Optimization of conditional value-at-risk

R. Tyrrell Rockafellar

Department of Applied Mathematics, University of Washington, 408 L Guggenheim Hall, Box 352420, Seattle, Washington 98195-2420, USA

Stanislav Uryasev

Department of Industrial and Systems Engineering, University of Florida, PO Box 116595, 303 Weil Hall, Gainesville, Florida 32611-6595, USA

A new approach to optimizing or hedging a portfolio of financial instruments to reduce risk is presented and tested on applications. It focuses on minimizing conditional value-at-risk (CVaR) rather than minimizing value-at-risk (VaR), but portfolios with low CVaR necessarily have low VaR as well. CVaR, also called mean excess loss, mean shortfall, or tail VaR, is in any case considered to be a more consistent measure of risk than VaR. Central to the new approach is a technique for portfolio optimization which calculates VaR and optimizes CVaR simultaneously. This technique is suitable for use by investment companies, brokerage firms, mutual funds, and any business that evaluates risk. It can be combined with analytical or scenario-based methods to optimize portfolios with large numbers of instruments, in which case the calculations often come down to linear programming or nonsmooth programming. The methodology can also be applied to the optimization of percentiles in contexts outside of finance.

1. INTRODUCTION

This paper introduces a new approach to optimizing a portfolio so as to reduce the risk of high losses. Value-at-risk (VaR) has a role in the approach, but the emphasis is on conditional value-at-risk (CVaR), which is also known as mean excess loss, mean shortfall, or tail VaR. By definition, with respect to a specified probability level β, the β-VaR of a portfolio is the lowest amount α such that, with probability β, the loss will not exceed α, whereas the β-CVaR is the conditional expectation of losses above that amount α. Three values of β are commonly considered: 0.90, 0.95, and 0.99. The definitions ensure that the β-VaR is never more than the β-CVaR, so portfolios with low CVaR must have low VaR as well.

A description of various methodologies for the modeling of VaR can be seen, along with related resources, at URL *http://www.gloriamundi.org/*. Most approaches to calculating VaR rely on linear approximation of the portfolio risks and assume a joint normal (or lognormal) distribution of the underlying market parameters (see, e.g., Duffie and Pan 1997, Jorion 1996, Pritsker 1997, JP Morgan 1996, Simons 1996, Beder 1995, Stambaugh 1996). Also, historical or Monte Carlo simulation-based tools are used when the portfolio contains nonlinear instruments such as options (Bucay and Rosen 1999, Jorion 1996, Mauser and Rosen 1999, Pritsker 1997, JP Morgan 1996, Beder 1995, Stambaugh 1996). Discussions of optimization problems involving VaR can be found in papers by Litterman (1997a,1997b), Kast *et al*. (1998), and Lucas and Klaassen (1998).

Although VaR is a very popular measure of risk, it has undesirable mathematical characteristics such as a lack of subadditivity and convexity (see Artzner *et al*. 1997, 1999). VaR is coherent only when it is based on the standard deviation of normal distributions (for a normal distribution, VaR is proportional to the standard deviation). For example, VaR associated with a combination of two portfolios can be deemed greater than the sum of the risks of the individual portfolios. Furthermore, VaR is difficult to optimize when it is calculated from scenarios. Mauser and Rosen (1999) and McKay and Keefer (1996) showed that VaR can be ill-behaved as a function of portfolio positions and can exhibit multiple local extrema, which can be a major handicap in trying to determine an optimal mix of positions or even the VaR of a particular mix. As an alternative measure of risk, CVaR is known to have better properties than VaR (see Artzner *et al*. 1997 and Embrechts *et al*. 1999). Recently, Pflug (2000) proved that CVaR is a coherent risk measure having the following properties: transition-equivariant, positively homogeneous, convex, monotonic w.r.t. stochastic dominance of order 1, and monotonic w.r.t. monotonic dominance of order 2. A simple description of the approach for minimization of CVaR and optimization problems with CVaR constraints can be found in the review paper by Uryasev (2000). Although CVaR has not become a standard in the finance industry, CVaR is gaining in the insurance industry (see Embrechts *et al*. 1997). Bucay and Rosen (1999) used CVaR in credit risk evaluations. A case study on application of the CVaR methodology to the credit risk is described by Andersson and Uryasev (1999). Measures similar to CVaR have been used before in the stochastic programming literature, although not in a financial mathematics context. The conditional expectation constraints and integrated chance constraints described by Prekopa (1995) may serve the same purpose as CVaR.

Minimizing the CVaR of a portfolio is closely related to minimizing VaR, as already observed from the definition of these measures. The basic contribution of this paper is a practical technique of optimizing CVaR and calculating VaR at the same time. It affords a convenient way of evaluating:

- linear and nonlinear derivatives (options, futures);
- market, credit, and operational risks;
- circumstances in any corporation that is exposed to financial risks.

It can be used for such purposes by investment companies, brokerage firms, mutual funds, and elsewhere.

In the optimization of portfolios, the new approach leads to solving a stochastic optimization problem. Many numerical algorithms are available for that (see, e.g., Birge and Louveaux 1997, Ermoliev and Wets 1988, Kall and Wallace 1995, Kan and Kibzun 1996, Pflug 1996, Prekopa 1995). These algorithms are able to make use of special mathematical features in the portfolio and can readily be combined with analytical or simulation-based methods. In cases where the uncertainty is modeled by scenarios and a finite family of scenarios is selected as an approximation, the problem to be solved can even reduce to one of linear programming. For applications of programming in finance, see, for instance, Zenios (1996) and Ziemba and Mulvey (1998).

2. DESCRIPTION OF THE APPROACH

Let $f(\boldsymbol{x}, \boldsymbol{y})$ be the loss associated with the decision vector $\boldsymbol{x}$, to be chosen from a certain subset X of $\mathbb{R}^n$, and the random vector $\boldsymbol{y}$ in $\mathbb{R}^m$. The vector $\boldsymbol{x}$ can be interpreted as representing a portfolio, with X as the set of available portfolios (subject to various constraints), but other interpretations are also possible. The vector $\boldsymbol{y}$ stands for the uncertainties, e.g. market variables, that can affect the loss. Of course, the loss might be negative and thus, in effect, constitute a gain.

For each $\boldsymbol{x}$, the loss $f(\boldsymbol{x}, \boldsymbol{y})$ is a random variable having a distribution in $\mathbb{R}$ induced by that of $\boldsymbol{y}$. The underlying probability distribution of $\boldsymbol{y}$ in $\mathbb{R}^m$ will be assumed for convenience to have density, which we denote by $p(\boldsymbol{y})$. However, as will be shown later, an analytical expression $p(\boldsymbol{y})$ for the implementation of the approach is not needed. It is sufficient to have an algorithm (code) which generates random samples from $p(\boldsymbol{y})$. A two-step procedure can be used to derive analytical expression for $p(\boldsymbol{y})$ or construct a Monte Carlo simulation code for drawing samples from $p(\boldsymbol{y})$ (see, e.g., JP Morgan 1996): (1) modeling of risk factors in $\mathbb{R}^{m_1}$, with $m_1 < m$; (2) based on the characteristics of instrument i $(i = 1, \ldots, n)$, the distribution $p(\boldsymbol{y})$ can be derived or code transforming random samples of risk factors to the random samples from density $p(\boldsymbol{y})$ can be constructed.

The probability of $f(\boldsymbol{x}, \boldsymbol{y})$ not exceeding a threshold α is then given by

$$\Psi(\boldsymbol{x}, \alpha) = \int_{f(\boldsymbol{x},\boldsymbol{y}) \leqslant \alpha} p(\boldsymbol{y})\, \mathrm{d}\boldsymbol{y}. \tag{1}$$

As a function of α for fixed $\boldsymbol{x}$, Ψ is the cumulative distribution function for the loss associated with $\boldsymbol{x}$. It completely determines the behavior of this random variable and is fundamental in defining VaR and CVaR. In general, $\Psi(\boldsymbol{x}, \alpha)$ is nondecreasing with respect to α and continuous from the right, but not necessarily from the left because of the possibility of jumps. We assume, however, in what follows that the probability distributions are such that no jumps occur, or, in other words, that $\Psi(\boldsymbol{x}, \alpha)$ is everywhere continuous with respect to α. This assumption, like the previous one about density in $\boldsymbol{y}$, is made for simplicity. Without it there are mathematical complications, even in the definition of CVaR, which would need more explanation. We prefer to leave such technical issues for a subsequent paper. In some common situations, the required continuity follows from properties of loss $f(\boldsymbol{x}, \boldsymbol{y})$ and the density $p(\boldsymbol{y})$ (see Uryasev 1995).

The β-VaR and β-CVaR values for the loss random variable associated with $\boldsymbol{x}$ and any specified probability level β in $(0, 1)$ will be denoted by $\alpha_\beta(\boldsymbol{x})$ and $\phi_\beta(\boldsymbol{x})$. In our setting, they are given by

$$\alpha_\beta(\boldsymbol{x}) = \min\{\alpha \in \mathbb{R} : \Psi(\boldsymbol{x}, \alpha) \geqslant \beta\} \tag{2}$$

and

$$\phi_\beta(\boldsymbol{x}) = (1-\beta)^{-1} \int_{f(\boldsymbol{x},\boldsymbol{y}) \geqslant \alpha_\beta(\boldsymbol{x})} f(\boldsymbol{x}, \boldsymbol{y})\, p(\boldsymbol{y})\, \mathrm{d}\boldsymbol{y}. \tag{3}$$

In the first formula, $\alpha_\beta(\boldsymbol{x})$ comes out as the left endpoint of the nonempty interval consisting of the values α such that $\Psi(\boldsymbol{x}, \alpha) = \beta$. (This follows from $\Psi(\boldsymbol{x}, \alpha)$ being continuous and nondecreasing with respect to α. The interval might contain more than a single point if Ψ has "flat spots".) In the second formula, the probability that $f(\boldsymbol{x}, \boldsymbol{y}) \geqslant \alpha_\beta(\boldsymbol{x})$ is therefore equal to $1 - \beta$. Thus, $\phi_\beta(\boldsymbol{x})$ comes out as the conditional expectation of the loss associated with $\boldsymbol{x}$ relative to that loss being $\alpha_\beta(\boldsymbol{x})$ or greater.

The key to our approach is a characterization of $\phi_\beta(\boldsymbol{x})$ and $\alpha_\beta(\boldsymbol{x})$ in terms of the function F_β on $X \times \mathbb{R}$ defined as follows:

$$F_\beta(\boldsymbol{x}, \alpha) = \alpha + (1-\beta)^{-1} \int_{y \in \mathbb{R}^m} [f(\boldsymbol{x}, \boldsymbol{y}) - \alpha]^+\, p(\boldsymbol{y})\, \mathrm{d}\boldsymbol{y}, \tag{4}$$

where

$$[t]^+ = \begin{cases} t & \text{when } t > 0, \\ 0 & \text{when } t \leqslant 0. \end{cases}$$

The crucial features of F_β, under the assumptions made above, are as follows. For background on convexity, which is a key property in optimization that in particular eliminates the possibility of a local minimum being different from a global minimum, see Rockafellar (1970) or Shor (1985), for instance.

THEOREM 1 *As a function of α, $F_\beta(\boldsymbol{x}, \alpha)$ is convex and continuously differentiable. The β-CVaR of the loss associated with any $\boldsymbol{x} \in X$ can be determined from the formula*

$$\phi_\beta(\boldsymbol{x}) = \min_{\alpha \in \mathbb{R}} F_\beta(\boldsymbol{x}, \alpha). \tag{5}$$

In this formula, the set consisting of the values of α for which the minimum is attained, namely

$$A_\beta(\boldsymbol{x}) = \operatorname*{argmin}_{\alpha \in \mathbb{R}} F_\beta(\boldsymbol{x}, \alpha), \tag{6}$$

is a nonempty closed bounded interval (perhaps reducing to a single point), and the β-VaR of the loss is given by

$$\alpha_\beta(\boldsymbol{x}) = \textit{left endpoint of } A_\beta(\boldsymbol{x}). \tag{7}$$

In particular, one always has

$$\alpha_\beta(\boldsymbol{x}) \in \operatorname*{argmin}_{\alpha \in \mathbb{R}} F_\beta(\boldsymbol{x}, \alpha) \quad \textit{and} \quad \phi_\beta(\boldsymbol{x}) = F_\beta\big(\boldsymbol{x}, \alpha_\beta(\boldsymbol{x})\big). \tag{8}$$

Theorem 1 is proved in the Appendix. Note that for computational purposes one could just as well minimize $(1-\beta)F_\beta(\boldsymbol{x}, \alpha)$ as minimize $F_\beta(\boldsymbol{x}, \alpha)$. This would avoid dividing the integral by $1-\beta$ and might be better numerically when $1-\beta$ is small.

The power of the formulas in Theorem 1 is apparent because continuously differentiable convex functions are especially easy to minimize numerically. Also revealed is the fact that β-CVaR can be calculated without first having to calculate the β-VaR on which its definition depends, which would be more complicated. The β-VaR may be obtained instead as a by-product, but the extra effort that this might entail (in determining the interval $A_\beta(\boldsymbol{x})$ and extracting its left endpoint, if it contains more than one point) can be omitted if β-VaR is not needed.

Furthermore, the integral in the definition (4) of $F_\beta(\boldsymbol{x}, \alpha)$ can be approximated in various ways. For example, this can be done by sampling the probability distribution of $\boldsymbol{y}$ according to its density $p(\boldsymbol{y})$. If the sampling generates a collection of vectors $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_q$, then the corresponding approximation to $F_\beta(\boldsymbol{x}, \alpha)$ is

$$\tilde{F}_\beta(\boldsymbol{x}, \alpha) = \alpha + \frac{1}{q(1-\beta)} \sum_{k=1}^{q} [f(\boldsymbol{x}, \boldsymbol{y}_k) - \alpha]^+. \tag{9}$$

The expression $\tilde{F}_\beta(\boldsymbol{x}, \alpha)$ is convex and piecewise linear with respect to α. Although it is not differentiable with respect to α, it can readily be minimized, either by line search techniques or by representation in terms of an elementary linear programming problem.

Other important advantages of viewing VaR and CVaR through the formulas in Theorem 1 are evident in the next theorem.

Theorem 2 *Minimizing the β-CVaR of the loss associated with $\boldsymbol{x}$ over all $\boldsymbol{x} \in X$ is equivalent to minimizing $F_\beta(\boldsymbol{x}, \alpha)$ over all $(\boldsymbol{x}, \alpha) \in X \times \mathbb{R}$, in the sense that*

$$\min_{\boldsymbol{x} \in X} \phi_\beta(\boldsymbol{x}) = \min_{(\boldsymbol{x}, \alpha) \in X \times \mathbb{R}} F_\beta(\boldsymbol{x}, \alpha), \tag{10}$$

where, moreover, a pair $(\boldsymbol{x}^*, \alpha^*)$ achieves the second minimum if and only if $\boldsymbol{x}^*$ achieves the first minimum and $\alpha^* \in A_\beta(\boldsymbol{x}^*)$. In particular, therefore, in circumstances where the interval $A_\beta(\boldsymbol{x}^*)$ reduces to a single point (as is typical), the minimization of $F(\boldsymbol{x}, \alpha)$ over $(\boldsymbol{x}, \alpha) \in X \times \mathbb{R}$ produces a pair $(\boldsymbol{x}^*, \alpha^*)$, not *necessarily unique, such that* $\boldsymbol{x}^*$ *minimizes the* β*-CVaR and* α^* *gives the corresponding* β*-VaR.*

Furthermore, $F_\beta(\boldsymbol{x}, \alpha)$ *is convex with respect to* $(\boldsymbol{x}, \alpha)$*, and* $\phi_\beta(\boldsymbol{x})$ *is convex with respect to* $\boldsymbol{x}$*, when* $f(\boldsymbol{x}, \boldsymbol{y})$ *is convex with respect to* $\boldsymbol{x}$*, in which case, if the constraints are such that X is a convex set, the joint minimization is an instance of convex programming.*

Again, the proof is furnished in the Appendix. According to Theorem 2, it is not necessary, for the purpose of determining an $\boldsymbol{x}$ that yields minimum β-CVaR, to work directly with the function $\phi_\beta(\boldsymbol{x})$, which may be hard to do because of the nature of its definition in terms of the β-VaR value $\alpha_\beta(\boldsymbol{x})$ and the often troublesome mathematical properties of that value. Instead, one can operate on the far simpler expression $F_\beta(\boldsymbol{x}, \alpha)$ with its convexity in the variable α and even, very commonly, with respect to $(\boldsymbol{x}, \alpha)$.

The optimization approach supported by Theorem 2 can be combined with ideas for approximating the integral in the definition (4) of $F_\beta(\boldsymbol{x}, \alpha)$ such as have already been mentioned. This offers a rich range of possibilities. Convexity of $f(\boldsymbol{x}, \boldsymbol{y})$ with respect to $\boldsymbol{x}$ produces convexity of the approximating expression $\tilde{F}_\beta(\boldsymbol{x}, \alpha)$ in (9), for instance.

The minimization of F_β over $X \times \mathbb{R}$ falls into the category of stochastic optimization, or more specifically stochastic programming, because of the presence of an "expectation" in the definition of $F_\beta(\boldsymbol{x}, \alpha)$. At least for the cases involving convexity, there is a vast literature on solving such problems (Birge and Louveaux 1997, Ermoliev and Wets 1988, Kall and Wallace 1995, Kan and Kibzun 1996, Pflug 1996, Prekopa 1995). Theorem 2 opens the door to applying stochastic programming approaches to the minimization of β-CVaR.

3. AN APPLICATION TO PORTFOLIO OPTIMIZATION

To illustrate the approach we propose, we now consider the case where the decision vector $\boldsymbol{x}$ represents a portfolio of financial instruments in the sense that $\boldsymbol{x} = (x_1, \ldots, x_n)$, with x_j being the position in instrument j and

$$x_j \geqslant 0 \quad \text{for } j = 1, \ldots, n, \ \text{ with } \sum_{j=1}^{n} x_j = 1. \tag{11}$$

Denoting by y_j the return on instrument j, we take the random vector to be $\boldsymbol{y} = (y_1, \ldots, y_n)$. The distribution of $\boldsymbol{y}$ constitutes a joint distribution of the various returns and is independent of $\boldsymbol{x}$; it has density $p(\boldsymbol{y})$.

The return on a portfolio $\boldsymbol{x}$ is the sum of the returns on the individual instruments in the portfolio, scaled by the proportions x_j. The loss, being the negative of this, is therefore given by

$$f(\boldsymbol{x}, \boldsymbol{y}) = -[x_1 y_1 + \cdots + x_n y_n] = -\boldsymbol{x}^{\mathsf{T}} \boldsymbol{y}. \tag{12}$$

As long as $p(\boldsymbol{y})$ is continuous with respect to $\boldsymbol{y}$, the cumulative distribution function for the loss associated with $\boldsymbol{x}$ will itself be continuous (see Kan and Kibzun 1996, Uryasev 1995).

Although VaR and CVaR are usually defined in terms of monetary value, here they are defined as percentage returns. We consider the case when there is a one-to-one correspondence between percentage return and monetary value (this may not be true for the portfolios with zero net investment). In this section, we compare the minimum CVaR methodology with the minimum variance approach, and so, to be consistent, we consider the loss in percentage terms.

The performance function on which we focus here in connection with β-VaR and β-CVaR is

$$F_\beta(\boldsymbol{x},\alpha) = \alpha + (1-\beta)^{-1} \int_{y\in\mathbb{R}^n} [-\boldsymbol{x}^{\mathsf{T}}\boldsymbol{y} - \alpha]^+ \, p(\boldsymbol{y})\,\mathrm{d}\boldsymbol{y}. \tag{13}$$

It is important to observe that, in this setting, $F_\beta(\boldsymbol{x},\alpha)$ is convex as a function of $\boldsymbol{x}$ and α, not just α. Often, it is also differentiable in these variables (see Kan and Kibzun 1996, Uryasev 1995). Such properties set the stage very attractively for implementation of the kinds of computational schemes suggested above.

For a closer look, let $\mu(\boldsymbol{x})$ and $\sigma(\boldsymbol{x})$ denote the mean and variance of the loss associated with portfolio $\boldsymbol{x}$; in terms of the mean $\boldsymbol{m}$ and variance V of $\boldsymbol{y}$, we have

$$\mu(\boldsymbol{x}) = -\boldsymbol{x}^{\mathsf{T}}\boldsymbol{m} \quad \text{and} \quad \sigma^2(\boldsymbol{x}) = \boldsymbol{x}^{\mathsf{T}} V \boldsymbol{x}. \tag{14}$$

Clearly, $\mu(\boldsymbol{x})$ is a linear function of $\boldsymbol{x}$, whereas $\sigma(\boldsymbol{x})$ is a quadratic function of $\boldsymbol{x}$. We impose the requirement that only portfolios that can be expected to return at least a given amount R will be admitted. In other words, we introduce the linear constraint

$$\mu(\boldsymbol{x}) \leqslant -R \tag{15}$$

and take the feasible set of portfolios to be

$$X = \big\{\, \boldsymbol{x} : \boldsymbol{x} \text{ satisfies (11) and (15)} \,\big\}. \tag{16}$$

This set X is convex (in fact "polyhedral", owing to linearity in all the constraints). The problem of minimizing F_β over $X \times \mathbb{R}$ is therefore one of convex programming, for the reasons laid out in Theorem 2.

Now consider the kind of approximation of F_β obtained by sampling the probability distribution in $\boldsymbol{y}$, as in (9). A sample set $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_q$ yields the approximate function

$$\tilde{F}_\beta(\boldsymbol{x},\alpha) = \alpha + \frac{1}{q(1-\beta)} \sum_{k=1}^{q} [-\boldsymbol{x}^{\mathsf{T}}\boldsymbol{y}_k - \alpha]^+. \tag{17}$$

The minimization of $\tilde{F}_\beta$ over $X \times \mathbb{R}$, in order to get an approximate solution to the minimization of F_β over $X \times \mathbb{R}$, can in fact be reduced to convex programming. In terms of auxiliary real variables u_k for $k = 1, \ldots, r$, it is equivalent to minimizing the linear expression

$$\alpha + \frac{1}{q(1-\beta)} \sum_{k=1}^{q} u_k$$

subject to the linear constraints (11), (15), and

$$u_k \geqslant 0 \quad \text{and} \quad \boldsymbol{x}^{\mathsf{T}} \boldsymbol{y}_k + \alpha + u_k \geqslant 0 \qquad \text{for } k = 1, \ldots, r.$$

Note that the possibility of such reduction to linear programming does not depend on $\boldsymbol{y}$ having a special distribution, such as a normal distribution; it works for nonnormal distributions just as well.

The discussion so far has been directed toward minimizing β-CVaR, or, in other words, the problem

(P1) $\qquad$ minimize $\phi_\beta(\boldsymbol{x})$ over $\boldsymbol{x} \in X$,

since that is what is accomplished, on the basis of Theorem 2, when F_β is minimized over $X \times \mathbb{R}$. The related problem of finding a portfolio that minimizes β-VaR (Kast *et al.* 1998, Mauser and Rosen 1999), i.e., one that solves the problem

(P2) $\qquad$ minimize $\alpha_\beta(\boldsymbol{x})$ over $\boldsymbol{x} \in X$,

is not covered directly. However, because $\phi_\beta(\boldsymbol{x}) \geqslant \alpha_\beta(\boldsymbol{x})$, solutions to (P1) should also be good from the perspective of (P2). According to Theorem 2, the technique of minimizing $F_\beta(\boldsymbol{x}, \alpha)$ over $X \times \mathbb{R}$ to solve (P1) also determines the β-VaR of the portfolio $\boldsymbol{x}^*$ that minimizes β-CVaR. This is not the same as solving (P2), but anyway it appears that (P1) is a better problem to be solving for risk management than (P2).

In this framework, it is useful also to compare (P1) and (P2) with a very popular problem, that of minimizing variance (see Markowitz 1952):

(P3) $\qquad$ minimize $\sigma^2(\boldsymbol{x})$ over $\boldsymbol{x} \in X$.

An attractive mathematical feature of problem (P3) is that it reduces to quadratic programming, but like (P2) its suitability has been questioned. Many other approaches could, of course, also be mentioned. The mean absolute deviation approach of Konno and Yamazaki (1991), the regret optimization approach of Dembo (1995) and Dembo and King (1992), and the minimax approach described by Young (1998) are notable in connection with the approximation scheme (17) for CVaR minimization because they also use linear programming algorithms.

These problems can yield, in at least one important case, the same optimal portfolio $\boldsymbol{x}^*$. We establish this fact next and then put it to use in numerical testing.

PROPOSITION 1 *Suppose that the loss associated with each* $\boldsymbol{x}$ *is normally distributed, as holds when* $\boldsymbol{y}$ *is normally distributed. If* $\beta \geqslant 0.5$ *and the constraint (15) is active at solutions to any two of the problems (P1), (P2), and (P3), then the solutions to those two problems are the same; a common portfolio* $\boldsymbol{x}^*$ *is optimal by both criteria.*

Proof. Using the analytical capabilities of the *Mathematica* package, under the normality assumption and with $\beta \geqslant 0.5$, we expressed the β-VaR and β-CVaR in terms of mean and variance by

$$\alpha_\beta(\boldsymbol{x}) = \mu(\boldsymbol{x}) + c_1(\beta)\sigma(\boldsymbol{x}), \quad \text{with } c_1(\beta) = \sqrt{2}\,\mathrm{erf}^{-1}(2\beta - 1) \tag{18}$$

and

$$\phi_\beta(\boldsymbol{x}) = \mu(\boldsymbol{x}) + c_2(\beta)\sigma(\boldsymbol{x}), \quad \text{with } c_2(\beta) = \left\{\sqrt{2\pi}\,\exp[\mathrm{erf}^{-1}(2\beta-1)]^2(1-\beta)\right\}^{-1}, \tag{19}$$

where erf^{-1} denotes the inverse of the error function, which is defined by

$$\mathrm{erf}(z) = \frac{2}{\sqrt{\pi}}\int_0^z \mathrm{e}^{-t^2}\,\mathrm{d}t.$$

When the constraint (15) is active at optimality, the set X can just as well be replaced in the minimization by the generally smaller set X' obtained by substituting the equation $\mu(\boldsymbol{x}) = -R$ for the inequality $\mu(\boldsymbol{x}) \leqslant -R$. For $\boldsymbol{x} \in X'$, however, we have

$$\alpha_\beta(\boldsymbol{x}) = -R + c_1(\beta)\sigma(\boldsymbol{x}) \quad \text{and} \quad \phi_\beta(\boldsymbol{x}) = -R + c_2(\beta)\sigma(\boldsymbol{x}),$$

where the coefficients $c_1(\beta)$ and $c_2(\beta)$ are positive. Minimizing either of these expressions over $\boldsymbol{x} \in X'$ is evidently the same as minimizing $\sigma(\boldsymbol{x})^2$ over $\boldsymbol{x} \in X'$. Thus, if the constraint (15) is active in two of the problems, then any portfolio $\boldsymbol{x}^*$ that minimizes $\sigma(\boldsymbol{x})$ over $\boldsymbol{x} \in X'$ is optimal for those two problems. □

Proposition 1 furnishes an opportunity of using quadratic programming solutions for problem (P3) as a benchmark in testing the method of minimizing β-CVaR by the sampling approximations in (17) and their reduction to linear programming. We carry this out for an example in which an optimal portfolio is to be constructed from three instruments: S&P 500, a portfolio of long-term US government bonds, and a portfolio of small-cap stocks, the returns on these instruments being modeled by a (joint) normal distribution. The calculations were conducted by Carlos Testuri as part of a project in the Stochastic Optimization course at the University of Florida.

The mean $\boldsymbol{m}$ of monthly returns and the covariance matrix V in this example are given in Tables 1 and 2, respectively. We took $R = 0.011$ in the constraint (15) on expected loss/return.

First, we solved the quadratic programming problem (P3) for these data elements, obtaining the portfolio $\boldsymbol{x}^*$ displayed in Table 3 as the unique optimal portfolio in the Markowitz minimum variance sense. The corresponding

TABLE 1. Portfolio mean return.

Instrument	Mean return
S&P	0.0101110
Gov. bond	0.0043532
Small cap	0.0137058

TABLE 2. Portfolio covariance matrix.

	S&P	Gov. bond	Small cap
S&P	0.00324625	0.00022983	0.00420395
Gov. bond	0.00022983	0.00049937	0.00019247
Small cap	0.00420395	0.00019247	0.00764097

TABLE 3. Optimal portfolio with the minimum variance approach.

S&P	Gov. bond	Small cap
0.452013	0.115573	0.432414

TABLE 4. VaR and CVaR obtained with the minimum variance approach.

	$\beta = 0.90$	$\beta = 0.95$	$\beta = 0.99$
VaR	0.067847	0.090200	0.132128
CVaR	0.096975	0.115908	0.152977

variance was $\sigma(\boldsymbol{x}^*)^2 = 0.00378529$ and the mean was $\mu(\boldsymbol{x}^*) = -0.011$; thus, the constraint (15) was active in this instance of (P3). Then, for the β-values 0.99, 0.95, and 0.90, we calculated the β-VaR and β-CVaR of this portfolio $\boldsymbol{x}^*$ from (18) and (19), obtaining the results given in Table 4.

With these values at hand for comparison purposes, we proceeded with our approach, based on Theorem 2, of solving the β-CVaR problem (P1) by minimizing $F_\beta(\boldsymbol{x}, \alpha)$ over $(\boldsymbol{x}, \alpha) \in X \times \mathbb{R}$. To approximate the integral in the expression (13) for $F_\beta(\boldsymbol{x}, \alpha)$, we sampled the return vector $\boldsymbol{y}$ according to its density $p(\boldsymbol{y})$ in the (multi)normal distribution $\mathcal{N}(\boldsymbol{m}, V) \in \mathbb{R}^3$. The samples produced approximations $\tilde{F}_\beta(\boldsymbol{x}, \alpha)$ as in (17). The minimization of $\tilde{F}_\beta(\boldsymbol{x}, \alpha)$ over $(\boldsymbol{x}, \alpha) \in X \times \mathbb{R}$ was converted in each case to a linear programming problem in the manner explained after (17). These approximate calculations yielded estimates $\boldsymbol{x}^*$ for the optimal portfolio in (P1) along with corresponding estimates α^* for their β-VaR and $\tilde{F}_\beta(\boldsymbol{x}^*, \alpha^*)$ for their β-CVaR.

The linear programming calculations were carried out using the CPLEX linear programming solver on a 300 MHz Pentium II machine. In generating the random samples, we worked with two types of "random" numbers: the pseudorandom sequence of numbers (conventional Monte Carlo approach) and the Sobol quasirandom sequence (Press *et al*. 1992, p. 310). For similar applications of the quasirandom sequences, see Birge (1995), Boyle *et al*. (1997), and Kreinin *et al*. (1998). The results for the pseudorandom sequence are shown in Table 5, while those for the quasirandom sequence are shown in Table 6.

In comparing the results in Table 5 for our minimum CVaR approach with pseudorandom sampling to those that correspond to the optimal portfolio under the minimum variance approach in Tables 3 and 4, we see that the CVaR values differ by only few percentage points, depending upon the number of samples, and likewise for the VaR values. However, the convergence of the CVaR estimates in Table 5 to the values in Table 4 (which Proposition 1 leads us to expect) is slow at best. This slowness might be attributable to the sampling errors in the Monte Carlo simulations. Besides, at optimality the variance, VaR, and CVaR appear to have low sensitivities to the changes in the portfolio positions.

TABLE 5. The portfolio, VaR, and CVaR from minimum CVaR approach: Monte Carlo simulations generated by pseudorandom numbers.

β	Sample size	S&P	Gov. bond	Small cap	VaR	VaR diff. (%)	CVaR	CVaR diff. (%)	Itera- tions	Time (min)
0.9	1000	0.35250	0.15382	0.49368	0.06795	0.154	0.09962	2.73	1157	0.0
0.9	3000	0.55726	0.07512	0.36762	0.06537	3.645	0.09511	−1.92	636	0.0
0.9	5000	0.42914	0.12436	0.44649	0.06662	1.809	0.09824	1.30	860	0.1
0.9	10000	0.48215	0.10399	0.41386	0.06622	2.398	0.09503	−2.00	2290	0.3
0.9	20000	0.45951	0.11269	0.42780	0.06629	−2.299	0.09602	−0.98	8704	1.5
0.95	1000	0.53717	0.08284	0.37999	0.09224	2.259	0.11516	−0.64	156	0.0
0.95	3000	0.54875	0.07839	0.37286	0.09428	4.524	0.11888	2.56	652	0.0
0.95	5000	0.57986	0.06643	0.35371	0.09175	1.715	0.11659	0.59	388	0.1
0.95	10000	0.47102	0.10827	0.42072	0.08927	−1.03	0.11467	−1.00	1451	0.2
0.95	20000	0.49038	0.10082	0.40879	0.09136	1.284	0.11719	1.11	2643	0.7
0.99	1000	0.41844	0.12848	0.45308	0.13454	1.829	0.14513	−5.12	340	0.0
0.99	3000	0.61960	0.05116	0.32924	0.12791	−3.187	0.14855	−2.89	1058	0.0
0.99	5000	0.63926	0.04360	0.31714	0.13176	−0.278	0.15122	−1.14	909	0.1
0.99	10000	0.45203	0.11556	0.43240	0.12881	−2.51	0.14791	−3.31	680	0.1
0.99	20000	0.45766	0.11340	0.42894	0.13153	−0.451	0.15334	0.24	3083	0.9

TABLE 6. The portfolio, VaR, and CVaR from minimum CVaR approach: simulations generated by quasirandom Sobel sequences.

β	Sample size	S&P	Gov. bond	Small cap	VaR	VaR diff. (%)	CVaR	CVaR diff. (%)	Itera-tions	Time (min)
0.9	1000	0.43709	0.12131	0.44160	0.06914	1.90	0.09531	−1.71	429	0.0
0.9	3000	0.45425	0.11471	0.43104	0.06762	−0.34	0.09658	−0.41	523	0.0
0.9	5000	0.44698	0.11751	0.43551	0.06784	−0.02	0.09664	−0.35	837	0.1
0.9	10000	0.45461	0.11457	0.43081	0.06806	0.32	0.09695	−0.02	1900	0.3
0.9	20000	0.46076	0.11221	0.42703	0.06790	0.08	0.09692	−0.06	4818	0.6
0.95	1000	0.43881	0.12065	0.44054	0.09001	−0.21	0.11249	−2.95	978	0.0
0.95	3000	0.43881	0.12065	0.44054	0.09001	−0.21	0.11511	−0.69	407	0.0
0.95	5000	0.46084	0.11218	0.42698	0.09036	0.18	0.11516	−0.64	570	0.1
0.95	10000	0.45723	0.11357	0.42920	0.09016	−0.05	0.11577	−0.12	1345	0.2
0.95	20000	0.45489	0.11447	0.43064	0.09023	0.03	0.11577	−0.12	1851	0.7
0.99	1000	0.52255	0.08846	0.38899	0.12490	−5.47	0.14048	−8.17	998	0.0
0.99	3000	0.43030	0.12392	0.44578	0.12801	−3.12	0.15085	−1.39	419	0.0
0.99	5000	0.45462	0.11457	0.43081	0.13073	−1.06	0.14999	−1.95	676	0.1
0.99	10000	0.39156	0.13881	0.46963	0.13288	0.57	0.15208	−0.59	1065	0.2
0.99	20000	0.46065	0.11225	0.42710	0.13198	−0.11	0.15211	−0.57	1317	0.5

The results obtained in Table 6 from our minimum CVaR approach with quasirandom sampling exhibit different and better behavior. There is relatively fast convergence to the values for the minimum variance problem. When the sample size is above 10 000, the differences in CVaR and VaR obtained with the minimum CVaR and the minimum variance approaches are less than 1%.

4. AN APPLICATION TO HEDGING

As a further illustration of our approach, we consider next an example where a Nikkei portfolio is hedged. This problem, out of Mauser and Rosen (1999), was provided to us by the research group of Algorithmics Inc. Mauser and Rosen (1999) considered two ways of hedging: parametric and simulation VaR techniques. In each case, the best hedge is calculated by one-instrument minimization of VaR, i.e., by keeping all but one of the positions in the portfolio fixed and varying that one, within a specified range, until the VaR of the portfolio appears to be as low as possible. Here, we show first that when the same procedure is followed but in terms of minimizing CVaR, the one-instrument hedges obtained are very close to the ones obtained in terms of minimizing VaR. We go on to show, however, that CVaR minimization has the advantage of being practical beyond the one-instrument setting. Positions of several, or even many, instruments may be adjusted simultaneously in a broader mode of hedging.

As in the application to portfolio optimization in the preceding section, the calculations could be reduced to linear programming by the kind of maneuver described after (16), which adds an extra variable for each scenario that is introduced. This might be advantageous for hedges involving the simultaneous adjustment of positions in a large number of instruments (say >1000), but we demonstrate here that, for hedges with relatively few instruments being adjusted, nonsmooth optimization techniques can compete with linear programming. With such techniques, there is no need to add extra variables, and the dimension of the problem stays the same regardless of how many scenarios are considered.

Table 7 shows a portfolio that implements a butterfly spread on the Nikkei index, as of 1 July 1997. In addition to common shares of Komatsu and Mitsubishi, the portfolio includes several European call and put options on these equities. This portfolio makes extensive use of options to achieve the desired payoff profile. Figure 1, reproduced from Mauser and Rosen (1999), shows the distribution of one-day losses over a set of 1000 Monte Carlo scenarios. It indicates that the normal distribution fits the data poorly. Therefore, minimum CVaR and minimum variance approaches could, in this case, lead to quite different optimal solutions.

For the 11 instruments in question, let $\boldsymbol{x}$ be the vector of positions in the portfolio to be determined, in contrast to $\boldsymbol{z}$, the vector of initial positions in Table 7 (the fifth column). These vectors belong to $\mathbb{R}^{11}$. In the hedging, we were only concerned, of course, with varying some of the positions in $\boldsymbol{x}$ away from those in $\boldsymbol{z}$, but we wanted to test out different combinations. This can be thought of in terms of selecting an index set J within $\{1, \ldots, 11\}$ to indicate the instruments that are open to adjustment. In the case of one-instrument hedging,

TABLE 7. Nikkei portfolio, reproduced from Mauser and Rosen (1999).

Instrument	Type	Days to maturity	Strike price (10^3 JPY)	Position (10^3)	Value (10^3 JPY)
Mitsubishi EC 6mo 860	Call	184	860	11.5	563340
Mitsubishi Corp	Equity	n/a	n/a	2.0	1720000
Mitsubishi Cjul29 800	Call	7	800	−16.0	−967280
Mitsubishi Csep30 836	Call	70	836	8.0	382070
Mitsubishi Psep30 800	Put	70	800	40.0	2418012
Komatsu Ltd	Equity	n/a	n/a	2.5	2100000
Komatsu Cjul29 900	Call	7	900	−28.0	−11593
Komatsu Cjun2 670	Call	316	670	22.5	5150461
Komatsu Cjun2 760	Call	316	760	7.5	1020110
Komatsu Paug31 760	Put	40	760	−10.0	−68919
Komatsu Paug31 830	Put	40	830	10.0	187167

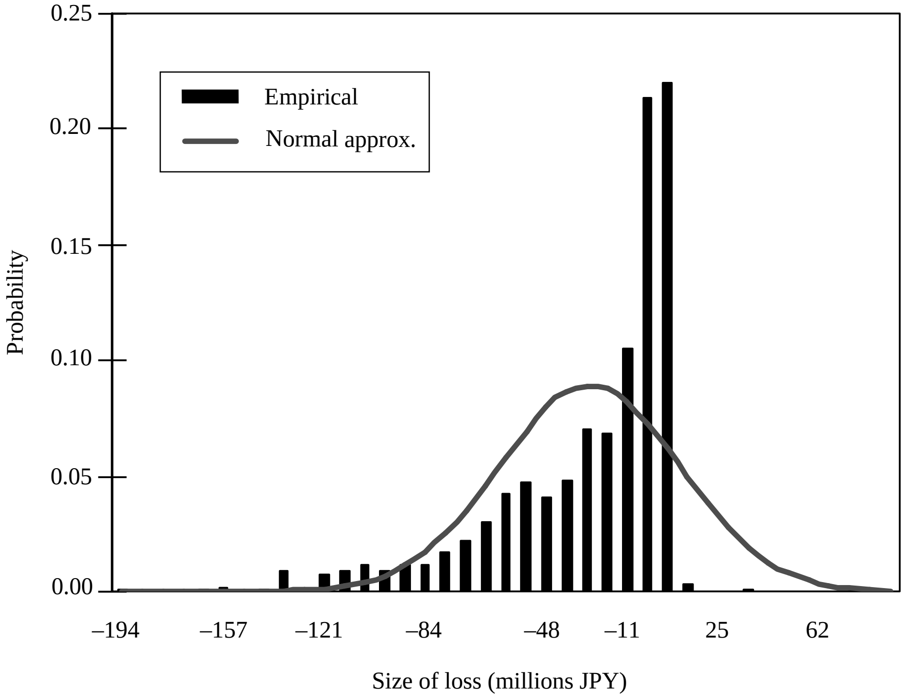

FIGURE 1. Distribution of losses for the Nikkei portfolio with best normal approximation (1000 scenarios), reproduced from Mauser and Rosen (1999).

for instance, we took J to specify a single instrument but consecutively went through different choices of that instrument.

Having selected a particular J, for the case when J contains more than one index, we imposed, on the coordinates x_j of $\boldsymbol{x}$, the constraints

$$-|z_j| \leqslant x_j \leqslant |z_j| \quad \text{for } j \in J, \tag{20}$$

and

$$x_j = z_j \quad \text{for } j \notin J, \tag{21}$$

thus taking

$$X = \{\boldsymbol{x} : \boldsymbol{x} \text{ satisfies (20) and (21)}\}. \tag{22}$$

The constraints (21) could, of course, be used to eliminate the variables x_j for $j \notin J$ from the problem, which we did in practice, but this formulation simplifies the notation and facilitates comparisons between different choices of J. The absolute values appear in (20) because short positions are represented by negative numbers.

Let $\boldsymbol{m}$ be the vector of initial prices (per unit) of the instruments in question and let $\boldsymbol{y}$ be the random vector of prices one day later. The loss to be dealt with in this context is the initial value of the entire portfolio minus its value one day later, namely

$$f(\boldsymbol{x}, \boldsymbol{y}) = \boldsymbol{x}^{\mathsf{T}}\boldsymbol{m} - \boldsymbol{x}^{\mathsf{T}}\boldsymbol{y} = \boldsymbol{x}^{\mathsf{T}}(\boldsymbol{m} - \boldsymbol{y}). \tag{23}$$

The corresponding function in our CVaR minimization approach is therefore

$$F_\beta(\boldsymbol{x}, \alpha) = \alpha + (1 - \beta)^{-1} \int_{y \in \mathbb{R}^{11}} [\boldsymbol{x}^{\mathrm{T}}(\boldsymbol{m} - \boldsymbol{y}) - \alpha]^{+}\, p(\boldsymbol{y})\, \mathrm{d}\boldsymbol{y}. \tag{24}$$

The problem to be solved, in accordance with Theorem 2, is that of minimizing $F_\beta(\boldsymbol{x}, \alpha)$ over $X \times \mathbb{R}$. This is the minimization of a convex function over a convex set.

To approximate the integral, we generated sample points $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_q$ and accordingly replaced $F_\beta(\boldsymbol{x}, \alpha)$ by

$$\tilde{F}_\beta(\boldsymbol{x}, \alpha) = \alpha + \frac{1}{q(1 - \beta)} \sum_{k=1}^{q} [\boldsymbol{x}^{\mathrm{T}}(\boldsymbol{m} - \boldsymbol{y}_k) - \alpha]^{+}, \tag{25}$$

an expression that is again convex in $(\boldsymbol{x}, \alpha)$ and, moreover, piecewise linear. Passing from there to the minimization of $\tilde{F}_\beta(\boldsymbol{x}, \alpha)$ over $X \times \mathbb{R}$, we could have converted the calculations to linear programming, but instead, as already explained, took the route of nonsmooth optimization. This involved working with the subgradient (or generalized gradient) set associated with $\tilde{F}_\beta$ at $(\boldsymbol{x}, \alpha)$, which consists of all vectors in $\mathbb{R}^{11} \times \mathbb{R}$ of the form

$$(\boldsymbol{0}, 1) + \frac{1}{q(1 - \beta)} \sum_{k=1}^{q} \lambda_k (\boldsymbol{m} - \boldsymbol{y}_k, -1) \quad \text{with} \begin{cases} \lambda_k = 1 & \text{if } \boldsymbol{x}^{\mathrm{T}}(\boldsymbol{m} - \boldsymbol{y}_k) - \alpha > 0, \\ \lambda_k \in [0, 1] & \text{if } \boldsymbol{x}^{\mathrm{T}}(\boldsymbol{m} - \boldsymbol{y}_k) - \alpha = 0, \\ \lambda_k = 0 & \text{if } \boldsymbol{x}^{\mathrm{T}}(\boldsymbol{m} - \boldsymbol{y}_k) - \alpha < 0. \end{cases} \tag{26}$$

We used the variable metric algorithm developed by Uryasev (1991) for nonsmooth optimization problems, taking $\beta = 0.95$, which made the initial β-VaR and β-CVaR values of the portfolio 657 816 and 2 022 060.

The results for the one-instrument hedging tests, where we followed our approach to minimize β-CVaR with $J = \{1\}$, then with $J = \{2\}$, and so forth, are presented in Table 8. The optimal hedges we obtained are close to the ones that were obtained by Mauser and Rosen (1999) by minimizing β-VaR. Because J consisted of a single index, $\boldsymbol{x}$ was just one-dimensional in these tests; minimization with respect to $(\boldsymbol{x}, \alpha)$ was therefore two-dimensional. The algorithm needed less than 100 iterations to find 6 correct digits in the performance function and variables.

For testing purposes, we employed the *Mathematica* version of the variable metric code on a 450 MHz Pentium II machine. (The Fortran and *Mathematica* versions of the code are available at *http://www.ise.ufl.edu/uryasev*.) The constraints were accounted for by nonsmooth penalty functions. Each run took less than one minute of computer time. The calculation time could be significantly improved using the algorithm implemented with Fortran or C; however, such computational studies were beyond the scope of this paper.

After finishing with the one-instrument tests, we tried hedging with respect

TABLE 8. Best hedge and corresponding VaR and CVaR with minimum CVaR approach: one-instrument hedges ($\beta = 0.95$).

Instrument	Best hedge	VaR	CVaR
Mitsubishi EC 6mo 860	7337.53	−205927	1183040
Mitsubishi Corp	−926.073	−1180000	551892
Mitsubishi Cjul29 800	−18978.6	−1170000	553696
Mitsubishi Csep30 836	4381.22	−1150000	549022
Mitsubishi Psep30 800	43637.1	−1150000	542168
Komatsu Ltd	−196.167	−1180000	551892
Komatsu Cjul29 900	−124939	−1200000	593078
Komatsu Cjun2 670	19964.9	−1220000	385698
Komatsu Cjun2 760	4745.20	−1200000	363556
Komatsu Paug31 760	31426.3	−1120000	538662
Komatsu Paug31 830	19356.3	−1150000	536500

the last 4 of the 11 instruments, simultaneously. The optimal hedge we determined in this way is indicated in Table 9. The optimization did not change the positions of Komatsu Cjun2 670 and Komatsu Paug31 760, but the positions of Komatsu Cjun2 760 and Komatsu Paug31 830 changed not only in magnitude but also in sign. In comparison with one-instrument hedging, we observe that multiple-instrument hedging considerably improved the VaR and CVaR. In this case, the final β-VaR equals $-1\,400\,000$ and the final β-CVaR equals 37 334.6, which is lower than best one-dimension hedge with $\beta - \text{VaR} = -1\,200\,000$ and $\beta - \text{CVaR} = 363\,556$ (see line 9 of Table 8). Six correct digits in the performance function and the positions were obtained after 400–800 iterations of the variable metric algorithm by Uryasev (1991), depending upon the initial parameters. It took about 4–8 minutes with a *Mathematica* version of the variable metric code on a 450 MHz Pentium II.

In contrast to the application in the preceding section, where we used linear programming techniques, the dimension of the nonsmooth optimization

TABLE 9. Initial positions and best hedge with minimum CVaR approach: simultaneous optimization with respect to four instruments ($\beta = 0.95$; VaR of best hedge equals $-1\,400\,000$, whereas CVaR equals 37334.6).

Instrument	Position in portfolio	Best hedge
Komatsu Cjun2 670	22500	22500
Komatsu Cjun2 760	7500	−527
Komatsu Paug31 760	−10000	−10000
Komatsu Paug31 830	10000	−10000

problem does not change with an increase in the number of scenarios. This may give some computational advantages for problems with a very large number of scenarios.

This example clearly shows, by the way, the superiority of CVaR over VaR in capturing risk. Portfolios are displayed that have positive β-CVaR but negative β-VaR for the same level of $\beta = 0.95$. The portfolio corresponding to the first line of Table 8, for instance, has β-VaR equal to $-205\,927$ but β-CVaR equal to 1 183 040. A negative loss is, of course, a gain.[1] The portfolio in question will thus result with probability 0.95 in a gain of 205 927 or more. That figure does not reveal, however, how serious the outcome might be the rest of the time. The CVaR figure says in fact that, in the cases where the gain of at least 205 927 is not realized, there is, on the average, a loss of 1 183 040.

5. CONCLUDING REMARKS

The paper considered a new approach for simultaneous calculation of VaR and optimization of CVaR for a broad class of problems. We showed that CVaR can be efficiently minimized using linear programming and nonsmooth optimization techniques. Although, formally, the method minimizes only CVaR, our numerical experiments indicate that it also lowers VaR because CVaR $\geqslant$ VaR.

We demonstrated with two examples that the approach provides valid results. These examples have relatively low dimensions and are offered here for illustrative purposes. Numerical experiments have been conducted for larger problems, but those results will be presented elsewhere in a comparison of numerical aspects of various linear programming techniques for portfolio optimization.

There is room for much improvement and refinement of the suggested approach. For instance, the assumption that there is a joint density of instrument returns can be relaxed. Furthermore, extensions can be made to optimization problems with value-at-risk constraints. Linear programming and nonsmooth optimization algorithms that utilize the special structure of the minimum CVaR approach can be developed. Additional research needs to be conducted on various theoretical and numerical aspects of the methodology.

APPENDIX

Central to establishing Theorems 1 and 2 is the following fact about the behavior with respect to α of the integral expression in the definition (4) of $F_\beta(\boldsymbol{x}, \alpha)$. We rely here on our assumption that $\Psi(\boldsymbol{x}, \alpha)$ is continuous with respect to α, which is equivalent to knowing that, regardless of the choice of $\boldsymbol{x}$, the set

[1] VaR may be negative because it is defined relative to zero, but not relative to the mean as in VaR based on the standard deviation.

of y with $f(x, y) = \alpha$ has probability zero, i.e.,

$$\int_{f(x,y)=\alpha} p(y)\,\mathrm{d}y = 0. \tag{27}$$

Lemma 1 *With x fixed, let $G(\alpha) = \int_{y \in \mathbb{R}^m} g(\alpha, y) p(y)\,\mathrm{d}y$, where $g(\alpha, y) = [f(x, y) - \alpha]^+$. Then G is a convex continuously differentiable function with derivative*

$$G'(\alpha) = \Psi(x, \alpha) - 1. \tag{28}$$

Proof. This lemma follows from Proposition 2.1 of Shapiro and Wardi (1994). □

Proof of Theorem 1. In view of the defining formula for $F_\beta(x, \alpha)$ in (4), it is immediate from Lemma 1 that $F_\beta(x, \alpha)$ is convex and continuously differentiable with derivative

$$\frac{\partial}{\partial \alpha} F_\beta(x, \alpha) = 1 + (1 - \beta)^{-1}[\Psi(x, \alpha) - 1] = (1 - \beta)^{-1}[\Psi(x, \alpha) - \beta]. \tag{30}$$

Therefore the values of α that furnish the minimum of $F_\beta(x, \alpha)$, i.e., the ones comprising the set $A_\beta(x)$ in (6), are precisely those for which $\Psi(x, \alpha) - \beta = 0$. They form a nonempty closed interval, inasmuch as $\Psi(x, \alpha)$ is continuous and nondecreasing in α with limit 1 as $\alpha \to \infty$ and limit 0 as $\alpha \to -\infty$. This further yields the validity of the β-VaR formula in (7). In particular, then, we have

$$\min_{\alpha \in \mathbb{R}} F_\beta(x, \alpha) = F_\beta(x, \alpha_\beta(x)) = \alpha_\beta(x) + (1 - \beta)^{-1} \int_{y \in \mathbb{R}^m} [f(x, y) - \alpha_\beta(x)]^+\, p(y)\,\mathrm{d}y.$$

But the integral here equals

$$\int_{f(x,y) \geqslant \alpha_\beta(x)} [f(x, y) - \alpha_\beta(x)]\, p(y)\,\mathrm{d}y = \int_{f(x,y) \geqslant \alpha_\beta(x)} f(x, y)\, p(y)\,\mathrm{d}y - \alpha_\beta(x) \int_{f(x,y) \geqslant \alpha_\beta(x)} p(y)\,\mathrm{d}y,$$

where the first integral on the right is by definition $(1 - \beta)\phi_\beta(x)$ and the second is $1 - \Psi(x, \alpha_\beta(x))$ by virtue of (27). Moreover, $\Psi(x, \alpha_\beta(x)) = \beta$. Thus,

$$\min_{\alpha \in \mathbb{R}} F_\beta(x, \alpha) = \alpha_\beta(x) + (1 - \beta)^{-1}[(1 - \beta)\phi_\beta(x) - \alpha_\beta(x)(1 - \beta)] = \phi_\beta(x).$$

This confirms the formula for β-CVaR in (5) and completes the proof of Theorem 1. □

Proof of Theorem 2. The initial claims, surrounding (10), are elementary consequences of the formula for $\phi_\beta(x)$ in Theorem 1 and the fact that the minimization of $F_\beta(x, \alpha)$ with respect to $(x, \alpha) \in X \times \mathbb{R}$ can be carried out by first minimizing over $\alpha \in \mathbb{R}$ for fixed $\boldsymbol{x}$ and then minimizing the result over $\boldsymbol{x} \in X$.

Justification of the convexity claim starts with the observation that $F_\beta(\boldsymbol{x}, \alpha)$ is convex with respect to $(\boldsymbol{x}, \alpha)$ whenever the integrand $[f(\boldsymbol{x}, \boldsymbol{y}) - \alpha]^+$ in the formula (4) for $F_\beta(\boldsymbol{x}, \alpha)$ is itself convex with respect to $(\boldsymbol{x}, \alpha)$. For each $\boldsymbol{y}$, this integrand is the composition of the function $(\boldsymbol{x}, \alpha) \mapsto f(\boldsymbol{x}, \boldsymbol{y}) - \alpha$ with the nondecreasing convex function $t \mapsto [t]^+$, so, by the rules given in Rockafellar (1970, Theorem 5.1), it is convex as long as the function $(\boldsymbol{x}, \alpha) \mapsto f(\boldsymbol{x}, \boldsymbol{y}) - \alpha$ is convex. The latter is true when $f(\boldsymbol{x}, \boldsymbol{y})$ is convex with respect to $\boldsymbol{x}$. The convexity of the function $\phi_\beta(\boldsymbol{x})$ follows from the fact that minimizing an extended-real-valued convex function of two vector variables (with infinity representing constraints) with respect to one of these variables results in a convex function of the remaining variable (Rockafellar 1970, pp. 38–39). □

Acknowledgements

We are grateful to Carlos Testuri, who conducted numerical experiments for the example on comparison of the minimum CVaR and the minimum variance approaches for the portfolio optimization. We also want to thank the research group of Algorithmics Inc. for the fruitful discussions and providing data for conducting numerical experiments with the Nikkei portfolio of options.

REFERENCES

Andersson, F., and Uryasev, S. (1999). Credit risk optimization with conditional value-at-risk criterion. Research Report 99-9, ISE Department, University of Florida.

Artzner, P., Delbaen F., Eber, J. M., and Heath, D. (1997). Thinking coherently. *Risk*, **10**(11), 68–71.

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

Beder, T. S. (1995). VAR: Seductive but dangerous. *Financial Analysts Journal*, **51**(5), 12–24.

Birge, J. R. (1995). Quasi–Monte Carlo methods approaches to option pricing. Technical Report 94-19, Department of Industrial and Operations Engineering, University of Michigan.

Birge, J. R., and Louveaux, F. (1997). *Introduction to Stochastic Programming*. Springer.

Boyle, P. P., Broadie, M., and Glasserman, P. (1997). Monte Carlo methods for security pricing. *Journal of Economic Dynamics and Control*, **21**(8/9), 1267–1321.

Bucay, N., and Rosen, D. (1999). Credit risk of an international bond portfolio: A case study. *ALGO Research Quarterly*, **2**(1), 9–29.

Dembo, R. S. (1995). Optimal portfolio replication. Algorithmics Technical Paper Series, 95–01.

Dembo, R. S. and King, A. J. (1992). Tracking models and the optimal regret distribution in asset allocation. *Applied Stochastic Models and Data Analysis*, **8**, 151–157.

Duffie, D., and Pan, J. (1997). An overview of value-at-risk. *Journal of Derivatives*, **4**, 7–49.

Embrechts, P., Kluppelberg, S., and Mikosch, T. (1997). *Extremal Events in Finance and Insurance*. Springer.

Embrechts, P., Resnick, S. I., and Samorodnitsky, G. (1999). Extreme value theory as a risk management tool. *North American Actuarial Journal*, **3**(2), 32–41.

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

Jorion, P. (1996). *Value at Risk: A New Benchmark for Measuring Derivatives Risk*. Irwin Professional Publishing.

JP Morgan (1996). *RiskMetrics Technical Manual*, 4th edn. JP Morgan, New York.

Kall, P., and Wallace, S.W. (1995). *Stochastic Programming*. Wiley.

Kast, R., Luciano, E., and Peccati, L. (1998). *VaR and Optimization*, Proceedings of the 2nd International Workshop on Preferences and Decisions, Trento, 1–3 July 1998.

Kan, Y. S., and Kibzun, A. I. (1996). *Stochastic Programming Problems with Probability and Quantile Functions*. Wiley.

Konno, H., and Yamazaki, H. (1991). Mean absolute deviation portfolio optimization model and its application to Tokyo Stock Market. *Management Science*, **37**, 519–531.

Kreinin, A., Merkoulovitch, L., Rosen, D., and Michael, Z. (1998). Measuring portfolio risk using quasi Monte Carlo methods. *ALGO Research Quarterly*, **1**(1), 17–25.

Litterman, R. (1997a). Hot spots and hedges, Part I. *Risk*. **10**(3), 42–45.

Litterman, R. (1997b). Hot spots and hedges, Part II. *Risk*, **10**(5), 38–42.

Lucas, A., and Klaassen, P. (1998). Extreme returns, downside risk, and optimal asset allocation. *Journal of Portfolio Management*, **25**(1), 71–79.

McKay, R., and Keefer, T. E. (1996). VaR is a dangerous technique. *Corporate Finance*, Searching for Systems Integration Supplement, September.

Markowitz, H. M. (1952). Portfolio selection. *Journal of Finance*, **7**(1), 77–91.

Mausser, H., and Rosen, D. (1999). Beyond VaR: From measuring risk to managing risk. *ALGO Research Quarterly*, **1**(2), 5–20.

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

Pflug, G. Ch. (2000). Some remarks on the value-at-risk and the conditional value-at-risk. In: *Probabilistic Constrained Optimization: Methodology and Applications* (ed. S. Uryasev). Kluwer.

Prekopa, A. (1995). *Stochastic Programming*. Kluwer.

Press, W. H., Teukolsky, S. A, Vetterling, W. T., and Flannery, B. P. (1992). *Numerical Recipes in C*. Cambridge University Press.

Pritsker, M. (1997). Evaluating value at risk methodologies. *Journal of Financial Services Research*, **12**(2/3), 201–242.

Rockafellar, R. T. (1970). *Convex Analysis*, Princeton Mathematics, Vol. 28. Princeton University Press.

Shapiro, A. and Wardi, Y. (1994). Nondifferentiability of the steady-state function in discrete event dynamic systems. *IEEE Transactions on Automatic Control*, **39**, 1707–1711.

Shor, N. Z. (1985). *Minimization Methods for Non-Differentiable Functions*. Springer.

Simons, K. (1996). Value-at-risk new approaches to risk management. *New England Economic Review*, Sept/Oct, 3–13.

Stambaugh, F. (1996). Risk and value-at-risk. *European Management Journal*, **14**(6), 612–621.

Uryasev, S. (1991). New variable-metric algorithms for nondifferential optimization problems. *Journal of Optimization Theory and Applications*, **71**(2), 359–388.

Uryasev, S. (1995). Derivatives of probability functions and some applications. *Annals of Operations Research*, **56**, 287–311.

Uryasev, S. (2000). Conditional value-at-risk: Optimization algorithms and applications. *Financial Engineering News*, **14**, February, 1–5.

Young, M.R. (1998). A minimax portfolio selection rule with linear programming solution. *Management Science*, **44**(5), 673–683.

Zenios, S. A. (ed.) (1996). *Financial Optimization*. Cambridge University Press.

Ziemba, W. T., and Mulvey, J. M. (eds.) (1998). *Worldwide Asset and Liability Modeling*. Cambridge University Press.