from (24) and (25). □

Of course, it must be remembered in Proposition 8 and Corollary 9 that not only the loss values z_k and their probabilities p_k, but also their ordering can depend on the choice of x, and so too then the index k_α, even though our notation omits that dependence for the sake of simplicity.

The case in Corollary 9 can very well come up in multistage stochastic programming models over scenario trees, for instance. In such optimization problems, the first stage may have only a few scenarios (see e.g. [19]), and CVaR will coincide then with maximum loss at that stage. Subsequent stages usually are represented with more scenarios and thus need the full force of the expressions in Proposition 8.

3 Minimization Rule and Coherence

We work now towards the goal of showing that the α-VaR and α-CVaR of the loss z associated with a choice x can be calculated simultaneously by solving an elementary optimization problem of convex type in one dimension. For this purpose we utilize, as we did in our original paper [39] in this subject, the special function

$$F_\alpha(x,\zeta) \;=\; \zeta + \frac{1}{1-\alpha} E\Big\{[f(x,y)-\zeta]^+\Big\}, \quad \text{where } [t]^+ = \max\{0,t\}. \tag{27}$$

The following theorem confirms that the minimization formula we originally developed in [39] under special assumptions on the loss distribution, such as the exclusion of discreteness, persists when the CVaR concept is articulated for general distributions in the manner of Definition 2. In contrast, no such formula holds for CVaR$^+$ or CVaR$^-$.

Theorem 10 (fundamental minimization formula)**.** *As a function of $\zeta \in I\!R$, $F_\alpha(x,\zeta)$ is finite and convex (hence continuous), with*

$$\phi_\alpha(x) \;=\; \min_\zeta F_\alpha(x,\zeta) \tag{28}$$

and moreover

$$\begin{aligned} \zeta_\alpha(x) \;&=\; \text{lower endpoint of } \; \operatorname{argmin}_\zeta F_\alpha(x,\zeta), \\ \zeta_\alpha^+(x) \;&=\; \text{upper endpoint of } \; \operatorname{argmin}_\zeta F_\alpha(x,\zeta), \end{aligned} \tag{29}$$

where the argmin refers to the set of ζ for which the minimum is attained and in this case has to be a nonempty, closed, bounded interval (perhaps reducing to a single point). In particular, one always has

$$\zeta_\alpha(x) \in \operatorname{argmin}_\zeta F_\alpha(x,\zeta), \qquad \phi_\alpha(x) = F_\alpha(x,\zeta_\alpha(x)). \tag{30}$$

Proof. The finiteness of $F_\alpha(x,\cdot)$ is a consequence of our assumption that the $E\{|f(x,y)|\} < \infty$ for each $x \in X$. Its convexity follows at once from the convexity of $[f(x,y)-\zeta]^+$ with respect to ζ. As a finite convex function, $F_\alpha(x,\cdot)$ has finite right and left derivatives at any ζ (see [38, Theorems 23.1, 24.1]). Our approach to proving the rest of the assertions in the theorem will rely on first establishing for these one-sided derivatives the formulas

$$\frac{\partial^+ F_\alpha}{\partial \zeta}(x,\zeta) = \frac{\Psi(x,\zeta_\alpha(x)) - \alpha}{1-\alpha}, \qquad \frac{\partial^- F_\alpha}{\partial \zeta}(x,\zeta) = \frac{\Psi(x,\zeta_\alpha(x)^-) - \alpha}{1-\alpha}. \tag{31}$$

We start by observing that

$$\frac{F_\alpha(x,\zeta') - F_\alpha(x,\zeta)}{\zeta' - \zeta} = 1 + \frac{1}{1-\alpha} E\Big\{\frac{[f(x,y)-\zeta']^+ - [f(x,y)-\zeta]^+}{\zeta'-\zeta}\Big\}. \tag{32}$$

When $\zeta' > \zeta$ we have

$$\frac{[f(x,y)-\zeta']^+ - [f(x,y)-\zeta]^+}{\zeta'-\zeta} \begin{cases} = -1 & \text{if } f(x,y) \geq \zeta', \\ = \ 0 & \text{if } f(x,y) \leq \zeta, \\ \in (-1,0) & \text{if } \zeta < f(x,y) < \zeta'. \end{cases}$$

Since $P\{y \,|\, f(x,y) > \zeta'\} = 1 - \Psi(x,\zeta')$ and $P\{y \,|\, \zeta < f(x,y) \leq \zeta'\} = \Psi(x,\zeta') - \Psi(x,\zeta)$, this yields the existence of a value $\rho(\zeta,\zeta') \in [0,1]$ for which

$$E\Big\{\frac{[f(x,y)-\zeta']^+ - [f(x,y)-\zeta]^+}{\zeta'-\zeta}\Big\} = -[1-\Psi(x,\zeta')] - \rho(\zeta,\zeta')[\Psi(x,\zeta') - \Psi(x,\zeta)].$$

Since furthermore $\Psi(x,\zeta') \searrow \Psi(x,\zeta)$ as $\zeta' \searrow \zeta$ (i.e., as $\zeta' \to \zeta$ with $\zeta' > \zeta$), it follows that

$$\lim_{\zeta' \searrow \zeta} E\Big\{\frac{[f(x,y)-\zeta']^+ - [f(x,y)-\zeta]^+}{\zeta'-\zeta}\Big\} = -[1-\Psi(x,\zeta)].$$

Applying this in (32), we obtain

$$\lim_{\zeta' \searrow \zeta} \frac{F_\alpha(x,\zeta') - F_\alpha(x,\zeta)}{\zeta'-\zeta} = 1 + \frac{1}{1-\alpha}[\Psi(x,\zeta) - 1] = \frac{\Psi(x,\zeta) - \alpha}{1-\alpha},$$

thereby verifying the first formula in (31). For the second formula in (31), we argue similarly that when $\zeta' < \zeta$ we have

$$\frac{[f(x,y)-\zeta']^+ - [f(x,y)-\zeta]^+}{\zeta'-\zeta} \begin{cases} = -1 & \text{if } f(x,y) \geq \zeta, \\ = \ 0 & \text{if } f(x,y) \leq \zeta', \\ \in (-1,0) & \text{if } \zeta' < f(x,y) < \zeta, \end{cases}$$

where $P\{y \,|\, f(x,y) \geq \zeta\} = 1 - \Psi(x,\zeta^-)$ and $P\{y \,|\, \zeta' < f(x,y) < \zeta\} = \Psi(x,\zeta^-) - \Psi(x,\zeta')$. Since $\Psi(x,\zeta') \nearrow \Psi(x,\zeta^-)$ as $\zeta' \nearrow \zeta$ (i.e., as $\zeta' \to \zeta$ with $\zeta' < \zeta$), we obtain

$$\lim_{\zeta' \nearrow \zeta} E\Big\{\frac{[f(x,y)-\zeta']^+ - [f(x,y)-\zeta]^+}{\zeta'-\zeta}\Big\} = -[1-\Psi(x,\zeta^-)],$$

and then in (32)

$$\lim_{\zeta' \nearrow \zeta} \frac{F_\alpha(x,\zeta') - F_\alpha(x,\zeta)}{\zeta' - \zeta} = 1 + \frac{1}{1-\alpha}[\Psi(x,\zeta^-) - 1] = \frac{\Psi(x,\zeta^-) - \alpha}{1-\alpha}.$$

That gives the second formula in (31).

Because of convexity, the one-sided derivatives in (31) are nondecreasing with respect to ζ, with the formulas assuring that

$$\lim_{\zeta\to\infty} \frac{\partial^+ F_\alpha}{\partial\zeta}(x,\zeta) = \lim_{\zeta\to\infty} \frac{\partial^- F_\alpha}{\partial\zeta}(x,\zeta) = 1$$

and on the other hand

$$\lim_{\zeta\to-\infty} \frac{\partial^+ F_\alpha}{\partial\zeta}(x,\zeta) = \lim_{\zeta\to-\infty} \frac{\partial^- F_\alpha}{\partial\zeta}(x,\zeta) = -\frac{\alpha}{1-\alpha}.$$

On the basis of these limits, we know that the level sets $\{\zeta \,|\, F_\alpha(x,\zeta) \le c\}$ are bounded (for any choice of $c \in I\!R$) and therefore that the minimum in (28) is attained, with the argmin set being a closed, bounded interval. The values of ζ in that set are characterized as the ones such that

$$\frac{\partial^- F_\alpha}{\partial\zeta}(x,\zeta) \;\le\; 0 \;\le\; \frac{\partial^+ F_\alpha}{\partial\zeta}(x,\zeta).$$

According to the formulas in (31), they are the values of ζ satisfying $\Psi(x,\zeta^-) \le \alpha \le \Psi(x,\zeta)$. The lowest such ζ is $\zeta_\alpha(x)$ by Definition 1, while the highest is $\zeta_\alpha^+(x)$ by Definition 2.

Thus, (29) and the first claim in (30) are correct. The truth of the second claim in (30) is immediate then from (28). □

Note: Very recently, and independently of our work, Acerbi and Tasche in [2] have likewise confirmed that our formula in [39] persists for CVaR in general. Their argument omits the details above, relying instead on observations about functions similar to our F_α that can be gleaned from exercises in classical probability texts.

Theorem 10 turns a powerful spotlight on the difference between CVaR and VaR, revealing the fundamental reason why CVaR is much better behaved than VaR when dependence on a choice of $x \in X$ must be handled. The reason is the fact, well known in optimization theory, that the optimal value in a problem of minimization, in this case $\phi_\alpha(x)$, is much more agreeable as a function of parameters than is the optimal solution set, which is here the argmin interval having $\zeta_\alpha(x)$ as its lower endpoint.

The special circumstances in Proposition 8 can be appreciated from the perspective of the minimization formula in Theorem 10 as well. The function $F_\alpha(x,\zeta)$ is in this case piecewise linear with derivative breakpoints at the loss values z_k. The argmin has to consist either of a single derivative breakpoint z_{k_α} or an interval $[z_{k_\alpha}, z_{k_{\alpha+1}}]$ between successive derivative breakpoints.

For the next result, we recall that a function $h(x)$ is *sublinear* if $h(x + x') \leq h(x) + h(x')$ and $h(\lambda x) = \lambda h(x)$ for $\lambda > 0$. The second of these two properties, called positive homogeneity, implies in particular that $h(0) = 0$. Sublinearity is equivalent to the combination of convexity with positive homogeneity; see [38]. Linearity is a special case of sublinearity.

Corollary 11 (convexity of CVaR). *If $f(x, y)$ is convex with respect to x, then $\phi_\alpha(x)$ is convex with respect to x as well. Indeed, in this case $F_\alpha(x, \zeta)$ is jointly convex in (x, ζ).*

Likewise, if $f(x, y)$ is sublinear with respect to x, then $\phi_\alpha(x)$ is sublinear with respect to x. Then too, $F_\alpha(x, \zeta)$ is jointly sublinear in (x, ζ).

Proof. The joint convexity of $F_\alpha(x, \zeta)$ in (x, ζ). is an elementary consequence of the definition of this function in (27) and the convexity of the function $(x, \zeta) \mapsto [f(x, y) - \zeta]^+$ when $f(x, y)$ is convex in x. The convexity of $\phi_\alpha(x)$ in x follows immediately then from the minimization formula (28). (In convex analysis, when a convex function of two vector variables is minimized with respect to one of them, the residual is a convex function of the other; see [38].)

The argument is for sublinearity is entirely parallel to the argument just given. Only the additional feature of positive homogeneity needs attention, according to the remark about sublinearity above. □

A case especially worth noting where the sublinearity in Corollary 11 is present is the one where $f(x, y)$ is actually linear with respect to x, i.e., of the form

$$f(x, y) = x_1 f_1(y) + \cdots + x_n f_n(y). \tag{33}$$

This case is common to numerous applications.

The observation that the minimization formula in Theorem 10 yields the convexity in Corollary 11 was made in our original paper [39]. We did not mention sublinearity there, but Pflug, in his follow-up article [32], noted that it too was a consequence of our formula.

Very close to Corollary 11 is an important fact about the coherence of CVaR as a risk measure, in the sense introduced by Artzner, Delbaen, Eber and Heath [6]. In the framework of those authors, a risk measure is a functional on a linear space of random variables. If we denote such random variables generically by z, thinking of them as losses, the axioms in [6] for *coherence* of a risk measure ρ amount to the requirement that ρ be sublinear,

$$\rho(z+z') \leq \rho(z) + \rho(z'), \qquad \rho(\lambda z) = \lambda \rho(z) \text{ for } \lambda \geq 0, \tag{34}$$

and in addition satisfy

$$\rho(z) = c \quad \text{when} \quad z \equiv c \text{ (constant) }, \tag{35}$$

along with

$$\rho(z) \leq \rho(z') \quad \text{when} \quad z \leq z', \tag{36}$$

where the inequality $z \leq z'$ refers to first-order stochastic dominance. (In [6], a stronger-seeming property than (35) is required, that $\rho(z+z') = c + \rho(z')$ when $z \equiv c$, but that follows from (35) and the subadditivity rule in (34).) Here our framework is different, due to the way we are modeling a loss as the joint outcome of a decision x and an underlying random vector y, but coherence can nonetheless be captured by viewing it (equivalently) as an assertion about the special case in (33).

Corollary 12 (coherence of CVaR)**.** *On the basis of Definition 3, α-CVaR is a coherent risk measure: when $f(x,y)$ is linear with respect to x, not only is $\phi_\alpha(x)$ sublinear with respect to x, but furthermore it satisfies*

$$\phi_\alpha(x) = c \quad \textit{when} \quad f(x,y) \equiv c \tag{37}$$

(thus accurately reflecting a lack of risk), and it obeys the monotonicity rule that

$$\phi_\alpha(x) \leq \phi_\alpha(x') \quad \textit{when} \quad f(x,y) \leq f(x',y). \tag{38}$$

Proof. In terms of $z = f(x,y)$ and $z' = f(x',y)$ in the context of the linearity in (33), these properties come out as the ones in (34), (35) and (36). The sublinearity of ϕ_α in the case of (33) has already been noted as ensured by Corollary 11. Like that, the additional properties (37) and (38) too can be seen as simple consequences of the fundamental minimization formula for ϕ_α in Theorem 10. □

Of course, the relations on the right sides of (37) and (38) should technically be interpreted as ones between random variables (with respect to y), rather than pointwise relations between functions of y. According to (38), for instance, a decision x that leads to an outcome at least as good as another decision x', no matter what happens, is deemed no riskier than x'.

Pflug, in [32], demonstrated that if a measure of risk were introduced in the framework of Artzner, Delbaen, Eber and Heath in [6] by the general expression derivable from the right side of our minimization formula, namely,

$$\rho(z) \;=\; \min_{\zeta\in\mathbb{R}} \Big\{ \zeta + \frac{1}{1-\alpha} E\Big\{[z-\zeta]^+\Big\}\Big\}, \tag{39}$$

it would be a coherent measure of risk. This conclusion tightly parallels Corollary 12, but here we are asserting that coherence holds for α-CVaR as the quantity introduced in Definition 3, not just for the functional defined by (39). For that assertion, the arguments behind Theorem 10, and with them the subtleties of α-CVaR as an "adjusted" conditional expectation that splits probability atoms, have a major role. The coherence of α-CVaR is a formidable advantage not shared by any other widely applicable measure of risk yet proposed.

Besides the properties already mentioned, Pflug uncovered others for the functional ρ in (39) that would likewise transfer to $\phi_\alpha(x)$. For this, we refer to his paper [32].

We close this section by pointing out still another feature of CVaR that distinguishes it from other common measures of risk for general loss distributions.

Proposition 13 (stability of CVaR)**.** *The value $\phi_\alpha(x)$ behaves continuously with respect to the choice of $\alpha \in (0,1)$ and even has left and right derivatives, given by*

$$\frac{\partial^-}{\partial\alpha}\phi_\alpha(x) = \frac{1}{(1-\alpha)^2}E\{[f(x,y)-\zeta_\alpha(x)]^+\}, \qquad \frac{\partial^+}{\partial\alpha}\phi_\alpha(x) = \frac{1}{(1-\alpha)^2}E\{[f(x,y)-\zeta_\alpha^+(x)]^+\}.$$

Proof. Fixing x, consider for each $\zeta \in \mathbb{R}$ the function of $\gamma \in \mathbb{R}$ defined by

$$\theta_\zeta(\gamma) = \zeta + \gamma E\Big\{[f(x,y)-\zeta]^+\Big\}, \tag{40}$$

and let

$$\theta(\gamma) = \min\nolimits_{\zeta\in\mathbb{R}} \theta_\zeta(\gamma). \tag{41}$$

In this way, we have through Theorem 10 that

$$\phi_\alpha(x) = \theta(\gamma) \quad \text{for } \gamma = 1/(1-\alpha), \tag{42}$$

with the minimum in (41) being attained when ζ belongs to the interval $[\zeta_\alpha(x), \zeta_\alpha^+(x)]$.

According to (41), θ is the pointwise minimum of the collection of functions θ_ζ. Those functions are affine, hence θ is concave. A finite, concave function on $\mathbb{R}$ is necessarily continuous and has left and right derivatives at every point. Under the pointwise minimization, the right derivative is the lowest of the slopes of the affine functions θ_ζ for which the minimum is attained, whereas the left derivative is the highest of such slopes. The slope of θ_ζ is given by the expectation in (40), which decreases as ζ increases. At $\gamma = 1/(1-\alpha)$, we therefore get the highest slope by taking $\zeta = \zeta_\alpha(x)$ and the lowest by taking $\zeta = \zeta_\alpha^+(x)$. Hence, at $\gamma = 1/(1-\alpha)$, the left and right derivatives of θ are $E\{[f(x,y) - \zeta_\alpha(x)]^+\}$ and $E\{[f(x,y) - \zeta_\alpha^+(x)]^+\}$, respectively.

The result now follows through (42) by considering the function $\alpha \mapsto \phi_\alpha(x)$ as the composition of θ with $\alpha \mapsto 1/(1-\alpha)$ and invoking the chain rule. □

4 CVaR in Optimization

In problems of optimization under uncertainty, CVaR can enter into the objective or the constraints, or both. A big advantage of CVaR over VaR in that context is the preservation of convexity, seen in Corollary 11. In numerical applications, the joint convexity of $F_\alpha(x,\zeta)$ with respect to both x and ζ in Corollary 10 is even more valuable than the convexity of $\phi_\alpha(x)$ in x. That is because the minimization of $\phi_\alpha(x)$ over $x \in X$, which can be adopted as a basic prototype in the management of risk when measured by α-CVaR, can be transformed into a much more tractable problem of minimizing $F_\alpha(x,\zeta)$ in both x and ζ.

Theorem 14 (optimization shortcut)**.** *Minimizing $\phi_\alpha(x)$ with respect to $x \in X$ is equivalent to minimizing $F_\alpha(x,\zeta)$ over all $(x,\zeta) \in X \times I\!R$, in the sense that*

$$\min_{x\in X} \phi_\alpha(x) \;=\; \min_{(x,\zeta)\in X\times \boldsymbol{R}} F_\alpha(x,\zeta), \tag{43}$$

where moreover

$$(x^*,\zeta^*) \in \operatorname*{argmin}_{(x,\zeta)\in X\times \boldsymbol{R}} F_\alpha(x,\zeta) \quad\Longleftrightarrow\quad x^* \in \operatorname*{argmin}_{x\in X} \phi_\alpha(x),\;\; \zeta^* \in \operatorname*{argmin}_{\zeta\in \boldsymbol{R}} F_\alpha(x^*,\zeta). \tag{44}$$

Proof. This rests on the principle in optimization that minimization with respect to (x,ζ) can be carried out by minimizing with respect to ζ for each x and then minimizing the residual with respect to x. In the situation at hand, we invoke Theorem 10 and in particular, in order to get the equivalence in (44), the fact there that the minimum of $F_\alpha(x,\zeta)$ in ζ (for fixed x) is always attained. □

Corollary 15 (VaR and CVaR calculation as a by-product)**.** *If (x^*,ζ^*) minimizes F_α over $X\times I\!R$, then not only does x^* minimize ϕ_α over X, but also*

$$\phi_\alpha(x^*) = F_\alpha(x^*,\zeta^*), \qquad\qquad \zeta_\alpha(x^*) \le \zeta^* \le \zeta_\alpha^+(x^*), \tag{45}$$

where actually $\zeta_\alpha(x^*) = \zeta^*$ if $\operatorname{argmin}_\zeta F_\alpha(x^*,\zeta)$ *reduces to a single point.*

The fact that the minimization of CVaR does *not* have to proceed numerically through repeated calculations of $\phi_\alpha(x)$ for various decisions x, may at first seem really surprising. It is a powerful attraction to working with CVaR, all the more so when compared with attempts to minimize VaR, which can be quite ill behaved and offers no such shortcut.

In the circumstance mentioned at the end of Corollary 15 where $\operatorname{argmin}_\zeta F_\alpha(x^*, \zeta)$ does not consist of just a single point, is possible to have $\zeta_\alpha(x^*) < \zeta^*$ in (45). Then the joint minimization in Theorem 14, in producing (x^*, ζ^*), although it yields the α-CVaR associated with x^*, does not immediately yield the α-VaR associated with x^*. That could well happen, for instance, in the scenario model of Proposition 8. But then, as noted earlier, $\operatorname{argmin}_\zeta F_\alpha(x^*, \zeta)$ is the interval between two consecutive points z_k in the discrete distribution of losses. In that case, therefore, $\zeta_\alpha(x^*)$ can nonetheless easily be obtained from the joint minimization: it is simply the highest $z_k \leq \zeta^*$.

Linear programming techniques can readily be utilized for the double minimization in Theorem 14 in the linear case in (33), as we have already illustrated in the more restricted setting adopted in [39]. This can be done similar to other linear programming approaches used in portfolio optimization with mean absolute deviation [27], maximum deviation [45], and mean regret [15]. Here, the significance of Theorem 14 and Corollary 15 lies in underscoring that the previous restrictions can be dropped.

The minimization of $\phi_\alpha(x)$ with respect to $x \in X$ is not the only way that CVaR can be utilized in risk management. It can also be brought in to "shape" the risk in an optimization model. For that purpose, several probability thresholds can be handled.

Theorem 16 (risk-shaping with CVaR)**.** *For any selection of probability thresholds α_i and loss tolerances ω_i, $i = 1, \ldots, l$, the problem*

$$\text{minimize } g(x) \text{ over } x \in X \text{ satisfying } \phi_{\alpha_i}(x) \leq \omega_i \text{ for } i = 1, \ldots, l, \tag{46}$$

where g is any objective function chosen on X, is equivalent to the problem

$$\begin{aligned} &\text{minimize } g(x) \text{ over } (x, \zeta_1, \ldots, \zeta_l) \in X \times I\!R \times \cdots \times I\!R \\ &\qquad \text{satisfying } F_{\alpha_i}(x, \zeta_i) \leq \omega_i \text{ for } i = 1, \ldots, l. \end{aligned} \tag{47}$$

Indeed, $(x^*, \zeta_1^*, \cdots, \zeta_l^*)$ solves the second problem if and only if x^* solves the first problem and the inequality $F_{\alpha_i}(x^*, \zeta_i^*) \leq \omega_i$ holds for $i = 1, \ldots, l$.

Moreover one then has $\phi_{\alpha_i}(x^*) \leq \omega_i$ for every i, and actually $\phi_{\alpha_i}(x^*) = \omega_i$ for each i such that $F_{\alpha_i}(x^*, \zeta_i^*) = \omega_i$ (i.e., such that the corresponding CVaR constraint is active).

Proof. Again, this relies on the minimization formula (28) in Theorem 10 and the assured attainment of the minimum there. The argument is very much like that for Theorem 14. Because

$$\phi_{\alpha_i}(x) \;=\; \min_{\zeta_i \in \boldsymbol{R}} F_{\alpha_i}(x, \zeta_i), \tag{48}$$

we have $\phi_{\alpha_i}(x) \leq \omega_i$ if and only if there exists ζ_i such that $F_{\alpha_i}(x, \zeta_i) \leq \omega_i$. □

When X and g are convex and $f(x, y)$ is convex in x, we know from Corollary 11 that the optimization problems in Theorems 14 and 16 are ones of *convex programming* and thus especially favorable for computation. In comparison, analogous problems in terms of VaR instead of CVaR could be highly unfavorable. Of course, a combination of the models in Theorems 14 and 16 could likewise be handled in such a manner, by taking $g(x) = \phi_{\alpha_0}(x)$ for some α_0.

Linear programming techniques can be used to compute answers in this setting as well. That is most evident when Y is a discrete probability space with elements y_k, $k = 1, \ldots, N$ having probabilities p_k, $k = 1, \ldots, N$. Then from (27) we have

$$F_{\alpha_i}(x, \zeta_i) = \zeta_i + \frac{1}{(1 - \alpha_i)} \sum_{k=1}^{N} p_k [f(x, y_k) - \zeta_i]^+. \tag{49}$$

The constraint $F_{\alpha_i}(x, \zeta_i) \leq \omega_i$ in Theorem 16 can be handled by introducing additional variables η_{ik} subject to the conditions

$$\eta_{ik} \geq 0, \qquad f(x, y_k) - \zeta_i - \eta_{ik} \leq 0, \tag{50}$$

and requiring that

$$\zeta_i + \frac{1}{(1 - \alpha_i)} \sum_{k=1}^{N} p_k \eta_k \;\leq\; \omega_i. \tag{51}$$

The minimization in the expanded problem (47) is converted then into the minimization of $g(x)$ over $x \in X$, the ζ_i's and all the new η_{ik}'s, with the constraints $F_{\alpha_i}(x, \zeta_i) \leq \omega_i$ being replaced by (50) and (51). When f is linear in x as in (33), these reconstituted constraints are linear.

This conversion is entirely parallel to the one we introduced in [39] for the expanded optimization problem with respect to x and ζ that appears in Theorem 14.

5 An Example of Portfolio Replication with CVaR Constraints

Putting together a portfolio in order to track a given financial index is a common and important undertaking. It fits in the framework of "portfolio replication" as a form of approximation, but of course the approximation criterion that is adopted must be one that focuses on risks associated with inaccuracies in the tracking. We present an example that demonstrates how

CVaR constraints can be used efficiently to control such risks. For other work on portfolio replication, see for instance [5, 9, 11, 13, 14, 16, 28, 42, 44].

Suppose we want to replicate an instrument I (e.g. the S&P100 index) using certain other instruments S_j, $j = 1, \ldots, n$. Denote by I_t the price of instrument I at time t, for $t = 1, \ldots, T$, and denote by p_{tj} the price of instrument S_j at time t. Let ν be amount of money to be on hand at the final time T. We denote by $\theta = \frac{\nu}{I_T}$ the number of units of the instrument I at time T. Let x_j, for $j = 1, \ldots, n$, be the number of units of instrument S_j in the proposed replicating portfolio. The value of that portfolio at time t is then $\sum_{j=1}^n p_{tj} x_j$. The absolute value of the relative deviation of the portfolio value from the target value θI_t is $|(\theta I_t - \sum_{j=1}^n p_{tj} x_j)/\theta I_t|$.

To put this into our earlier framework, we think of the price vectors $p_t = (p_{t1}, \ldots, p_{tn})$ for $t = 1, \ldots, T$ as observations of a random element $y \in \mathbb{R}^n$, but now write p instead of y (and have indexing $t = 1, \ldots, T$ instead of $k = 1, \ldots, N$). These observed vectors p_t give a finite distribution of p in which $p = p_t$ has probability $1/T$. We take the loss associated with a decision x to be the relative shortfall

$$f(x, p) = \Big(\theta I_t - \sum_{j=1}^n p_{tj} x_j\Big) \Big/ \theta I_t, \tag{52}$$

and introduce, as the expression to be minimized, the expectation of $|f(x, p)|$, i.e., the average of the absolute values of the relative deviations $|f(x, p_t)|$ for $t = 1, \ldots, T$. In addition, we impose a constraint on the CVaR amount $\phi_\alpha(x)$ associated with the loss $f(x, p)$ in order to control large deviations of the portfolio value *below* the target value.

In the pattern of the expanded problem (47) in Theorem 16, but with only one CVaR constraint, our portfolio replication problem comes out then as follows:

$$\text{mininimize} \quad g(x) = \frac{1}{T} \sum_{t=1}^T \Big| \Big(\theta I_t - \sum_{j=1}^n p_{tj} x_j\Big) \Big/ \theta I_t \Big| \tag{53}$$

subject to the constraints

$$\sum_{j=1}^n p_{jT} x_j = \nu, \tag{54}$$

$$0 \le x_j \le \gamma_j, \qquad j = 1, \ldots, n, \tag{55}$$

(which realize in this setting the constraint $x \in X$ in the general discussion earlier) and

$$\zeta + \frac{1}{(1-\alpha)T} \sum_{t=1}^T \Big[\Big[\Big(\theta I_t - \sum_{j=1}^n p_{tj} x_j\Big) \Big/ \theta I_t \Big] - \zeta \Big]^+ \le \omega. \tag{56}$$

The minimization takes place with respect to both $x = (x_1, \ldots, x_n)$ and the variable ζ. The expression on the left side of (56) is $F_\alpha(x, \zeta)$; thus, (56) corresponds to requiring $\phi_\alpha(x) \le \omega$.

For any choice of α and ω, this problem can be solved by conversion to linear programming, more or less in the manner already explained above. The performance function g is handled by introducing still more variables $\eta_{t0} \geq 0$ constrained by

$$\left(\theta I_t - \sum_{j=1}^n p_{tj} x_j\right) \Big/ \theta I_t - \eta_{t0} \leq 0,$$
$$-\left(\theta I_t - \sum_{j=1}^n p_{tj} x_j\right) \Big/ \theta I_t + \eta_{t0} \leq 0,$$

and minimizing the expression $(1/T) \sum_{t=1}^T \eta_{t0}$.

Several important issues in the modeling, such as transaction costs and how to select the stocks to be included in the replicating portfolio, are beyond the scope of this paper. However, that does not undermine the basic idea of the CVaR approach, which we proceed to lay out.

Calculations for this example were conducted using LP solver of CPLEX package.

In our numerical experiments, we aimed at replicating the S&P100 index using 30 of the stocks that belong to that index (namely, the ones with ticker symbols: GD, UIS, NSM, ORCL, CSCO, HET, BS, TXN, HM, INTC, RAL, NT, MER, KM, BHI, CEN, HAL, BDK, HWP, LTD, BAC, AVP, AXP, AA, BA, AGC, BAX, AIG, AN, AEP). These stocks were the instruments S_j. The experiments were conducted in two stages:

Stage 1 (in-sample calculations): the problem (53)–(56) was solved using in-sample historical data on stock prices.

Stage 2 (out-of-sample calculations): replicating properties of the portfolio were verified by using the out-of-sample historical data just after the in-sample replicating period.

For the in-sample calculations, we used the closing prices for 600 days (from 10.21.1996 to 03.08.1999). For the out-of-sample calculations we considered 100 days (from 03.09.1999 to 07.28.1999). The confidence level in CVaR constraint (56) was taken to be $\alpha = 0.9$, so that the CVaR constraint would control the largest 10% of relative deviations (underperformance of the portfolio compared to the index).

We solved the replication problem (53)–(56) for several values of the risk-tolerance level ω in the CVaR constraint (ω was varied from 0.02 to 0.001). To verify out-of-sample goodness of fit we calculated the values of performance function (53) and the CVaR in (56) for the out-of-sample dataset. The results of the calculations are presented in Table 1 and Figures 6–13. The analysis of these results follows.

In-sample calculations. Imposing the CVaR constraint ought to lead to a deterioration in the value of the in-sample objective function (the average absolute value of the relative deviation). Indeed, decreasing the value of ω causes an increase in the value of objective function in in-sample

confidence level ω	in-sample (600 days) objective function, in %	out-of-sample (100 days) objective function, in %	out-of-sample CVaR, in %
0.02	0.71778	2.73131	4.88654
0.01	0.82502	1.64654	3.88691
0.005	1.11391	0.85858	2.62559
0.003	1.28004	0.78896	2.16996
0.001	1.48124	0.80078	1.88564

Table 1: Calculation results for various risk levels ω in the CVaR constraint.

region (Column 2 of Table 1). This is seen in Figure 6 (continuous thick line) and is an evident consequence of the fact that decreasing the value of ω diminishes the feasible set. At the risk-tolerance level $\omega = 0.02$, the constraint on CVaR in (56) is inactive; at $\omega \leq 0.01$ that constraint is active. The dynamics of relative absolute deviation (in-sample) for an instance when the CVaR constraint is active (at $\omega = 0.005$) and an instance when it is inactive (at $\omega = 0.02$) are shown in Fig. 7. This figure reveals that the CVaR constraint has reduced underperformance of the portfolio versus the index in the in-sample region: the dotted curve corresponding to the active CVaR constraint is lower than solid curve corresponding to the inactive CVaR constraint. The dynamics of portfolio and index values for cases when the CVaR constraint is active (at $\omega = 0.005$) and inactive (at $\omega = 0.02$) are shown in Fig. 8 and Fig. 9, respectively. These figures demonstrate that the portfolio fits the index quite well for both active and inactive CVaR constraints.

At $\omega = 0.005$ and the optimal portfolio point x^*, we got $\zeta^* = 0.001538627671$ and the CVaR value of the left side in (56) equal to 0.005. In this case the probability of the VaR point itself is 14/600, which means that 14 time points have the same deviation 0.001538627671. To verify our optimization result at the optimal portfolio x^*, we manually calculated:

$$\text{VaR}{=}0.001538627671, \quad \text{CVaR} = 0.005$$
$$\text{CVaR}^- = 0.004592779726, \quad \text{CVaR}^+{=}0.005384596925.$$

We found that ζ^* =VaR and the left side of the inequality (56) is CVaR= $\omega = 0.005$. In the case under consideration, the losses of 54 scenarios exceed VaR. The probability of exceeding the VaR, i.e, the probability of the interval $(\zeta_\alpha(x^*), \infty)$, was

$$1 - \Psi(x^*, \zeta_\alpha(x^*)) = 54/600 < 1 - \alpha,$$

whereas

$$\lambda_\alpha(x^*) = [\Psi(x^*, \zeta_\alpha(x^*)) - \alpha]/[1-\alpha] = [546/600 - 0.9]/[1-0.9] = 0.1.$$

In accordance with formula (20), we got

$$\text{CVaR} = \lambda_\alpha(x^*)\ \text{VaR} + (1-\lambda_\alpha(x))\ \text{CVaR}^+,$$

$$(\text{CVaR} = 0.1 * 0.001538627671 + 0.9 * 0.005384596925 = 0.005).$$

Also, because, $\Psi(x, \zeta_\alpha(x^*)) > \alpha$

$$\text{CVaR}^- < \text{CVaR} < \text{CVaR}^+.$$

In several runs we observed that the optimal ζ^* may overestimate the VaR because of the nonuniqueness of the optimal solution, i.e., instances of a nontrivial argmin interval in (29). (In our case of a discrete distribution, ζ^* can equal the value of the first loss possibility beyond the VaR.) Also, when the CVaR constraint (56) is not active, the optimal ζ^* may be quite far from the VaR and the value on the left of (56) may likewise be quite far from the CVaR.

Out-of-sample calculations. Table 1 shows that the CVaR calculated in the out-of-sample region decreases when value of ω decreases (Column 4). This means that we improved both in-sample and the out-of-sample "large deviations" by imposing the constraint (56). The index and optimal portfolio values in the out-of-sample region when the CVaR constraint is active (at $\omega = 0.005$) are shown in Fig. 10, and when it is not active (at $\omega = 0.02$), are shown in Fig. 11. The relative absolute deviation in the out-of-sample region for the active cases (at $\omega = 0.005$) and the inactive (at $\omega = 0.02$) are displayed in Fig. 12.

An improvement of the CVaR for both in-sample and out-of-sample regions was also observed for other data intervals, for instance, for 600 in-sample days from 11.28.1997 to 04.13.2000 and 100 out-of-sample days from 04.14.2000 to 09.06.2000.

Column 3 of Table 1 demonstrates that imposing the in-sample CVaR constraint brings about an improvement of the objective function in the out-of-sample data region (in contrast to the in-sample increase of the objective function); see Figure 1. However, a decrease in the objective function in the out-of-sample region was not observed for several other datasets.

Acknowledgments. We are grateful to Alexander Golodnikov from the Glushkov Institute in Kiev, and to Grigori Zrazhevski from Kiev University for conducting numerical experiments for the portfolio replication problem.

Figure 6: In-sample objective function, out-of-sample objective function, out-of-sample CVaR for various risk levels ω in CVaR constraint.

Figure 7: Relative discrepancy in in-sample region, CVaR constraint is active (ω=0.005) and inactive (ω=0.02)

Figure 8: Index and optimal portfolio values in in-sample region, CVaR constraint is active (ω=0.005).

Figure 9: Index and optimal portfolio values in in-sample region, CVaR constraint is inactive (ω=0.02)

Figure 10: Index and optimal portfolio values in out-of-sample region, CVaR constraint is active (ω =0.005).

Figure 11: Index and optimal portfolio values in out-of-sample region, CVaR constraint is inactive (ω=0.02)

Figure 12: Relative discrepancy in out-of-sample region, CVaR constraint is active (ω=0.005) and inactive (ω =0.02)

References

[1] C. Acerbi, C. Nordio and C. Sirtori, *Expected shortfall as a tool for financial risk management.* Working paper (2001), can be downloaded from http://www.gloriamundi.org.

[2] C. Acerbi and D. Tasche, *On the coherence of expected shortfall.* Working paper (2001), can be downloaded from http://www.gloriamundi.org.

[3] J.V. Andersen and D. Sornette, *Have Your Cake and Eat It Too: Increasing Returns While Lowering Large Risks.* Working Paper, University of Los Angeles, (1999), can be downloaded from http://www.gloriamundi.org.

[4] F. Andersson, H. Mausser, D. Rosen, S. Uryasev, Credit risk optimization with conditional value-at-risk, *Mathematical Programming*, Series B, December, 2000, relevant Research Report 99-9 can be downloaded from www.ise.ufl.edu/uryasev/pubs.html#t.

[5] C. Andrews, D. Ford and K. Mallinson, The Design of Index Funds and Alternative Methods of Replication, *The Investment Analyst* 82, (October 1986), 16-23.

[6] P. Artzner, F. Delbaen, J.-M. Eber, D. Heath, Coherent measures of risk, *Mathematical Finance*, 9 (1999), 203–228.

[7] S. Basak and A. Shapiro, *Value-at-Risk Based Management: Optimal Policies and Asset Prices.* Working Paper, Wharton School, University of Pennsylvania, (1998), can be downloaded from http://www.gloriamundi.org.

[8] F. Bassi, P. Embrechts, M. Kafetzaki, Risk management and quantile estimation, in *Practical Guide to Heavy Tails* (R. Adler, F. Feldman, M. Taqqu, eds.), Birkhäuser, Boston, (1998), 111–130.

[9] J.E. Beasley, N. Meade, T.-J. Chang, *Index Tracking,* Working Paper, Imperial College London, (1999).

[10] J.R. Birge and F. Louveaux, *Introduction to Stochastic Programming.* Springer, New York, (1997).

[11] I.R.C. Buckley and R. Korn, Optimal Index Tracking under Transaction Costs and Impulse Control, *International Journal of Theoretical and Applied Finance*, (1998), 315-330.

[12] A. Checklov, S. Uryasev, M. Zabarankin, Portfolio optimization with drawdown constraints, submitted to *Applied Mathematical Finance*, relevant Research Report 2000-5 can be downloaded from www.ise.ufl.edu/uryasev/pubs.html#t.

[13] G. Connor and H. Leland, Cash Management for Index Tracking, *Financial Analysts Journal* 51(6) (November/December 1995), 75–80.

[14] H. Dalh, A. Meeraus, and S.A. Zenios, Some Financial Optimization Models: I risk Management, in *Financial optimization*, S.A. Zenios ed., Cambridge University Press, (1993), 3-36.

[15] R.S. Dembo and A.J. King, Tracking Models and the Optimal Regret Distribution in Asset Allocation. *Applied Stochastic Models and Data Analysis.* Vol. 8, (1992) 151–157.

[16] R. Dembo and D. Rosen, The Practice of Portfolio Replication. A Practical Overview of Forward and Inverse Problems, *Annals of Operations Research* 85 (1999), 267–284.

[17] P. Embrechts, A. McNeil, D. Straumann, *Correlation and dependency in risk management: properties and pitfalls*, To appear in "Risk Management: Value at Risk and Beyond." Ed. M. Dempster. Cambridge University Press, Cambridge, (2001).

[18] P. Embrechts, C. Klüppelberg, T. Mikosch, *Modelling Extremal Events for Insurance and Finance*, Springer, New York, (1997).

[19] Yu. Ermoliev and R. J-B Wets (Eds.), *Numerical Techniques for Stochastic Optimization*, Springer Series in Computational Mathematics, 10 (1988).

[20] A.A. Gaivoronski, and G. Pflug, *Value at Risk in portfolio optimization: properties and computational approach*, NTNU, Department of Industrial Economics and Technology Management, Working paper, (July 2000).

[21] C. Gourieroux, J.P. Laurent, and O. Scaillet, *Sensitivity Analysis of Values-at-Risk.*, Working paper, Universite de Louvan, (January 2000), can be downloaded from http://www.gloriamundi.org.

[22] H. Grootweld and W.G. Hallerbach, *Upgrading VaR from Diagnostic Metric to Decision Variable: A Wise Thing to Do?*, Report 2003 Erasmus Center for Financial Research, (June 2000).

[23] Ph. Jorion, *Value at Risk: A New Benchmark for Measuring Derivatives Risk.* Irwin Professional Pub. (1996).

[24] P. Kall, and S.W. Wallace, *Stochastic Programming.* Willey, Chichester, (1994).

[25] Y.S. Kan and A.I. Kibzun, *Stochastic Programming Problems with Probability and Quantile Functions*, John Wiley & Sons, (1996) 316.

[26] R. Kast, E. Luciano, and L. Peccati, *VaR and Optimization*, 2nd International Workshop on Preferences and Decisions, Trento, (July 1998).

[27] Konno, H. and H. Yamazaki, Mean Absolute Deviation Portfolio Optimization Model and Its Application to Tokyo Stock Market, *Management Science,* **37**, (1991) 519-531.

[28] H. Konno and A. Wijayanayake, *Minimal Cost Index Tracking under Nonlinear Transaction Costs and Minimal Transaction Unit Constraints,* Tokyo Institute of Technology, CRAFT Working paper 00-07, (2000).

[29] H.M. Markowitz, Portfolio Selection. *Journal of Finance.* Vol.7, 1, (1952) 77–91.

[30] H. Mausser and D. Rosen, Beyond VaR: From Measuring Risk to Managing Risk, *Algo Research Quarterly*, Vol. 1, No. 2,(1998), 5–20.

[31] H. Mausser and D. Rosen, Efficient Risk/Return Frontiers for Credit Risk, *Algo Research Quarterly*, Vol. 2, No. 4,(1999), 35–47.

[32] G. Pflug, *Some remarks on the value-at-risk and the conditional value-at-risk*, in "Probabilistic Constrained Optimization: Methodology and Applications" (S. Uryasev ed.), Kluwer Academic Publishers, 2000.

[33] G.Ch. Pflug, *Optimization of Stochastic Models: The Interface Between Simulation and Optimization*, Kluwer Academic Publishers, (1996).

[34] J. Palmquist, S. Uryasev, and P. Krokhmal, Portfolio optimization with conditional value-at-risk criterion, *Journal of Risk*, forthcoming (relevant Research Report 99-14 can be downloaded from www.ise.ufl.edu/uryasev/pal.pdf).

[35] M. Pritsker, Evaluating Value at Risk Methodologies, *Journal of Financial Services Research*, 12:2/3, (1997) 201–242.

[36] A. Prekopa, *Stochastic Programming*, Kluwer Academic Publishers, 1995.

[37] A. Puelz, *Value-at-Risk Based Portfolio Optimization.* Working paper, Southern Methodist University, (November 1999).

[38] R. T. Rockafellar, *Convex Analysis*, Princeton University Press, 1970; available since 1997 in paperback in the series *Princeton Landmarks in Mathematics and Physics*.

[39] R. T. Rockafellar, S. Uryasev, Optimization of conditional value-at-risk, *Journal of Risk* 2 (2000), 21–41.

[40] R. T. Rockafellar and R. J-B Wets, *Variational Analysis*, Grundlehren der Math. Wissenschaften 317, Springer Verlag, 1997.

[41] R. Rubinstein and A. Shapiro, *Discrete Event Systems: Sensitivity Analysis and Stochastic Optimization via the Score Function Method*, Willey, Chichester,(1993).

[42] A. Rudd, Optimal Selection of Passive Portfolios, *Financial Management*, (Spring 1980), 57-66.

[43] D. Tasche, *Risk contributions and performance measurement*, Working paper, Munich University of Technology, (July 1999).

[44] W.M. Toy and M.A. Zurack, Tracking the Euro-Pac Index, *The Journal of Portfolio Management*, 15(2) (Winter 1989), 55–58.

[45] M.R. Young, A Minimax Portfolio Selection Rule with Linear Programming Solution. *Management Science*, Vol.44, No. 5, (1998) 673–683.

[46] S.A. Zenios (Ed.), *Financial Optimization*, Cambridge Univ Press, (1993).

[47] T.W. Ziemba and M.J. Mulvey (Eds.), *Worldwide Asset and Liability Modeling* , Cambridge Press, Publications of the Newton Institute, (1998).