FULL LENGTH PAPER

Series A

CrossMark

Varying confidence levels for CVaR risk measures and minimax limits

Edward Anderson[1] · Huifu Xu[2] · Dali Zhang[3]

Abstract

Conditional value at risk (CVaR) has been widely studied as a risk measure. In this paper we add to this work by focusing on the choice of confidence level and its impact on optimization problems with CVaR appearing in the objective and also the constraints. We start by considering a problem in which CVaR is minimized and investigate the way in which it approximates the minimax robust optimization problem as the confidence level is driven to one. We make use of a consistent tail condition which ensures that the CVaR of a random function will converge uniformly to its supremum as the confidence level increases, and establish an error bound for the CVaR optimal solution under second order growth conditions. The results are extended to a minimization problem with a constraint on the CVaR value which in the limit as the confidence level approaches one coincides with a problem having semi-infinite constraints. We study the sample average approximation scheme for the CVaR constraints and establish an exponential rate of convergence for the sample averaged optimal solution. We propose a procedure to explore the possibility of varying the confidence level to a lower value which can give an advantage when there is a need to find good solutions to CVaR-constrained problems out of sample. Our numerical results demonstrate that using the optimal solution to an adjusted problem with lower confidence level can lead to better overall performance.

Mathematics Subject Classification 90C15 · 62D99 · 65K05

✉ Dali Zhang
zhangdl@sjtu.edu.cn

Edward Anderson
edward.anderson@sydney.edu.au

Huifu Xu
H.Xu@soton.ac.uk

[1] University of Sydney Business School, University of Sydney, Sydney, NSW 2006, Australia

[2] School of Mathematics, University of Southampton, Southampton SO17 1BJ, UK

[3] Sino-US Global Logistics Institute, Antai College of Economics and Management, Shanghai Jiao Tong University, Shanghai 200030, China

1 Introduction

Conditional value at risk (CVaR), sometimes called Expected Shortfall, has received a great deal of attention as a measure of risk. In a financial context it has a number of advantages over the value at risk (VaR) measure that is commonly used, and CVaR has been proposed as the primary tool for banking capital regulation in the Basel III standard [1]. Nevertheless there remains some controversy over its effectiveness. As we will discuss below, since it is a tail risk measure, it is harder to estimate accurately than a quantile measure like VaR. However the difficulties with CVaR estimation can be partly overcome by reducing the confidence level (which we will call β). The usual confidence level for VaR is $\beta = 0.99$, and the argument for a high value is in part based on the need to "see into the tails". Since VaR does not explicitly take account of the size of any losses that are less probable than $1 - \beta$, taking a lower value of β can be seen as ignoring these extreme risks. CVaR on the other hand is impacted by all losses no matter how unlikely, and it is therefore appropriate to use a lower value of β than for VaR; an example of this is that the proposals under Basel III suggest $\beta = 0.975$.

In practice risk measures are often estimated on the basis of empirical data, rather than on the basis of a model of the underlying process. A sample-based estimate for CVaR, since it is based only on the worst few results, becomes more and more difficult as the β level approaches 1 and there are fewer data points available for the estimate. The end result is an estimator with a variance that, for a fixed sample size, usually increases as β approaches 1. This provides another reason for looking carefully at what happens as the value of β is changed.

Our interest in this paper is in optimization models in which a decision variable x is chosen either to minimize CVaR as an objective, or as a constraint. In particular, we explore how the choice of confidence level β affects these optimization problems because little attention has been given to this issue in the literature despite the huge importance of CVaR as a risk measure. We primarily focus on two questions. First we explore the convergence behavior of the optimal value and optimal solution as β is driven to 1. We will consider the uniform convergence of CVaR for a random function and its impact on the underlying optimization problems. The other question we consider is whether there is an advantage from choosing lower β values within the optimization framework and how we can achieve the computational benefits of this approach.

We start with the simple problem of choosing the decision variable x to minimize CVaR:

$$\text{MnCV}(\beta) : \min_{x \in X} \ \text{CVaR}_\beta(g(x, \xi)),$$

where $g : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ is a continuous loss function, and $\xi : \Omega \to Y \subset \mathbb{R}^m$ is a random vector defined on the probability space $(\Omega, \mathcal{F}, P)$ with support set Y.

In practice, it may not be easy to determine what value of confidence level should be chosen, i.e. how extreme the risks are that should be considered in the risk minimization. Confidence levels β that are very close to 1 correspond to more con-

servative behavior, in which we focus on more and more unlikely events. At its limit, $\text{CVaR}_\beta(g(x, \xi))$ converges to the supremum of $g(x, \xi)$. Consequently, we obtain the following minimax problem

$$\text{MnMx} : \min_{x \in X} \sup_{y \in Y} \ g(x, y).$$

Surprisingly little has been written about the relationship between $\text{MnCV}(\beta)$ and MnMx in terms of the optimal values and optimal solutions. We will come back to discuss the relationship between these problems in detail in Sects. 2 and 3.

We note that minimax problems in this form occur in a large number of applications in economics and engineering (see [2,3]). If we view y as an uncertain parameter, then we can see this as a robust minimization problem where an optimum decision on x is made in a way that protects against the impact of uncertainty in y. This kind of robust formulation dates back to the early work of Soyster [4]. Over the past decade, robust optimization has rapidly developed into a new area of optimization with substantial applications in operations research, finance, engineering and computer science (see the monograph [2]). In broad terms we consider a framework in which the problem MnMx is viewed as a robust version of the original risk minimization problem, while in the other direction we can see $\text{MnCV}(\beta)$ as an approximation of MnMx.

In practical applications, we may not have complete information about the distribution of ξ, or its support set Y. Thus the solution of $\text{MnCV}(\beta)$ must be carried out through sampling from the distribution for ξ. This may happen, for example, when the risk optimization model generates ξ from a simulation. The model also applies when we have access only to a historical set of values, $\xi_1, \xi_2, \ldots, \xi_N$. In the same way, for the minimax problem MnMx, we suppose that the possible values of $y \in Y$ are not given directly, but can be obtained by sampling from the states of nature Y.

In this situation we will estimate the value of $\text{CVaR}_\beta(g(x, \xi))$ simply from looking at the average of the highest $1 - \beta$ proportion of the sample losses that occur $g(x, \xi_i)$, $i = 1, 2, \ldots, N$. For values of β greater than $1 - 1/N$ this will mean simply looking at $\max g(x, \xi_i)$. We note that there are other methods that have been proposed for the estimation of CVaR, such as kernel-based methods [5] or a semi-parametric approach based on extreme value theory [6]. However our simpler approach is supported by some results from Chen [7] who argues for the use of this approach when a point estimation of CVaR is required. Moreover the simple approach will be adequate for establishing some specific bounds on convergence.

In the case that $\text{MnCV}(\beta)$ is used to approximate MnMx, we may use any convenient distribution for ξ provided that the support set coincides with Y.

The second type of problem we consider is to minimize a deterministic function under CVaR constraint:

$$\text{MnCnCV}(\beta) : \begin{cases} \min_{x \in X} \ h(x) \\ \text{s.t.} \ \ \text{CVaR}_\beta(g(x, \xi)) \leq U, \end{cases}$$

where $h : \mathbb{R}^n \to \mathbb{R}$ is a continuous function and U is a prescribed maximal level of CVaR that a decision maker may accept. Here and later on we write "s.t." for "subject

to" for brevity. Again we look into what happens to this problem as we increase β, reflecting a more and more conservative decision maker. In particular, we investigate its relationship with the following optimization problem with a semi-infinite constraint:

$$\text{MnCnMx} : \begin{cases} \min_{x \in X} h(x) \\ \text{s.t.} \ \ g(x, y) \leq U, \quad \text{for all } y \in Y, \end{cases}$$

where as before Y is the support set of the random variable ξ. If we view y as an uncertain parameter, then the semi-infinite constraints can also be understood as a robust constraint of the form $\sup_{y \in Y} g(x, y) \leq U$.

We may (redefining g as necessary) take $U = 0$ and obtain a standard optimization problem subject to the condition that a solution is feasible for all possible instances of the uncertain parameter y. Unless we specify otherwise we will assume that $U = 0$ in what follows. This problem has been considered by Calafiore and Campi [8] and special cases with applications in engineering design can be found in [9–11].

For the constrained risk problem $\text{MnCnCV}(\beta)$ we will need to determine the value of β that is appropriate: for example we may need to decide between minimizing h subject to the 97.5% CVaR being less than $10,000 or minimizing h subject to the 99% CVaR being less than $20,000.

We will investigate the case where the solution of $\text{MnCnCV}(\beta)$ is estimated through a sample, $\xi_1, \xi_2, \ldots \xi_N$, and this means that the choice of β value will affect the degree of accuracy in the estimation of the feasible region of the problem, with lower values of β giving more accurate estimations. We will explore the idea of opting for lower values of β in order to avoid the optimization being dependent on too small a number of samples.

The use of a sampling approach is also found in Calafiore and Campi [8,12] and [13]. These authors have investigated the problem MnCnMx when $h(x)$ is linear and $g(x, \xi)$ is convex in x. Specifically, they consider taking a Monte Carlo sample from the set of uncertain parameters Y and then approximating the semi-infinite constraints with a finite number of sample indexed constraints. They show that the resulting randomized solution fails to satisfy only a small proportion of the original constraints if a sufficient number of samples is drawn. An explicit bound on the measures of the original constraints that may be violated by the randomized solution is derived. The approach has been shown to be numerically efficient and it has been extensively applied to various stochastic and robust programs including chance constrained stochastic programs and multistage stochastic programs, see [14–16] and references therein.

It is straightforward to recast the minimum CVaR problem $\text{MnCV}(\beta)$ as a constrained risk problem. To see this observe that we may write $\text{MnCV}(\beta)$ equivalently as

$$\text{MnCV}(\beta)' : \begin{cases} \min_{x \in X, z \in \mathbb{R}} z \\ \text{s.t.} \quad \text{CVaR}_\beta(g(x, \xi) - z) \leq 0, \end{cases}$$

where z is a new variable and we use the fact that $\text{CVaR}_\beta(g(x, \xi)) - z = \text{CVaR}_\beta(g(x, \xi) - z)$. This is the form required for the constrained risk problem

MnCnCV(β) where the decision variables (x, z) are chosen from $X \times \mathbb{R}$. There is a technical problem that occurs because $X \times \mathbb{R}$ is not compact, but we can solve this by minimizing over $x \in X$, $z \in [-C, C]$ with a sufficiently large positive constant C. Since the constrained risk problem MnCnCV(β) is so important in practice, and subsumes the minimum CVaR problem, we will concentrate our analysis on this problem.

We believe that the main contributions of the paper can be summarized as follows.

– We derive a sufficient condition to ensure that the CVaR of a random function converges uniformly to its supremum and demonstrate that the condition may be satisfied by a large class of random functions (Proposition 1). A deterministic error bound in terms of β is derived for the difference between CVaR and the supremum (Theorem 1). The results are used to investigate the relationship between MnCV(β) and MnMx and the relationship between MnCnCV(β) and MnCnMx. Specifically, under second order growth conditions, we derive an error bound for the difference between the optimal solutions to MnCV(β) and MnMx (Theorem 2), and under a Slater condition, we establish an error bound for the difference between the optimal values of MnCnCV(β) and MnCnMx (Theorem 4).

– In the case when the true probability is not easily available for computing the characteristics of the underlying random functions, we study the sample average approximation scheme for the CVaR constraints and establish an exponential rate of convergence for the optimal solution obtained from this scheme as the sample size increases (Theorem 7). Moreover, since using a confidence level near 1 makes the computation of CVaR from samples unreliable, we explore the possibility of varying the confidence level to a lower value in problem MnCnCV(β) without significantly affecting its optimal solution (Theorems 5 and 6).

– For the MnCnCV(β) model, we propose a procedure which solves an adjusted problem with lower confidence level and a correspondingly lower value for the upper bound on risk. We use a resource allocation problem in project management both to illustrate our main theoretical results and also to demonstrate numerically that using the optimal solution to this adjusted problem can lead to better overall performance. We also show how this technique can be used in a portfolio optimization model.

The rest of the paper is organized as follows. In Sect. 2, we investigate uniform approximation of CVaR of a random function to its supremum and sufficient conditions for such approximation. In Sect. 3, we discuss the relationship between MnCV(β) and MnMx in terms of the optimal solutions and optimal values, and extend the discussion to MnCnCV(β) and MnCnMx. In Sect. 4 we give a discussion of how MnCnCV(β) can be solved using a sample average approximation scheme. Finally in Sect. 5, we apply the proposed models and established theoretical results to a resource allocation problem in project management. We also report some numerical test results for both the project management problem and a portfolio optimization problem.

2 Uniform CVaR approximation and error bound

In this section, we discuss uniform approximation of $CVaR_\beta(g(x,\xi))$ to $\sup_{y\in Y} g(x,y)$ and its error bound; this is to pave the way for investigating the optimization problems outlined in Sect. 1.

Let g be a continuous function defined as in MnCV(β). We assume that $\xi : \Omega \to Y$ continuous random vector defined on probability space $(\Omega, \mathcal{F}, P)$ with density function $\rho(y)$, and we treat $g(x, \xi(\omega))$ for $x \in X \subset \mathbb{R}^n$ as a random variable induced by $\xi(\omega)$.

As a function of α for fixed x, we let $F(x, \alpha)$ be the cumulative distribution function for $g(x, \xi(\omega))$, the loss associated with x. Hence

$$F(x, \alpha) = \Pr(g(x, \xi) \leq \alpha) = \int_{\{y \,:\, g(x,y)\leq\alpha\}} \rho(y)dy. \tag{1}$$

Since we are interested in the problem MnMx as a limit when $\beta \to 1$ and this involves the supremum of g, we need to be careful to link the set of all values that $\xi(\omega)$ can take and the density function ρ. Essentially we need to avoid cases with isolated points in Y, which occur with probability zero, since this may lead to $\sup_{y\in Y} g(x, y)$ being greater than $\sup\{\alpha : F(x, \alpha) < 1\}$. To rule out these irrelevant points, we can use a number of conditions, but for clarity we will assume that for any $y \in Y$, and any open set A containing y, $\Pr(\xi \in A) > 0$.

Following [17], we may define the value at risk (VaR) of $g(x, \xi)$ at a level $\beta \in (0, 1)$ as

$$VaR_\beta(g(x, \xi)) := \inf\{\alpha : F(x, \alpha) \geq \beta\}.$$

We will often consider x as fixed and write $F_x(\cdot)$ for $F(x, \cdot)$. In the case that $F_x(\alpha)$ is strictly monotonically increasing with respect to α, we will have a well defined inverse function for F_x and then

$$VaR_\beta(g(x, \xi)) = F_x^{-1}(\beta).$$

Consequently we can define the conditional value at risk at a confidence level β as

$$CVaR_\beta(g(x, \xi)) := \frac{1}{1-\beta} \int_{\{y:\, g(x,y)\geq VaR_\beta(g(x,\xi))\}} g(x, y)\rho(y)dy.$$

Since in this case the probability that $g(x, y) \geq VaR_\beta(g(x, \xi))$ is equal to $(1 - \beta)$ we can interpret this expression as the conditional expectation of the loss associated with x given that this loss is $VaR_\beta(g(x, \xi))$ or greater. Since we are assuming an inverse function for F_x then we can write

$$CVaR_\beta(g(x, \xi)) = \frac{1}{1-\beta} \int_\beta^1 F_x^{-1}(t)dt,$$

see for instance [18, Theorem 2.34].

Now we wish to explore the properties of $\text{VaR}_\beta(g(x, \xi))$ and $\text{CVaR}_\beta(g(x, \xi))$, as well as their approximations to $\sup_{y \in Y} g(x, y)$. First observe that immediately from the definition

$$\text{VaR}_\beta(g(x, \xi)) \leq \text{CVaR}_\beta(g(x, \xi)) \leq \sup_{y \in Y} g(x, y).$$

Our first result shows pointwise and uniform approximation of $\text{VaR}_\beta(g(x, \xi))$ to $\sup_{y \in Y} g(x, y)$. It is convenient to write $g_x(\xi)$ for $g(x, \xi)$ and we let $g_x^* = \sup_{y \in Y} g_x(y)$. There has been previous work [19] that establishes a connection between VaR_β and CVaR_β as $\beta \to 1$ with these two quantities being related through a ratio that depends on the tail index of the underlying distribution. Our results are different in that they relate to the difference between CVaR_β and the supremum, g_x^*.

Theorem 1 (Pointwise and uniform CVaR approximation) *Let $x \in X$ be fixed. Suppose that the support set of ξ coincides with Y. Then the following assertions hold.*

(i) *If for every $y \in Y$ and open set A containing y, $\text{Pr}(\xi \in A) > 0$, then*

$$\lim_{\beta \to 1} \text{VaR}_\beta(g_x(\xi)) = \lim_{\beta \to 1} \text{CVaR}_\beta(g_x(\xi)) = g_x^*. \tag{2}$$

(ii) *If, in addition, there exists α_0 such that $F_x(\alpha)$ is continuously differentiable with a non-increasing positive derivative for all $\alpha \in (\alpha_0, g_x^*)$, then for $\beta > F_x(\alpha_0)$, the function $\text{VaR}_\beta(g_x(\xi))$ is convex in β and*

$$\text{VaR}_{(1+\beta)/2}(g_x(\xi)) \leq \text{CVaR}_\beta(g_x(\xi)) \leq \frac{1}{2}\left(\text{VaR}_\beta(g_x(\xi)) + g_x^*\right). \tag{3}$$

(iii) *If, $g_x^* < \infty$ for every $x \in X$ and there exist positive constants $\beta_0 \in (0, 1)$, K and τ (independent of x) such that*

$$1 - F_x(\alpha) \geq K\left(g_x^* - \alpha\right)^\tau, \quad \text{for all } \alpha \in (\text{VaR}_{\beta_0}(g_x(\xi)), g_x^*), \quad \forall x \in X, \tag{4}$$

then for $\beta \in (\beta_0, 1)$,

$$\Delta_\beta(x) := g_x^* - \text{CVaR}_\beta(g(x, \xi)) \leq \frac{1}{K^{1/\tau}} \frac{\tau}{1+\tau}(1-\beta)^{1/\tau}, \quad \forall x \in X. \tag{5}$$

The proof to Theorem 1 is given in the "Appendix". Note that parts (i) and (ii) of this theorem hold when $g_x^* = \infty$, and concern pointwise convergence of $\text{CVaR}_\beta(g(x, \xi))$ to $\sup_{y \in Y} g(x, y)$ as β is driven to 1 whereas part (iii) gives uniform approximation and an error bound. Part (i) of the theorem is enough to establish that MnMx is the limit of MnCV(β) as $\beta \to 1$. Later we will look in more detail at this convergence.

Essentially the bounds given in parts (ii) and (iii) of the theorem rely on the good behavior of the cumulative distribution function F_x. The condition for part (ii) that the derivative of $F_x(\alpha)$ is decreasing (not necessarily strictly) for α large enough simply

requires that the density function of $g_x(\xi)$ has a largest mode less than the right hand end of its range (beyond which it is automatically decreasing).

Condition (4) is a kind of growth condition of the cumulative distribution function at $\beta = 1$. It requires the cumulative distribution function to approach 1 faster than some power of the distance to g_x^*. This will hold whenever the corresponding density function is bounded as $\alpha \to g_x^*$, (even less of a restriction than part (ii) requires). When ξ is univariate the parameter τ is determined by the way in which the density function of $g_x(\xi)$ approaches zero when ξ approaches its limit, but the relationship is complex when ξ is of higher dimension. Throughout the paper, we say that g has consistent tail behavior if it satisfies condition (4).

We can compare the bounds in parts (ii) and (iii) of Theorem 1. By the first inequality of (3),

$$g_x^* - \mathrm{CVaR}_\beta(g(x,\xi)) \le g_x^* - \mathrm{VaR}_{(1+\beta)/2}(g_x(\xi)).$$

If (4) holds, then using the fact that $\mathrm{VaR}_\beta(g_x(\xi)) = F_x^{-1}(\beta)$, we have

$$g_x^* - \mathrm{VaR}_\beta(g_x(\xi)) \le \left(\frac{1-\beta}{K}\right)^{1/\tau}$$

and hence

$$g_x^* - \mathrm{CVaR}_\beta(g(x,\xi)) \le \frac{1}{K^{1/\tau}}\left(1 - \frac{1+\beta}{2}\right)^{1/\tau} = \left(\frac{1-\beta}{2K}\right)^{1/\tau}. \qquad (6)$$

We can compare this with the bound appearing in part (iii) and note that $\frac{\tau}{1+\tau} < \left(\frac{1}{2}\right)^{1/\tau}$ when $\tau > 1$, so that part (iii) gives the stronger bound in this case. When $\tau < 1$, (4) implies that $F_x(\alpha)$ is bounded above by a function with increasing derivative with respect to α. This in turn implies that the density function cannot be decreasing as the random variable approaches it's upper limit, and this will contradict the requirements for part (ii).

Now we give an example to explain the main conditions and results in Theorem 1.

Example 1 Consider the case where $g(x,\xi) = \xi x$ with $x \in [0, 1]$ and ξ having a density function $\rho(y) = 2 - 2y$ over its support set $Y = [0, 1]$. Then $g_x^*(x) = x$ for $x \in [0, 1]$. When $x = 0$, $g(0,\xi) = 0$ which is deterministic. In what follows, we consider the case when $x \in (0, 1]$. It is easy to derive that for $\alpha < x$,

$$F_x(\alpha) = 1 - (1 - \alpha/x)^2$$

and $\mathrm{VaR}_\beta(g_x(\xi)) = x\left[1 - (1-\beta)^{1/2}\right]$. Thus

$$\mathrm{CVaR}_\beta(g_x(\xi)) = \frac{x}{1-\beta}\int_{1-(1-\beta)^{1/2}}^{1} y(2 - 2y)dy = x\left[1 - \frac{2}{3}(1-\beta)^{1/2}\right]. \qquad (7)$$

Since $g_x^* = x$, we obtain from the equation above that

$$g_x^* - \mathrm{CVaR}_\beta(g_x(\xi)) = \frac{2}{3}x(1-\beta)^{1/2} \leq \frac{2}{3}(1-\beta)^{1/2}, \quad \forall x \in (0, 1], \qquad (8)$$

which illustrates Theorem 1 (iii) with $K = 1$ and $\tau = 2$. In this case, the bound is exact. Moreover, since $\frac{1}{2} < \frac{2}{3} < \frac{\sqrt{2}}{2}$, we have from (7)

$$x\left(1 - \sqrt{\frac{1-\beta}{2}}\right) \leq \mathrm{CVaR}_\beta(g(x, \xi)) \leq x\left(1 - \frac{1}{2}\sqrt{1-\beta}\right),$$

which illustrates Theorem 1 (ii).

Let

$$\Delta(\beta) := \sup_{x \in X} \Delta_\beta(x). \qquad (9)$$

Then Theorem 1 (iii) can be written as

$$\Delta(\beta) \leq \frac{1}{K^{1/\tau}} \frac{\tau}{1+\tau}(1-\beta)^{1/\tau}. \qquad (10)$$

We now turn to discuss sufficient conditions for the consistent tail behavior of $g(x, \xi)$. The proposition below shows that when the loss function g is well-behaved, consistent tail behavior follows naturally from the behavior of the underlying random variable ξ.

Proposition 1 *Let Y be compact. If g has the property that there are positive constants C_1 and ν_1 (independent of x) with*

$$g(x, y_1) - g(x, y_2) < C_1 \|y_1 - y_2\|^{\nu_1} \qquad (11)$$

for all $x \in X$, $y_1, y_2 \in Y$; and if, in addition, there are positive constants C_2, ν_2 and δ_0 with

$$\Pr(\|\xi_0 - \xi\| < \delta) \geq C_2 \delta^{\nu_2} \qquad (12)$$

for every fixed ξ_0 and $\delta \in (0, \delta_0)$; then $g(x, \xi)$ has consistent tail behavior on X.

The proof to Proposition 1 is given in the "Appendix". The condition (12) is very weak: it can be guaranteed whenever the density function of ξ is lower bounded by a positive real valued function which is analytic. It is easily seen to hold when the density function is bounded away from zero around ξ_0, and so it is effectively a condition on the way that the density approaches zero at the boundary of its support. Condition (11) is known as uniform calmness in variational analysis, see [20, Chapter 8].

Before we conclude this subsection, we note that the CVaR risk measure can be reformulated as the result of a minimization. Let

$$\Phi_\beta(x, \eta) := \eta + \frac{1}{1-\beta} \int_{y \in Y} (g(x, y) - \eta)_+ \rho(y) dy, \tag{13}$$

where $(t)_+ = \max(0, t)$. Rockafellar and Uryasev [17] proved that

$$\text{CVaR}_\beta(g(x, \xi)) = \min_{\eta \in \mathbb{R}} \Phi_\beta(x, \eta), \tag{14}$$

and this allows us to reformulate the problem $\text{MnCV}(\beta)$ as

$$\min_{x \in X} \text{CVaR}_\beta(g(x, \xi)) = \min_{(x, \eta) \in X \times \mathbb{R}} \Phi_\beta(x, \eta). \tag{15}$$

This kind of reformulation is discussed comprehensively by Rockafellar and Uryasev, see e.g. [17, Theorem 2] and [21, Theorem 14], see also Ogryczak and Ruszczyński [22].

3 Approximation of the optimization problems

With the CVaR approximations in Sect. 2, we are now ready to discuss approximation of $\text{MnCV}(\beta)$ to MnMx, and $\text{MnCnCV}(\beta)$ to MnCnMx.

3.1 Approximation of MnCV(β) to MnMx

Let $X^*(\beta)$ denote the set of optimal solutions of the minimum CVaR problem $\text{MnCV}(\beta)$ and X^* the set of optimal solutions of the minimax problem MnMx. We write $d(x, \mathcal{D}) := \inf_{x' \in \mathcal{D}} \|x - x'\|$ for the distance from a point x to a set $\mathcal{D}$. For two compact sets $\mathcal{C}$ and $\mathcal{D}$,

$$\mathbb{D}(\mathcal{C}, \mathcal{D}) := \sup_{x \in \mathcal{C}} d(x, \mathcal{D})$$

denotes the deviation of $\mathcal{C}$ from $\mathcal{D}$ and $\mathbb{H}(\mathcal{C}, \mathcal{D}) := \max(\mathbb{D}(\mathcal{C}, \mathcal{D}), \mathbb{D}(\mathcal{D}, \mathcal{C}))$ denotes the Hausdorff distance between $\mathcal{C}$ and $\mathcal{D}$. We investigate the relationship between $X^*(\beta)$ and X^*, specifically, we estimate $\mathbb{D}(X^*(\beta), X^*)$. We do so by making use of a result from [23, Lemma 3.1].

Theorem 2 *Assume that $g(x, \xi)$ satisfies consistent tail condition* (4) *with parameters K and τ. Then*

$$\overline{\lim_{\beta \to 1}} X^*(\beta) \subset X^*, \tag{16}$$

where $\overline{\lim}$ denotes the outer limit of a set-valued mapping, i.e. the set of all cluster points of the sequence of the solution sets as $\beta \to 1$. Moreover, if the minimax

problem MnMx *satisfies the growth condition at* X^*, *i.e., there exists positive constants* $K_0, p > 0$ *such that*

$$\sup_{y \in Y} g(\hat{x}, y) \geq \min_{x \in X} \left(\sup_{y \in Y} g(x, y) \right) + K_0 d(\hat{x}, X^*)^p, \quad \text{for all } \hat{x} \in X, \quad (17)$$

then

$$\mathbb{D}(X^*(\beta), X^*) \leq \left[\frac{3}{K_0 K^{1/\tau}} \frac{\tau}{1 + \tau} (1 - \beta)^{1/\tau} \right]^{\frac{1}{p}}. \quad (18)$$

The proof to Theorem 2 is given in the "Appendix". The bound (18) gives us the minimal rate that $X^*(\beta)$ may approximate X^*. The tightness of the bound depends on problem. Here we give an example varied from Example 1 to compare the bound with the actual error.

Example 2 Let $g(x, \xi) = x^2 \xi - 2x$, where ξ is distributed with density $\rho(y) = 2 - 2y$ on the range $(0, 1)$. Similar to the calculations we made for Example 1, we have

$$F_x(\alpha) = 1 - \left(1 - \frac{\alpha + 2x}{x^2} \right)^2.$$

Since $g_x^* = x^2 - 2x$, we have

$$1 - F_x(\alpha) = \frac{1}{x^4} (x^2 - 2x - \alpha)^2 = \frac{1}{x^4} (g_x^* - \alpha)^2.$$

If we restrict x to take values from $X = [0, 1]$, then the inequality above implies

$$1 - F_x(\alpha) \geq (g_x^* - \alpha)^2,$$

so we can set parameters $K = 1$ and $\tau = 2$ in the consistent tail condition in Theorem 1 (iii). Moreover, looking at MnMx with $X = [0, 1]$, we can easily the unique optimal solution $x^* = 1$.

Let us now look into MnCV(β). Using the same approach as Example 1, we can derive

$$\text{VaR}_\beta(g_x(\xi)) = x^2 (1 - (1 - \beta)^{1/2}) - 2x,$$
$$\text{CVaR}_\beta(g_x(\xi)) = x^2 (1 - (2/3)(1 - \beta)^{1/2}) - 2x,$$

and hence

$$g_x^* - \text{CVaR}_\beta(g_x(\xi)) = (2x^2/3)(1 - \beta)^{1/2}.$$

Thus $\text{CVaR}_\beta(g_x(\xi))$ is minimized at

$$x^*(\beta) = \frac{1}{1 - (2/3)(1 - \beta)^{1/2}}.$$

Thus the actual error between $x^*(\beta)$ and x^* is

$$\|x^*(\beta) - x^*\| = \left| \frac{1}{1 - (2/3)(1 - \beta)^{1/2}} - 1 \right| = \frac{2}{3}(1 - \beta)^{1/2} + o((1 - \beta)^{1/2}).$$

We now turn to the bound in (18). To this end, we need to check the growth condition (17) at $x^* = 1$. It is easy to verify that the condition is satisfied with $K_0 = 1$ and $p = 2$. This gives

$$\left[\frac{3}{K_0 K^{1/\tau}} \frac{\tau}{1 + \tau} (1 - \beta)^{1/\tau} \right]^{\frac{1}{p}} = \sqrt{2}(1 - \beta)^{\frac{1}{4}}.$$

3.2 Approximation of MnCnCV(β) to MnCnMx

Now we turn to the constrained risk problem MnCnCV(β). Using the alternative representation of CVaR in (14), this problem can be written as

$$\begin{aligned} \min_{x \in X} \ & h(x) \\ \text{s.t.} \ & \min_{\eta \in \mathbb{R}} \ \Phi_\beta(x, \eta) \leq U \end{aligned} \tag{19}$$

with $U = 0$. Throughout this section, we assume that X is a convex compact set in $\mathbb{R}^n$, $g \colon \mathbb{R}^{n \times m} \to \mathbb{R}$ is a continuous function and Y is a closed subset of $\mathbb{R}^m$. Our interest is in looking at the difference between MnCnCV(β) and the robust feasibility problem MnCnMx. Thus we want to compare the solutions to the following semi-infinite convex system of inequalities:

$$g(x, y) \leq 0, \quad \text{for all } y \in Y, \tag{20}$$

where $x \in X$, and its CVaR approximation, $\text{CVaR}_\beta(g(x, \xi)) \leq 0$ which we can rewrite using the equivalence (14) as follows:

$$\min_{\eta \in \mathbb{R}} \left(\eta + \frac{1}{1 - \beta} \mathbb{E}[(g(x, \xi) - \eta)_+] \right) \leq 0. \tag{21}$$

We will discuss the difference between $\mathcal{G}$, which we define as the $x \in X$ satisfying (20), and $\mathcal{G}(\beta)$ which is defined as the $x \in X$ satisfying (21). These are the feasible sets over which the optimization of h takes place in the underlying problems MnCnCV(β) and MnCnMx. Clearly $\mathcal{G} \subset \mathcal{G}(\beta)$ and so $\mathcal{G}(\beta)$ provides an outer approximation of $\mathcal{G}$. We would like to know the excess of $\mathcal{G}(\beta)$ over $\mathcal{G}$ for $\beta \in (0, 1)$. The theorem below addresses this through the use of the Hausdorff distance $\mathbb{H}$ between the sets involved.

Problem ([20]) is said to satisfy the *Slater constraint qualification* if there exists a positive number $\bar{\delta}$ and a point $\bar{x} \in X$ such that

$$\max_{y \in Y} g(\bar{x}, y) \leq -\bar{\delta}. \tag{22}$$

Theorem 3 *Assume that $g(x, \xi)$ satisfies the consistent tail condition ([4]) with parameters K and τ.*

(i) *If $\mathcal{G}$ is non-empty, then for any $\varepsilon > 0$, there exists a $\beta_0 \in (0, 1) > 0$ such that when $\beta \in (\beta_0, 1)$,*

$$\mathbb{H}(\mathcal{G}(\beta), \mathcal{G}) \leq \varepsilon.$$

(ii) *If for each $y \in Y$, $g(\cdot, y)$ is convex on X, X is a convex set and ([20]) satisfies the Slater condition ([22]), then there exists a positive constant C such that for any $\beta \in (0, 1)$*

$$\mathbb{H}(\mathcal{G}(\beta), \mathcal{G}) \leq \frac{C}{K^{1/\tau}} \frac{\tau}{1 + \tau} (1 - \beta)^{1/\tau}.$$

The proof to the above results is given in the "Appendix". Note that the positive constant C may be estimated through Robinson's theorem in [24]. Indeed, if we let D be the diameter of $\mathcal{G}(\beta)$ and

$$\delta := -\min_{x \in \mathcal{G}} \mathrm{CVaR}_\beta(g(x, \xi)). \tag{23}$$

Then we can set $C := (\delta - \gamma)^{-1} D$, where γ is any positive number smaller than δ. Note also that since $\mathcal{G}$ is usually unknown, then the minimization in ([23]) may be taken over X. In the case when ([20]) satisfies the Slater constraint qualification, the δ value estimated from ([23]) is strictly positive for all $\beta > 0$. Moreover, since $\mathcal{G}(\beta) \subset X$, the diameter of $\mathcal{G}(\beta)$ is upper bounded by that of X. This means that we may choose a positive constant C which is independent of β.

The theorem says that the solution set of the CVaR system coincides with that of the semi-infinite system when β is driven to 1 and under the Slater constraint qualification, we may quantify the distance between the two solution sets.

Now we are ready to consider implications for the constrained risk problem MnCnCV(β). Note that MnCnCV(β) can be seen as a relaxation of the robust feasibility problem obtained by replacing the constraint $g(x, y) \leq U$, $\forall y \in Y$ with the constraint $g(x, y) \leq U + \Delta_\beta(x)$, $\forall y \in Y$ where $\Delta_\beta(x)$ is defined by ([9]). The next result looks at the way that solutions to the problem MnCnCV(β) approach those of MnCnMx as $\beta \to 1$.

Theorem 4 *Let $\hat{\vartheta}$ and $\hat{X}$ denote the optimal value and set of optimal solutions of MnCnMx. Let $\hat{\vartheta}(\beta)$ and $\hat{X}(\beta)$ denote the optimal value and set of optimal solutions of MnCnCV(β). Assume: (a) h is Lipschitz continuous with modulus L, (b) MnCnMx satisfies the Slater constraint qualification ([22]) and $g(x, \xi)$ satisfies the consistent tail condition ([4]) for parameters K and τ. Then*

(i) $\hat{\vartheta}(\beta)$ converges to $\hat{\vartheta}$ as $\beta \to 1$;
(ii) $\overline{\lim}_{\beta \to 1} \hat{X}(\beta) \subset \hat{X}$;
(iii) if for each $y \in Y$, $g(\cdot, y)$ is convex on X, X is a convex set and (20) satisfies the Slater condition (22), then

$$|\hat{\vartheta}(\beta) - \hat{\vartheta}| \leq \frac{LC}{K^{1/\tau}} \frac{\tau}{1 + \tau} (1 - \beta)^{1/\tau}, \tag{24}$$

for all β close to 1, where C is defined as in Theorem 3.

The proof of Theorem 4 is given in the "Appendix". Parts (i) and (ii) ensure consistency of optimal value and set of optimal solutions when β converges to 1 while Part (iii) quantifies the difference between the two optimal values.

3.3 Varying β in CVaR constrained problems

Up to this point we have been primarily concerned with the limiting behavior in our optimization problems as the confidence level β is pushed towards 1. Now we will change our perspective and look at the possibility of reducing the level of β in order to ensure that our estimates are based on a larger sample when we are using sample data. In the next section we will see how badly behaved sample-based CVaR estimates can be in the limit as β approaches 1, this provides motivation for applying a lower value of β to obtain a more stable estimate of CVaR and hence better performance of the optimization procedure.

We write $g_\beta(x) = \text{CVaR}_\beta(g(x, \xi))$ and $\overline{g}(x) = \sup_{\xi \in Y} g(x, \xi)$. We wish to solve the base problem MnCnMx:

$$\text{MnCnMx} : \begin{cases} \min_{x \in X} h(x) \\ \text{s.t. } \overline{g}(x) \leq 0. \end{cases}$$

We are interested in showing that an optimal solution x^* for our base problem MnCnMx is close to the optimal solution x_β^* for the approximating problem

$$\text{MnCnMx}(\beta) : \begin{cases} \min_{x \in X} h(x) \\ \text{s.t. } g_\beta(x) \leq U_\beta, \end{cases}$$

where $U_\beta = g_\beta(x^*)$. We will come back to discuss how U_β may be estimated in practice at the end of Sect. 4. The use of a different right hand side in the constraint distinguishes this problem from the problem MnCnCV(β) in which we set $U_\beta = 0$. Varying U_β will improve the approximation.

Let $\mathcal{F} = \{x \in X : \overline{g}(x) \leq 0\}$ be the feasible region for MnCnMx. We will establish bounds on the difference between the optimal solutions in the theorem below.

Theorem 5 Assume: (a) $\overline{g}(x) \leq 0$ satisfies the Slater condition (22), (b) $g(x, \xi)$ is convex in x for each fixed ξ, (c) h is globally Lipschitz continuous over $\mathcal{F}$ with modulus

L and satisfies the second order growth condition at x*, so that there is a positive number r such that

$$h(x) \geq h(x^*) + r\|x - x^*\|^2, \forall x \in \mathcal{F},$$

(d) there are positive numbers C and τ with

$$\sup_{x \in X} |g_\beta(x) - \overline{g}(x)| \leq C(1 - \beta)^{1/\tau}.$$

Then

$$\left\|x_\beta^* - x^*\right\| < \frac{2CD}{\overline{\delta}}(1 - \beta)^{1/\tau} + \sqrt{\frac{2LCD}{r\overline{\delta}}(1 - \beta)^{1/\tau}},$$

where D is the diameter of $\mathcal{F}$ and $\overline{\delta}$ is given as in (22).

The technical proof of the above result is given in the "Appendix". One limitation of this result is that the feasible set must be bounded in order for *D* to be finite, and even when the feasible set is bounded *D* may be large. In the next result we establish a tighter bound, without dependence on the size of the feasible set, in the special case that *h* and g_β are smooth enough to have well-defined derivatives at *x** and *ξ* is univariate.

Theorem 6 *Assume: (a) ξ is univariate, continuously distributed with density ρ, and there are positive numbers C and τ with*

$$\sup_{\xi \in Y} \xi - CVaR_\beta(\xi) \leq C(1 - \beta)^{1/\tau},$$

(b) g(x, ξ) is increasing in ξ differentiable in x with integrably bounded derivatives $\frac{\partial g}{\partial x_i}(x, \xi)$; (c) h, g_β and $\overline{g}(x)$ are convex functions in x, continuously differentiable at x*; (d) g_β is strongly convex at x*, that is,

$$g_\beta(x) \geq g_\beta(x^*) + (x - x^*)^T \nabla g_\beta(x^*) + K_c \left\|x - x^*\right\|^2, \tag{25}$$

for some positive constant K_c, (e) there are unique optimal solutions with binding constraints in both problems MnCnMx(β) *and* MnCnMx, *(f) the mixed derivatives $\frac{\partial^2 g}{\partial x_i \partial \xi}(x^*, \xi)$ exist and are bounded. Then*

$$\left\|x_\beta^* - x^*\right\| \leq \frac{C\sqrt{n}}{K_c} \max_{i=1,\ldots,n} \sup_{\xi \in Y} \left|\frac{\partial^2 g}{\partial x_i \partial \xi}(x^*, \xi)\right| (1 - \beta)^{1/\tau}. \tag{26}$$

The proof of Theorem 6 is given in the "Appendix". We provide an example to explain the theorem.

Example 3 Consider the case that $g(x, \xi) = \frac{\xi}{1+x_1} + x_2^2 - 1$ with ξ being uniformly distributed over $Y = [0, 1]$ and $X = [0, \infty) \times [0, 1]$. Let $h(x) = x_1^2 + x_2^2$. By definition

$$g_\beta(x) = \text{CVaR}_\beta(g(x, \xi)) = \frac{1 + \beta}{2(1 + x_1)} + x_2^2 - 1$$

and $\bar{g}(x) = \frac{1}{1+x_1} + x_2^2 - 1$. Thus problem *MnCnMx* can be written as

$$\text{MnCnMx} : \begin{cases} \min_{x \in X} x_1^2 + x_2^2 \\ \text{s.t.} \ \frac{1}{1+x_1} + x_2^2 \leq 1. \end{cases}$$

Obviously the feasible set $\mathcal{F}$ of *MnCnMx* is unbounded and the optimal solution to $x^* = (0, 0)_T$ is located at the boundary of $\mathcal{F}$. Consequently we can formulate MnCnMx as

$$\text{MnCnMx}(\beta) : \begin{cases} \min_{x \in X} x_1^2 + x_2^2 \\ \text{s.t.} \ \frac{1+\beta}{2(1+x_1)} + x_2^2 \leq \frac{1}{2}(1 + \beta). \end{cases}$$

The optimal solution $x_\beta^* = (0, 0)^\top$ and the constraint is active at x_β^*. In what follows, we verify the conditions of Theorem 6. Observe first that $\sup_{\xi \in Y} \xi = 1$, $\text{CVaR}_\beta(\xi) = (1 + \beta)/2$ and

$$\sup_{\xi \in Y} \xi - \text{CVaR}_\beta(g(x, \xi)) = \frac{1}{2}(1 - \beta).$$

Thus condition (a) is satisfied with $C = \frac{1}{2}$ and $\tau = 1$. Moreover, it is easy to see that g is monotonically increasing in ξ and $\nabla_x g(x, \xi) = (-\xi/(1 + x_1)^2, 2x_2)^\top$ which is integrably bounded, this verifies condition (b). Conditions (c) and (e) are obvious. Condition (d) can also be verified with $K_c = \frac{1}{8}$ as $g_\beta(x)$ is strongly convex at x_β^*. Finally condition (f) is satisfied with $\frac{\partial^2 g}{\partial x_1 \partial \xi}(x^*, \xi) = -\frac{1}{(1+x_1^*)^2} = -1$ and $\frac{\partial^2 g}{\partial x_2 \partial \xi}(x^*, \xi) = 0$. With all conditions being verified, we obtain from Theorem 6 that

$$\left\| x_\beta^* - x^* \right\| = 0 < \frac{C\sqrt{n}}{K_c} \max_{i=1,\dots,n} \sup_{\xi \in Y} \left| \frac{\partial^2 g}{\partial x_i \partial \xi}(x^*, \xi) \right| (1 - \beta)^{1/\tau} = \frac{4\sqrt{2}}{1 + \beta}(1 - \beta).$$

4 Convergence with sample-based methods

We now return to our discussion of numerical methods for solving problem MnCnCV(β). Note that (19) can be reformulated by combining the two minimizations:

$$\min_{x \in X, \eta \in \mathbb{R}} h(x)$$
$$\text{s.t.} \quad \eta + \frac{1}{1-\beta} \mathbb{E}[(g(x, \xi) - \eta)_+] \le 0. \tag{27}$$

This reformulation is well-known, see Rockafellar and Uryasev [21, Theorem 16].

Problem (27) is a nonlinear stochastic optimization problem with deterministic objective and a stochastic constraint. The main challenge here is to handle the expected value $\mathbb{E}[(g(x, \xi) - \eta)_+]$. As we discussed in the introduction the usual approach is based on a sample from ξ. Let $\xi^1, \ldots, \xi^N$ be an independent and identically distributed (i.i.d.) sample of ξ. We consider the following sample average approximation (SAA) for $\mathrm{MnCnCV}(\beta)$ [using formulation (27)]:

$$\mathrm{MnCnCVSA}(\beta) : \begin{cases} \min_{x, \eta} h(x) \\ \text{s.t. } \eta + \frac{1}{(1-\beta)N} \sum_{j=1}^{N} (g(x, \xi^j) - \eta)_+ \le 0, \\ x \in X, \eta \in \mathbb{R}. \end{cases}$$

For a fixed sample, $\mathrm{MnCnCVSA}(\beta)$ is a deterministic nonlinear programming problem and can be easily solved by standard methods for nonlinear programming.

At this point, it is helpful to link our SAA approach to Calafiore and Campi's randomization approach in [8,12] in which the following problem is taken as a sample-based approximation to MnCnMx:

$$\mathrm{MnCnMxSA} : \begin{cases} \min_{x \in X} h(x) \\ \text{s.t. } g(x, \xi^j) \le 0, \text{ for } j = 1, \ldots, N, \end{cases}$$

where $\xi^1, \ldots, \xi^N$ are randomly taken from set Y. A clear benefit of the randomization approach is to replace the continuum of the constraints of the problem MnCnMx with a finite number of constraints. In doing so, we obtain an outer approximation to the feasible set of the true problem, and hence the optimal value of MnCnMxSA gives rise to a lower bound for the optimal value of MnCnMx. From a practical point of view, an important issue when using this kind of approximation scheme concerns feasibility of the optimal solution of MnCnMxSA to its true counterpart. It has been shown that the solution satisfies most of the constraints of MnCnMx if the number of points N is sufficiently large, see [8] and [14]. Moreover, [14] demonstrated an exact universal bound on the number of samples required to ensure a particular value for the probability that the constraints of the original problem are violated by more than ϵ.

Analogous to MnCnMxSA, our approximation scheme $\mathrm{MnCnCVSA}(\beta)$ also gives an outer approximation to MnCnMx in terms of the feasible set and provides a lower bound for the optimal value provided the sample size is sufficiently large. However the two formulations are quite different: we note that there is a single non-smooth convex constraint in $\mathrm{MnCnCVSA}(\beta)$ whereas MnCnMxSA has N smooth constraints.

Let (x^N, η^N) be an optimal solution which is obtained from solving the sample average approximation $\mathrm{MnCnCVSA}(\beta)$ with sample size N. In the next result we

estimate the probability of x^N which is based on a sample, deviating by more than an amount ε from the real optimal solution for the constrained risk problem MnCnCV(β). This result does not require $g(x, \xi)$ to be bounded. Then in part (ii) of the Theorem we use the consistent tail behavior condition to establish the equivalent result where we compare the sample average approximation MnCnCVSA(β) with the optimal solution to the robust problem MnCnMx as β approaches 1.

Theorem 7 *As before let $\hat{X}$ and $\hat{X}(\beta)$ denote the set of optimal solutions of MnCnMx and MnCnCV(β). Assume: (a) the feasible set of problem MnCnMx is non-empty; (b) $g(x, \xi)$ is convex as a function of x; (c) we may choose a measurable function $\theta(\xi)$ as the Lipschitz constant for g, so that*

$$|g(x', \xi) - g(x, \xi)| \le \theta(\xi)\|x' - x\|$$

for all $\xi \in Y$ and all $x', x \in X$, and the moment generating function of $\theta(\xi)$ (as a function of t) is finite for t in a neighborhood of zero; (d) for every x the moment generating function for $g(x, \xi)$, is finite for t in a neighborhood of zero; (e) $h(x)$ is Lipschitz continuous on X with modulus L and it satisfies some growth condition on $\mathcal{G}(\beta)$, that is, there exists $\delta_0 > 0$ such that

$$R(\delta) := \inf_{x \in \mathcal{G}(\beta), d(x, \hat{X}(\beta)) \ge \delta} h(x) - \hat{\vartheta}(\beta) > 0 \tag{28}$$

for any $\delta \in (0, \delta_0]$. Then

(i) *for any positive number ε, there exist positive constants $C(\varepsilon)$ and $\alpha(\varepsilon)$ (independent of N) such that*

$$\Pr(d(x^N, \hat{X}(\beta)) \ge \varepsilon) \le C(\varepsilon)e^{-\alpha(\varepsilon)N}$$

for N sufficiently large, where x^N is an optimal solution to MnCnCVSA(β).

(ii) *If, in addition, problem MnCnMx satisfies the Slater constraint qualification (22), and $g(x, \xi)$ has consistent tail behavior on X, then for any $\varepsilon > 0$ there exist positive constants β_0, $C(\varepsilon)$ and $\alpha(\varepsilon)$ (independent of N) such that for $\beta > \beta_0$*

$$\Pr(d(x^N, \hat{X}) \ge \varepsilon) \le C(\varepsilon)e^{-\alpha(\varepsilon)N}$$

for N sufficiently large.

The proof of the above technical results is given in the "Appendix".

It might be helpful to make some comments about the conditions. Conditions (c) and (d) involving moment generating functions simply mean that the probability distributions for the associated random variables $\theta(\xi)$ and $g(x, \xi)$ die exponentially fast in the tails. In particular, they will be satisfied when the random variable has a distribution supported on a bounded subset of $\mathbb{R}$. Condition (c) requires f to be Lipschitz continuous in x, but this is implied by the convexity of $g(x, \xi)$ as a function of x. These conditions are standard for deriving exponential rate of convergence, see for example

[25]. Condition (e) requires the objective function h to satisfy certain growth condition when x deviates from solution set $\hat{X}(\beta)$. This is implied by similar growth condition when x deviates from $\hat{X}$. Growth conditions are often needed to derive stability of optimal solutions, see for instance [26].

The theorem says that x^N converges to an optimal solution of MnCnMx in distribution and it does so at an exponential rate as the sample size increases. The proof exploits the uniform law of large numbers for random functions. Note that x^N is not necessarily a feasible solution of MnCnMx but it approaches feasibility as $\beta \to 1$.

This result allows us to use a sample-based CVaR constrained problem MnCnCVSA(β) to approximate the optimal solution to the robust problem MnCnMx. In doing this, the accuracy of the result is improved by taking β close to 1, but the sample based approximation then becomes worse because the estimate of CVaR is based on a smaller number of observations. There is an alternative approach making use of the approximations of MnCnMx(β) discussed in Theorems 5 and 6. This involves an appropriate adjustment of the constraint right hand side to $g_\beta(x^*)$, and hence requires an initial estimate of the optimal solution x^*.

Thus we are led to a two stage procedure to solve MnCnMx.

(a) Generate a sample ξ^j, $j = 1, 2, \ldots N$, and find the optimal solution x_0^* to the sample average robust problem MnCnMxSA.
(b) For a given value of β use ξ^j, $j = 1, 2, \ldots N$ to estimate $U_0 = \text{CVaR}_\beta(g(x_0^*, \xi))$ and solve the sample-based CVaR problem MnCnMxSA(β) with U_0 as right hand side to give an optimal solution x_1^*.

In the next section, we will discuss the application of this approach to a particular problem arising in project management and show that x_1^* can be a better solution than x_0^*. The second stage does not require more samples to be taken, but it makes more use of the information contained in these samples.

In assessing whether the second stage is worthwhile we need to consider two factors. First, note that neither of x_0^* or x_1^* is guaranteed to be feasible for the original problem, MnCnMx, since both are based on sample approximations. Ideally the two stage procedure will not reduce the probability of feasibility. The second aspect is the quality of the solution as measured by the objective function $h(x)$, or by the distance to the true solution. We aim to achieve a solution which is close to the exact optimum and no worse in terms of the objective.

The use of a CVaR approximation with a value of β that is not close to 1, will enable the solution to be calculated based on a greater proportion of the sample set, and this gives an opportunity to improve the solution in moving from x_0^* to x_1^*.

4.1 Asymptotic behavior of sample average approximated CVaR

It is well known that estimating CVaR from sample data is subject to high volatility. In general understanding the tail behavior of a distribution on the basis of a sample is very challenging (and require large sample sizes). Work by Heyde and Kou [27] shows that even determining whether the underlying distribution has power or exponential type tails can be very difficult. In this section we will characterize the variance of

the CVaR estimates in more detail. Since these estimates are embedded within our optimization procedure, we can expect to see poor performance of the sample based optimization procedure as $\beta \to 1$. In fact Lim et al. [28] show that the optimization procedure introduces an extra bias when applying this approach to a portfolio optimization problem. In this section we will widen our analysis to include cases where the values of $g(x, \xi)$ are unbounded (and so cannot have the property of consistent tail behavior).

Extreme value theory and the Picklands-Balkema-de Haan (PBdH) theorem [29] show that, if a distribution for a random variable ξ is in a set MDA(H_κ) (the maximum domain of attraction for a generalized extreme value distribution with parameter κ), then the excess distribution over u, defined by

$$F_{>u}(t) = \Pr(\xi - u \le t \mid \xi > u) \text{ for } t > 0, \tag{29}$$

has a generalized Pareto distribution in the limit of large u with shape parameter κ. In other words there are parameters κ and $\gamma > 0$ such that the cumulative distribution function for the excess distribution above u approaches $G_{\kappa, \gamma}(t)$ given by

$$G_{\kappa, \gamma}(t) = \begin{cases} 1 - (1 + \kappa t/\gamma)^{-1/\kappa} & \text{for } \kappa \ne 0, \\ 1 - \exp(-t/\gamma) & \text{for } \kappa = 0. \end{cases} \tag{30}$$

When $\kappa \ge 0$ the distribution is defined over the range $t \ge 0$. The $\kappa < 0$ case corresponds to a distribution for F with a maximum value $t_{\max} = -\gamma/\kappa$.

To apply these ideas in our setting we consider the tail behavior of $g(x, \xi(\omega))$ which has a cumulative distribution function $F_x(\cdot)$. The condition that F is in some set MDA(H_κ) is hardly a restriction, since it applies to all distributions we are ever likely to observe in practice. The PBdH theorem in effect distinguishes between three cases for the tail behavior of $g(x, \xi)$: either there is a finite maximum to the distribution ($\kappa < 0$), or the tail index is infinity (like a normal distribution) ($\kappa = 0$), or there are fat tails so that the tail index is less than infinity ($\kappa > 0$).

We can use this theory and a result of Chen [7] (see also Brazauskas et al. [30]) to establish the following result for a sample based estimate of CVaR. The standard CVaR estimator based on a sample, $\xi^1, \xi^2, \ldots, \xi^N$ is given by

$$\widehat{\text{CVaR}}_\beta = \widehat{v}_\beta + \frac{1}{N - [N\beta]} \sum_{j=1}^{N} (\xi^j - \widehat{v}_\beta)_+$$

where we write $[N\beta]$ for the integer part of $N\beta$ and $\widehat{v}_\beta$ is the sample estimator of VaR$_\beta(\xi)$ (thus $\widehat{v}_\beta = \xi^{([N\beta]+1)}$ where $\xi^{(r)}$ is the r'th order statistic of $\{\xi^i\}_{i=1}^N$). As we mentioned earlier there are other more complex estimates available (e.g. [31]) avoiding the problem of bias [7]. Other estimates are helpful when $N\beta$ is far from an integer, but our asymptotic result does not require this since we will consider $N \to \infty$.

Proposition 2 *Suppose that a CVaR estimate is made from independent samples* $g(x, \xi^1), g(x, \xi^2), \ldots, g(x, \xi^N)$ *using the estimator* $\widehat{\text{CVaR}}_\beta$ *and* $g(x, \xi)$ *has an*

MDA(H_κ) distribution. Then the asymptotic variance of the estimator as $N \to \infty$ is approximated by

$$\gamma \left(\frac{1}{(1 - 2\kappa)} + \beta \right) \frac{(1 - \beta)^{-2\kappa - 1}}{N}$$

for some constant γ, and is exactly this expression if $g(x, \xi)$ follows a generalized Pareto distribution (with shape parameter $\kappa < 0.5$) for values above $VaR_\beta(g(x, \xi))$.

The proof of Proposition 2 is given in the "Appendix". We have formulated this result in a way that clarifies the β dependence of the variance of the CVaR estimator. This is less clear cut in the similar results obtained by Necir et al. [32] (see also [31]). Notice that the dependence on β is different when $\kappa \leq -0.5$. This corresponds to distributions with a finite maximum g_x^* where, for α close enough to g_x^*, $F_x(\alpha) = 1 - k_0(g_x^* - \alpha)^\theta$ for some choice of scaling k_0 and $\theta = -1/\kappa < 2$. We can say that the density approaches zero at its upper bound faster than linearly. In this case the variance of the estimator tends to decrease as $\beta \to 1$, but for values of $\kappa > -0.5$ the variance will increase.

5 Applications

5.1 An application to project management

To illustrate the theoretical results established in Theorems 1 and 2, we apply them to a resource allocation problem occurring in project management. One reason for considering this particular application is that the essential supremum of the underlying functions is bounded which enables us to apply the error bound in Theorem 1 (iii).

Suppose that a project is completed when the last of K different activities are completed, with a requirement to complete these sequentially. The distribution of time required for an activity is assumed to be known, but the assignment of additional resources by working overtime can reduce the total days needed. [33] and [34] give some discussion of different models for time reduction in a stochastic environment and how resources can be optimized in this context.

It is desired to minimize expected completion time subject to a CVaR risk constraint. We suppose that there are K activities and the elapsed time for the k'th activity is given by $\xi_k/(1 + x_k)$ for $k = 1, 2, \ldots, K$ where ξ_k is a random variable giving the total time required for this activity and x_k is the proportion of overtime worked on this activity. The cost of additional overtime for activity k is $c_k x_k$ and there is a maximum budget of B available for overtime. Hence the feasible allocations for x are given by

$$x \in X = \left\{ (x_1, x_2, \ldots x_K) \mid x_i \geq 0, \sum_{k=1}^{K} c_k x_k \leq B \right\}.$$

5.1.1 MnMx and MnCV(β)

The problem of minimizing the CVaR for project completion becomes

$$\min_{x \in X} \text{ CVaR}_\beta \left(\sum_{k=1}^{K} \frac{\xi_k}{(1+x_k)} \right).$$

In practice, the activity durations ξ_k are often correlated as they may be affected by the same external factors such as the weather. The durations of ξ_i often follow a beta distribution. The beta distribution has often been used as a bounded distribution alternative to the normal distribution. The beta distribution has been proposed as a good candidate for modeling bounded continuous random variables using only the information provided by an expert not only in project management but also in different fields, such as resource assessment, construction duration, and engineering, among others [35]. When the parameters of the beta distribution are not easily estimated, it has been suggested that the triangular distribution can be used as a proxy for the beta distribution [36].

For the simplicity of discussion, here we concentrate on the case where each ξ_k is independent and has a distribution with a triangular density on (a_k, b_k) with mode at $m_k \in (a_k, b_k)$.

Let $g(x, \xi) = \sum_{k=1}^{K} \xi_k/(1 + x_k)$. Then $g_x^* = \sum_{k=1}^{K} b_k/(1 + x_k)$. It is easy to see that

$$
\begin{aligned}
1 - F_x(\alpha) &= \Pr\left(\sum_{k=1}^{K} \frac{\xi_k}{1 + x_k} > \alpha \right) \\
&= \Pr\left(\sum_{k=1}^{K} \frac{\xi_k - b_k}{1 + x_k} > \alpha - \sum_{k=1}^{K} \frac{b_k}{1 + x_k} \right) \\
&\geq \prod_{k=1}^{K} \Pr\left(\frac{\xi_k - b_k}{1 + x_k} > \frac{\alpha}{K} - \sum_{k=1}^{K} \frac{b_k}{K(1 + x_k)} \right) \\
&= \prod_{k=1}^{K} \Pr\left(\frac{\xi_k}{(1 + x_k)} > \frac{b_k}{(1 + x_k)} - \frac{g_x^* - \alpha}{K} \right),
\end{aligned}
$$

the last term gives the probability that each of the K components of g comes within $\left(g_x^* - \alpha \right)/K$ of its maximum value. Now for a triangular distribution we have, for $\delta_k < b_k - m_k$,

$$\Pr(\xi_k > b_k - \delta) = \frac{\delta^2}{(b_k - a_k)(b_k - m_k)}.$$

So, for α large enough,

$$1 - F_x(\alpha) \geq \left(g_x^* - \alpha\right)^{2K} \frac{1}{K^{2K}} \prod_{k=1}^{K} \frac{(1 + x_k)^2}{(b_k - a_k)(b_k - m_k)}$$

$$> \left(g_x^* - \alpha\right)^{2K} \frac{1}{K^{2K}} \prod_{k=1}^{K} \frac{1}{(b_k - a_k)(b_k - m_k)}.$$

This establishes the inequality of Theorem 1 part (iii) without dependence on x, so we have consistent tail behavior. We can also check the conditions of Proposition 1 to establish the same result.

5.1.2 MnCnMx and MnCnCV (β)

We now turn to consider MnCnCV. Given activity durations that are triangular, the expected durations are $\mu_k = (a_k + b_k + m_k)/3$. Hence the constrained risk problem MnCnCV(β) becomes

$$\min_{x \in X} \sum_{k=1}^{K} \frac{\mu_k}{(1 + x_k)}$$
$$\text{s.t.} \ \ \mathrm{CVaR}_\beta \left(\sum_{k=1}^{K} \frac{\xi_k}{(1 + x_k)} \right) \leq U,$$

where $X = \{(x_1, x_2, \ldots x_K) \mid x_i \geq 0, \sum_{k=1}^{K} c_k x_k \leq B\}$. Writing $g(x, \xi) = \sum_{k=1}^{K} \xi_k/(1 + x_k) - U$ we reach the form we require.

The robust feasibility problem MnCnMx requires a guarantee that the completion time does not exceed U, and replaces the constraint in this problem with

$$\sum_{k=1}^{K} \frac{b_k}{1 + x_k} - U \leq 0. \tag{31}$$

Thus $\mathcal{G} = \{x \in X : \sum_{k=1}^{K} b_k/(1 + x_k) \leq U\}$. In this case, provided B and U are chosen so that both the budget and risk constraint are tight, we can find the optimal solution $\hat{x}$ with

$$\hat{x}_k = \sqrt{\frac{\mu_k + \lambda_2 b_k}{\lambda_1 c_k}} - 1, \quad \text{for } k = 1, \ldots, K,$$

where λ_1 and λ_2 are the simultaneous solutions of

$$\sqrt{\lambda_1} \left(B + \sum_{k=1}^{K} c_k \right) = \sum_{k=1}^{K} \sqrt{c_k(\mu_k + \lambda_2 b_k)},$$

$$\sqrt{\lambda_1} \sum_{k=1}^{K} b_k \sqrt{\frac{c_k}{\mu_k + \lambda_2 b_k}} = U.$$

The constrained risk problem MnCnCV(β) involves the CVaR level in the constraint and cannot be easily solved explicitly. However when K, the number of activities, is large, use of the central limit theorem allows us to approximate the term $g(x, \xi) = \sum_{k=1}^{K} \xi_k/(1 + x_k)$ with a normal distribution, even though it continues to have a bounded range. Note that the variance of ξ_k is

$$V_k = \left(a_k^2 + b_k^2 + m_k^2 - a_k b_k - a_k m_k - b_k m_k \right)/18,$$

so $g(x, \xi)$ has mean $\mu(x) = \sum_{k=1}^{K} \mu_k/(1 + x_k)$ and standard deviation $\sigma(x) = \left(\sum_{k=1}^{K} V_k/(1 + x_k)^2 \right)^{1/2}$. Since $g(x, \xi)$ is approximately normal we can estimate CVaR$_\beta$ from the mean and standard deviation $\mu(x) + c_2(\beta)\sigma(x)$ (see [17] for an expression for the constant $c_2(\beta)$).

We have already shown the consistent tail behavior of $g(x, \xi)$; we also observe that the objective function $h(x) = \sum_{k=1}^{K} \mu_k/(1 + x_k)$ is Lipschitz; and the Slater condition for the constraint is easily checked for specific examples. Hence Theorem 4 applies to the convergence behavior as $\beta \to 1$, and the value of the optimal solution to the constrained risk problem MnCnCV(β) approaches the value for the robust problem MnCnMx with constraint (31).

5.1.3 A numerical example

Now we wish to check numerically the performance of the two-stage procedure for a sample based approximation to MnCnMx. We generate a set of 10,000 project management problems with $K = 10$ and $a_k, b_k, m_k, k = 1, 2, \ldots 10$ being generated by taking m_k uniform on [9, 15] and $a_k = \gamma_k^a m_k$ with γ_k^a uniform in [0.85, 0.95] and $b_k = \gamma_k^b m_k$ with γ_k^b uniform in [1.10, 1.20]. Moreover for each problem we generate a set of costs c_k uniform from [1, 5], and set $U = 20$ and $B = 130$.

An example of a problem generated in this way has the values of parameters a, b, m and c as given below

a	11.33	8.79	9.78	10.19	13.51	9.95	8.88	9.71	11.61	8.07
b	14.81	11.07	12.78	12.59	16.83	12.35	11.72	12.50	15.06	10.39
m	13.10	9.87	11.20	10.85	14.46	10.58	9.92	10.65	13.59	9.05
c	1.97	3.72	1.40	3.16	2.72	1.66	2.87	4.09	2.71	1.87

To solve the project management problem we suppose that we have access to a sample of 100 data points, ξ_k^j, $k = 1, 2, \ldots 10$, $j = 1, 2, \ldots 100$ generated randomly from the appropriate triangular distributions. We estimate b_k and μ_k with

$$\widehat{b}_k = \max_{j=1,\ldots,100} \xi_k^j, \quad \widehat{\mu}_k = \frac{1}{100} \sum_{j=1}^{100} \xi_k^j.$$

Then the sample based robust optimization problem MnCnMxSA becomes

$$\min_{x \in X} \sum_{k=1}^{K} \frac{\widehat{\mu}_k}{(1 + x_k)}$$

$$\text{s.t.} \sum_{k=1}^{K} \frac{\widehat{b}_k}{1 + x_k} \leq U$$

where $X = \{(x_1, x_2, \ldots x_K) \mid x_i \geq 0, \sum_{k=1}^{K} c_k x_k \leq B\}$. We write x_0^* for the optimal solution of this problem (and x^* for the actual optimal solution based on b_k and μ_k).

We can estimate the probability of a new sample taken from the original distribution failing to be feasible for x_0^*, i.e. we find

$$Z_0 = \Pr\left(\sum_{k=1}^{K} \xi_k/(1 + x_{0k}^*) > U\right).$$

This will be very small and thus hard to estimate using a sample based approach. For a failure of feasibility each of the ξ_k values will need to be close to its upper limit. Observe that since $\xi_k < b_k$, $k = 1, 2, \ldots K$, the inequality $\sum_{k=1}^{K} \xi_k/(1 + x_{0k}^*) > U$ implies

$$\xi_j > (1 + x_{0j}^*)\left(U - \sum_{k \neq j} \xi_k/(1 + x_{0k}^*)\right) > (1 + x_{0j}^*)\left(U - \sum_{k \neq j} b_k/(1 + x_{0k}^*)\right) = q_j.$$

Provided $q_j > m_j$, we will have $\Pr(\xi_j > q_j) = \frac{(b_j - q_j)^2}{(b_j - m_j)(b_j - a_j)}$. So we can estimate Z_0 by choosing each ξ_k from the triangular distribution truncated below q_k with density $f(x) = 2(b_k - x)/(b_k - q_k)^2$ and then scale by dividing by $\Pi_{k=1}^{K} \Pr(\xi_k > q_k)$. This is estimated from a new sample of size 10,000 ξ vectors.

For each problem we take the optimal solution x_0^* and generate a new optimal solution for MnCnCVSA(β) with the right hand side of the constraint replaced by the estimated CVaR value at x_0^*, $\mathrm{CVaR}_\beta(g(x_0^*, \xi))$. Thus the constraint becomes

$$\mathrm{CVaR}_\beta\left\{\sum_{k=1}^{K} \frac{\xi_k^j}{(1 + x_k)}, \ j = 1, 2 \ldots N\right\} \leq \mathrm{CVaR}_\beta\left\{\sum_{k=1}^{K} \frac{\xi_k^j}{(1 + x_{0k}^*)}, \ j = 1, 2 \ldots, N\right\}.$$

The solution is x_1^*.

We make three tests to see whether the solution x_1^* is better than x_0^*. Ideally we will achieve three properties. First the solution obtained will have improved feasibility, so that the probability of a sample in which the project is not completed by time U is reduced. In other words

$$Z_1 = \Pr\left(\sum_{k=1}^{K} \xi_k / (1 + x_{1k}^*) > U\right) < Z_0.$$

Second the optimization problem will have an improved objective value; that is $\sum_{k=1}^{K} E(\xi_k)/(1 + x_{1k}^*) < \sum_{k=1}^{K} E(\xi_k)/(1 + x_{0k}^*)$. Finally the solution obtained will be closer to the correct value $\|x^* - x_1^*\| < \|x^* - x_0^*\|$ where x^* is the optimal solution for our base problem MnCnCX.

We estimate Z_1 in the same way as Z_0 by generating another test sample of size 10,000 to see how many of this sample are infeasible. Figure 1 shows the comparison between Z_0 and Z_1 for the complete set of 10,000 sample problems and different values of β.

Figure 2 shows the values of objectives

$$\sum_{k=1}^{K} \mathbb{E}(\xi_k)/(1 + x_{0k}^*) \quad \text{and} \quad \sum_{k=1}^{K} \mathbb{E}(\xi_k)/(1 + x_{1k}^*).$$

We can observe that in the case that x_1^* is better it will achieve a lower objective values than x_0^*.

Finally the third set of results is about the precision of the approximation solutions, where we compare $\|x^* - x_1^*\|$ and $\|x^* - x_0^*\|$. In order to find an exact value for x^* we use a sample based approach but increase the size of the sample to 50,000.

We observe that there is a clear improvement from using x_1^* instead of x_0^* for each of the three criteria as shown in Figs. 1, 2 and 3. The improvement in feasibility is achieved when moving β down to a value of 0.94 and is maintained when β is reduced further to 0.9. The values of the objective function and the optimizing decision vector x, shown in Figs. 2 and 3, show improvement as β is reduced all the way down to $\beta = 0.9$.
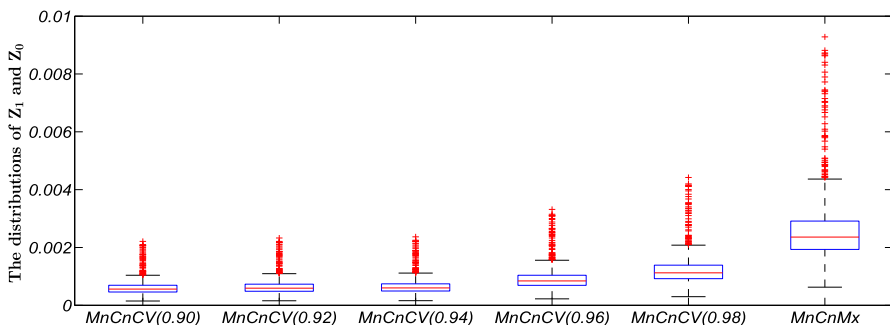

Fig. 1 The improvement of feasibility: Z_1 for out-of-sample tests at x_1^* in MnCnCV(β) and Z_0 for out-of-sample tests at x_0^* in MnCnMx
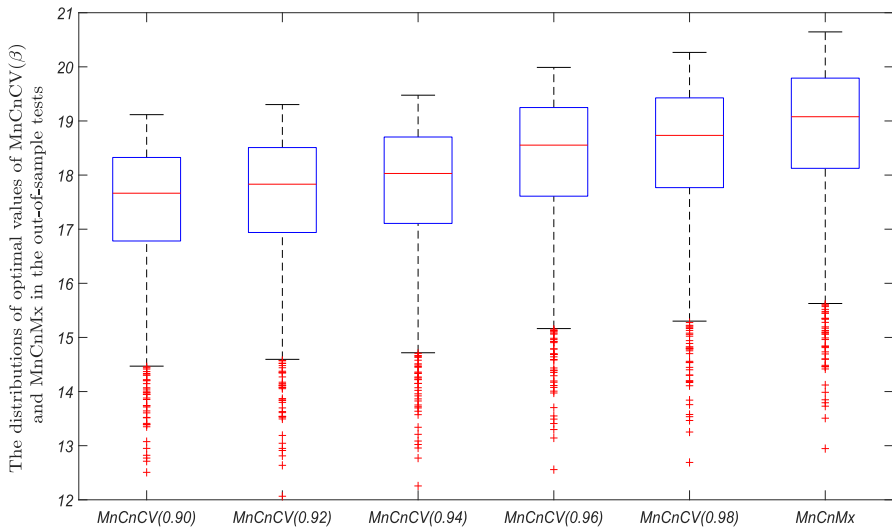
Fig. 2 The improvement in the objective values: the averaged objective value of out-of-sample tests at x_1^* in MnCnCV(β) and at x_0^* in MnCnMx
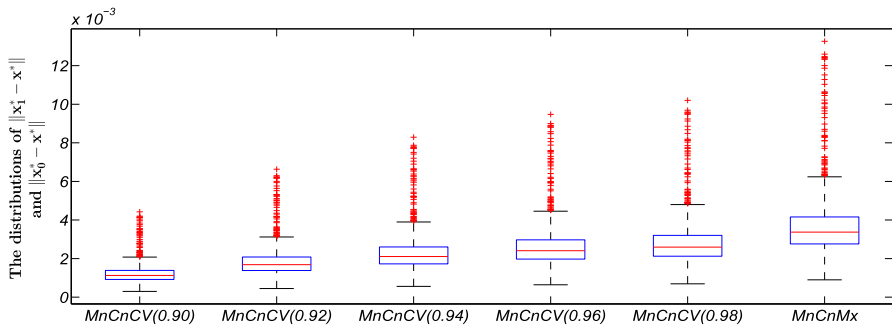

Fig. 3 The distance of x_1^* in MnCnCv(β) and x_0^* in MnCnMx to true solution x^*

5.2 An application to portfolio optimization

Portfolio optimization is a natural area of application for our ideas and has been extensively studied. Originally portfolio optimization was most commonly formulated using a Markowitz approach concentrating on the mean-variance efficient frontier. More recently it is common to consider either mean-VaR or mean-CVaR problems. The idea behind a mean-CVaR approach is simple: we want to find a portfolio of assets which achieves a good result both in terms of mean return and also in terms of risk as measured by CVaR. A portfolio with more risk will allow better mean returns, and there will be a Pareto optimal solution that minimizes risk for a given level of return. We will consider the equivalent problem MnCnCV(β) where we maximize return subject to a constraint that CVaR risk is bounded by U. Solving the problem for different values of U would map the efficient frontier.

There is a substantial literature dealing with both VaR and CVaR versions of portfolio optimization, but we will mention only papers that bear directly on the problem of high volatility of CVaR estimates discussed by Lim et al. [28] and others. Some authors (Zhu and Fukushima [37], Wozabal [38]) have proposed using a robust formulation of the problem to deal with the uncertainty in the underlying distribution. An alternative approach proposed by El Karoui et al. [39] is to use a regularization method. These are relatively complex methods, and we will not attempt to make comparisons with our proposal.

5.2.1 Varying β in constrained risk problems

Recall the analysis on varying β in Sect. 3.3. Our aim in this section is to approximate the solution of a constrained risk problem MnCnCV(β_0) by changing to a new (lower) confidence level β_1. However, to achieve a good approximation we will need to change the value of the constraint right-hand side at the same time from U_0 to U_1. The amount by which we should change U is determined by the characteristics of the tail behavior of the distribution of $g(x, \xi)$. We will not require bounds on the values of $g(x, \xi)$ for this procedure.

In the project management example of Sect. 5.1 we were looking for ways to solve the base problem MnCnMx through the MnCnCV approximation. Here our target problem is already in the form MnCnCV(β) and MnCnMx does not appear. Nevertheless we can still use the underlying idea of adjusting the confidence level downwards to make use of more of the sample. Instead of starting with x_0^* as an estimate of the solution of MnCnMx, we begin with x_0^* as an estimate of the solution of MnCnCV(β).

Thus we are led to a two-stage procedure. We suppose that there is a given confidence level β_0 and the decision maker is concerned to bound the risk CVaR$_{\beta_0}$ to less than U_0. The process starts by finding an estimate of the optimal solution x_0^* for the problem MnCnCV(β_0) with U_0; then having done this we calculate $U_1 = \text{CVaR}_{\beta_1} g(x_0^*, \xi)$ and finally solve the problem MnCnCV(β_1) with U_1 to as the constraint right hand side to produce a new optimal solution x_1^*. Because we will use the sample average approximation for each of the solutions we will expect some difference between x_0^* and x_1^* even though the sample $\xi^1, \ldots, \xi^N$ is the same for both problems.

We suppose that we have a portfolio problem in which there are n stocks and write w_i for the weight on stock i. Thus the version of MnCnCV(β) we wish to solve is

$$
\min_{\mathbf{w} \in W_n} \mathbb{E}\left[-\sum_{i=1}^n w_i R_i \right]
$$
$$
\text{s.t.} \quad \text{CVaR}_\beta \left(-\sum_{i=1}^n w_i R_i \right) \le U. \tag{32}
$$

Here R_i is a random variable giving the return of stock i and the set of feasible weights is defined through lower bounds on each component (for example this may be used to prohibit short selling), $W_n = \{\mathbf{w} \mid \mathbf{w}^\top \mathbf{1} = 1, \mathbf{w}^L \le \mathbf{w}\}$, where the components in the n-dimensional vector $\mathbf{w}^L := (0, 0, \ldots, 0)'$ is used to denote the lower bounds of investment weight for the corresponding stocks w_i for $i = 1, 2, \ldots, n$.

Now we translate this into a sample average problem supposing that we have a sample of observations $\tilde{r}_i^j$, $j = 1, 2, \ldots, N$, from the return R_i. Then the corresponding version of MnCnCVSA(β) is

$$\min_{\mathbf{w} \in W_n, \eta \in \mathbb{R}} \quad -\frac{1}{N} \sum_{j=1}^{N} \sum_{i=1}^{n} w_i \tilde{r}_i^j$$

$$\text{s.t.} \quad \eta + \frac{1}{(1-\beta)N} \sum_{j=1}^{N} \left(\left(-\sum_{i=1}^{n} w_i \tilde{r}_i^j \right) - \eta \right)_+ \leq U.$$

We call this the risk constrained portfolio problem (RCPP) and write the optimal portfolio as $w^*(\beta, U)$.

We will test the two stage approach in detail on synthetic data.

5.2.2 A numerical example

In order to test the method proposed above we use synthetic data based on stock returns for 66 stocks drawn from the S&P-100 index over the period from April 2011 to April 2015.[1] The data used is taken from www.finance.google.com. We use a t-distribution to model the log returns of each stock. The one-dimensional t-distribution is widely used in modelling univariate financial data because, in comparison to the Normal, it easily incorporates a heavy tail with a single extra parameter (the degree of freedom). For instance, Markowitz and Usmen [40,41] investigated the rich family of Pearson distributions and identified the t-distribution with about 4.5 degrees of freedom as the best fit to daily log-return data of the S&P500. We refer the readers to Hu and Kercheval [42] and Platen and Rendek [43] for recent results and empirical evidences. When using t-distributions in the numerical tests on synthetic data, the values of the mean and variance are estimated from the real data set, but we will vary the degree of freedom as part of our experiments.

Our interest is in improving performance out of sample by moving from $\beta = 0.975$ to $\beta = 0.95$. We begin by taking a sample of size 1000, $\tilde{r}^j$, $j = 1, 2 \ldots 1000$, of vectors of returns from the 66 stocks modelled assuming that returns from different stocks are independent and that log returns for each stock have a t-distribution. We solve the RCPP problem with $\beta = 0.975$ and CVaR$_{0.975}$ constrained to be less than 5% (i.e. a U value of -0.95). Call this optimal portfolio w_0^*. Then we follow the procedure described earlier by setting $U_1 = \text{CVaR}_{0.95}\{-w_0^{*\top} \tilde{r}^j, j = 1, 2, \ldots, 1000\}$ and then solving RCPP with $\beta = 0.95$ and U_1 to give an alternative optimal portfolio w_1^*. Note that the RCPP problem is easy to solve using linear programming techniques [44].

The next step is to compare the performance of the two portfolios out of sample. We do this by drawing 200 different out of sample data sets (each with 1000 observations). For each of these 200 data sets we calculate the mean return and the CVaR$_{0.975}$ values for both w_0^* and w_1^*. The results that we obtain depend on the choices of w_0^* and w_1^* which will vary according to the initial sample of 1000 points. A box plot for a

---

[1] We fit 81 stocks in the S&P-100 index by t-distributions and select 66 stocks with their degrees of freedom in the range of [2.5, 6.0].

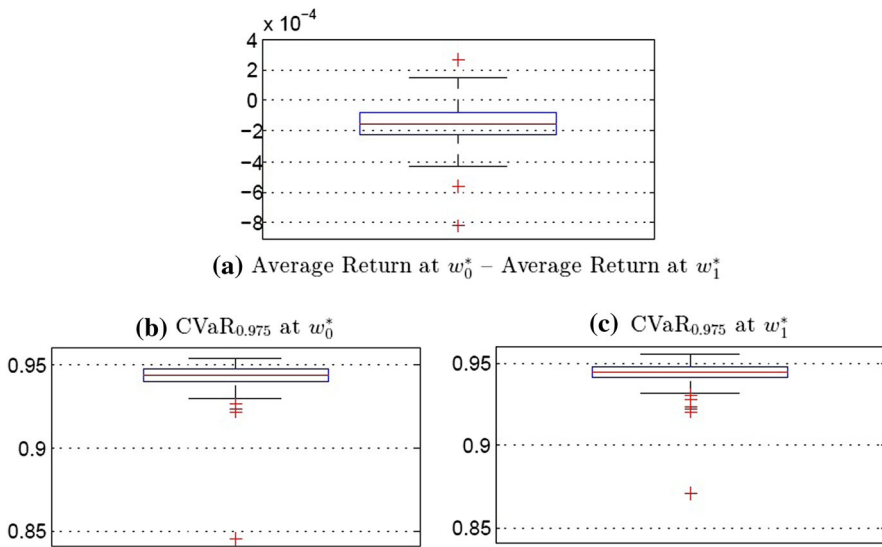**(a)** Average Return at w_0^* – Average Return at w_1^*

Fig. 4 The mean and CVaR$_{0.975}$ of returns in 200 out-of-sample tests for w_0^* and w_1^* from *MnCnCV* problems on stock sets with degrees of freedom being 3

typical example of the results we obtain is shown in Fig. 4. We describe this as a single experiment.

We repeat this experiment 50 times with different random initial samples. Most often the mean return from w_1^* is better than that from w_0^* while the CVaR values may move in either direction. More details are given in Table 1, which shows the percentage of experiments in which the return is improved under w_1^* and the percentage of times that improvement is significant at the 5% level using a paired sample t test. The table also shows the percentage of time that w_1^* leads to a lower level of CVaR$_{0.975}$ (again a paired sample t test is used to identify significant improvements). The first column of the table is generated by fixing the degrees of freedom to equal 3 for the log return t-distribution for all 66 stocks. The second and third columns show increased values for the degrees of freedom parameter (the mean and variance parameters for each stock remain at the values that are estimated from the original data set). As degrees of freedom increase the t-distribution becomes closer to the normal distribution and the tails become thinner.

The experiments here indicate that there is a sensitivity to the thickness of the tails (controlled here by the degrees of freedom parameter). For heavy tails (with degrees of freedom = 3) there is a clear benefit in mean return for a given level of risk through the use of the new procedure (i.e. using w_1^*). There is significantly better performance in more than two thirds of cases and this is achieved with no overall worsening of the risks when measured through CVaR$_{0.975}$. However, whether or not the new approach is used, we see from the table that levels of risks are higher than the 5% losses that are targeted, and are more than 6.

As the degrees of freedom are increased and the tails become thinner, the improvement in mean returns disappears. In its place we find that the new procedure generates solutions with lower average risk. With degrees of freedom = 6 we find that there

Table 1 Comparison of performance of w_0^* and w_1^* on synthetic data

	Deg. of freedom $= 3$	Deg. of freedom $= 5$	Deg. of freedom $= 6$
% of experiments where mean return from w_1^* is better than w_0^*	86%	56%	48%
% of experiments where mean return from w_1^* is significantly (p value < 0.05) better than w_0^*	68%	48%	30%
Mean CVaR$_{0.975}$ losses under w_0^* (measured as percent)	6.17	5.60	5.37
Mean CVaR$_{0.975}$ losses under w_1^* (measured as percent)	6.16	5.57	5.34
% of experiments where CVaR$_{0.975}$ with w_1^* is significantly (p value < 0.05) better than with w_0^*	52%	60%	70%

is really no advantage in w_1^* from the point of view of mean returns, but there is a small improvement in CVaR levels. In fact in 70% of the experiments CVaR levels are lower under w_1^* than they are under w_0^*. Note that in this thinner tailed case the CVaR risks are also much closer to the targeted level of 5% losses, whether or not the new procedure is used.

6 Conclusion

This paper explores the effect of changing the confidence level β when using a CVaR risk measure within an optimization problem. The setting of this confidence level in practice requires a judgement that trades off the greater confidence that a high level of β brings, against an inappropriate focus on extremely unlikely scenarios (which are associated with very high levels of loss). We can put this dilemma into concrete terms by considering a manager who needs to make a choice between operating with a constraint that losses in the worst 5% of cases have an average value of no more than \$1 million, or a constraint that losses in the worst 1% of cases have an average value of no more than \$3 million.

We begin with a study of the convergence behavior in risk-based optimization as the confidence level β approaches 1. It is well known that CVaR is a risk measure which ranges between the expected value (risk neutral) and the extreme value (most conservative risk aversion). However the convergence behavior has not received much attention in the literature so far and the results we obtain in Theorem 1 do not seem to have been given before.

Though we have formulated the problem in terms of the asymptotic behavior as β approaches 1, it is equally possible to see our results as relating to the use of CVaR approximations in a context where the original problem is one of robust optimization.

A complexity in this discussion is that the solution of these problems in practice uses a sample average to approximate the CVaR values, and these sample approximations will be worse and worse behaved as the confidence level approaches 1. On the

one hand this shows that our asymptotic results are quite strong since they overcome this difficulty. On the other hand it also suggests that lower values of β may bring some advantage in practice since the sample approximations are better behaved. We explore this possibility through a numerical investigation of certain project management problems. For these examples we show that there is a clear advantage in using a CVaR based approach after adjusting the right hand side, using a two stage procedure. In fact with a sample size of 100, β levels down to 0.9 appear to work well. We also show an advantage in using a related two stage approach in portfolio optimization.

Appendix

Proof of Theorem 1 Part (i). Suppose first that $g_x^* < \infty$. Then it is enough to show that $\lim_{\beta \to 1} \text{VaR}_\beta(g(x, \xi)) = g_x^*$. Observe that, since g_x^* is a supremum, for any $\varepsilon > 0$, there is a $\omega \in \Omega$ such that $y_0 = \xi(\omega)$ and $g(x, y_0) > g_x^* - \varepsilon$. Let $A = (g(x, y_0) - \varepsilon/2, g(x, y_0) + \varepsilon/2)$, then since g_x is continuous, $g_x^{-1}(A)$ is open and we can deduce that $\Pr(\xi \in g_x^{-1}(A)) > 0$. Thus

$$\beta_0 := F(x, g(x, y_0) + \varepsilon/2) > F(x, g(x, y_0) - \varepsilon/2).$$

Since $F(x, \alpha)$ is increasing in α, this implies

$$\text{VaR}_{\beta_0}(g(x, \xi)) \geq g(x, y_0) - \varepsilon/2 \geq g_x^* - 3\varepsilon/2.$$

As ε is arbitrary this shows the result we need. When $\sup_{y \in Y} g(x, y) = \infty$ we can use a similar approach to find values of β with $\text{VaR}_\beta(g(x, \xi)) \geq M$ for any integer M.

Part (ii). By the classical implicit function theorem, $F_x^{-1}(\beta)$ is continuously differentiable and

$$\frac{dF_x^{-1}(\beta)}{d\beta} = \frac{1}{dF_x(\alpha)/d\alpha},$$

where $\beta = F(x, \alpha)$. The assumption that $dF_x(\alpha)/d\alpha$ is monotonically decreasing is then enough to show that $F_x^{-1}(\beta)$ has an increasing derivative and hence is convex (i.e. that $\text{VaR}_\beta(g(x, \xi))$ is convex with respect to β). By the definition of CVaR

$$\text{CVaR}_\beta(g(x, \xi)) = \frac{1}{1 - \beta} \left(\int_\beta^{(1+\beta)/2} F_x^{-1}(t)dt + \int_{(1+\beta)/2}^1 F_x^{-1}(t)dt \right)$$

$$= \frac{1}{1 - \beta} \int_0^{(1-\beta)/2} \left[F_x^{-1}\left(\frac{1+\beta}{2} - s \right) + F_x^{-1}\left(\frac{1+\beta}{2} + s \right) \right] ds$$

$$\geq \frac{1}{1-\beta} \int_0^{(1-\beta)/2} 2F_x^{-1}\left(\frac{1+\beta}{2}\right) ds = F_x^{-1}\left(\frac{1+\beta}{2}\right)$$

and this establishes the first inequality of (3).

To show the second inequality of (3) we first observe that it is trivial in the case that $\sup_{y \in Y} g(x, y) = \infty$. So we can assume that f achieves its maximum value, and this is given by $F_x^{-1}(1)$. Now consider a change of variable $t = \beta + (1 - \beta)s$ and then, using the convexity of $F_x^{-1}(t)$, we have

$$\frac{1}{1-\beta} \int_\beta^1 F_x^{-1}(t)dt \leq \int_0^1 \left[sF_x^{-1}(1)+(1-s)F_x^{-1}(\beta)\right] ds \leq \frac{1}{2}\left[F_x^{-1}(1) + F_x^{-1}(\beta)\right]$$

$$= \frac{1}{2}\left[\mathrm{VaR}_\beta(g(x, \xi)) + \mathrm{VaR}_1(g(x, \xi))\right].$$

Part (iii). If we set $t = F_x(\alpha)$, then the condition on $F_x(\alpha)$ can be rewritten

$$1 - t \geq K \left(g_x^* - F_x^{-1}(t)\right)^\tau.$$

Hence we have

$$g_x^* - \mathrm{CVaR}_\beta(g(x, \xi)) = \frac{1}{1-\beta} \int_\beta^1 (g_x^* - F_x^{-1}(t))dt \leq \frac{1}{1-\beta} \int_\beta^1 \left(\frac{1-t}{K}\right)^{1/\tau} dt$$

$$= \frac{1}{K^{1/\tau}} \frac{1}{1 + (1/\tau)}(1 - \beta)^{1/\tau}.$$

The proof is complete. □

Proof of Proposition 1 By (11), we have

$$\Pr\left(\|\xi^* - \xi\| < \frac{1}{C_1}z^{1/v_1}\right) \leq \Pr(g(x, \xi^*) - g(x, \xi) < z)$$

since the event on the left hand side implies the event on the right hand side. Through (12), we obtain

$$\Pr(g(x, \xi) > g_x^* - z) \geq \frac{C_2}{C_1}z^{v_2/v_1}. \tag{33}$$

provided that $\frac{1}{C_1}z^{1/v_1} \leq \delta_0$. Thus

$$1 - F_x(\alpha) = \Pr(g(x, \xi) > \alpha) \geq \frac{C_2}{C_1}(g_x^* - \alpha)^{v_2/v_1}$$

provided $\frac{1}{C_1}(g_x^* - \alpha)^{1/v_1} \leq \delta_0$. Hence this inequality holds for

$$\alpha \in \left(g_x^* - (C_1\delta_0)^{v_1}, g_x^*\right)$$

The final step is to show that we can find a single β_0 such that $\text{CVaR}_{\beta_0}(g(x, \xi)) > g_x^* - (C_1\delta_0)^{\nu_1}$ for every $x \in X$. We choose $\beta_0 = 1 - \frac{C_2}{C_1}(C_1\delta_0)^{\nu_2}$ having adjusted δ_0 downwards if necessary to ensure that $\beta_0 > 0$. So

$$\frac{1}{C_1}\left(\left(\frac{C_1}{C_2}(1-\beta_0)\right)^{\nu_1/\nu_2}\right)^{1/\nu_1} = \delta_0$$

and hence from (33) for every $x \in X$,

$$\Pr\left(g(x, \xi) > g_x^* - \left(\frac{C_1}{C_2}(1-\beta_0)\right)^{\nu_1/\nu_2}\right) \geq 1 - \beta_0.$$

Thus from the definition of VaR

$$\text{VaR}_{\beta_0}(g(x, \xi)) \geq g_x^* - \left(\frac{C_1}{C_2}(1-\beta_0)\right)^{\nu_1/\nu_2} = g_x^* - (C_1\delta_0)^{\nu_1}.$$

Hence we have $\text{CVaR}_{\beta_0}(g(x, \xi)) > g_x^* - (C_1\delta_0)^{\nu_1}$ and the result is established. □

Proof of Theorem 2 Applying [23, Lemma 3.1] to $\text{MnCV}(\beta)$ (treating it as a perturbation of MnMx), we know that for any $\epsilon > 0$, there exists a $\delta > 0$ such that $\mathbb{D}(X^*(\beta), X^*) \leq \epsilon$ when $\Delta_\beta(x) \leq \delta$. This shows (16) because

$$\lim_{\beta \to 0} \sup_{x \in X} \Delta_\beta(x) = 0$$

from Theorem 1. Moreover, under the growth condition (17) and the consistent tail condition (4), we obtain (18) by virtue of [23, Lemma 3.1]. □

Proof of Theorem 3 Part (i). The conclusion follows from [45, Lemma 4.2 (i)]. Here we provide a proof for completeness. Let ϵ be a fixed small positive number. Define

$$R(\epsilon) := \inf_{\{x \in X, d(x, \mathcal{G}) \geq \epsilon\}} \sup_{y \in Y} g(x, y). \tag{34}$$

Then $R(\epsilon) > 0$ as we take an infimum over x values outside of $\mathcal{G}$. Let $\Phi_\beta(x, \eta)$ be defined as in (13). Under the condition that $g(x, y)$ has consistent tail behavior on X, then Theorem 1 implies that $\min_{\eta \in \mathbb{R}} \Phi_\beta(x, \eta)$ approximates $\sup_{y \in Y} g(x, y)$ uniformly with respect to x over X as $\beta \to 1$, i.e. if we choose β sufficiently close to 1 then

$$\sup_{x \in X}\left[\sup_{y \in Y} g(x, y) - \min_{\eta \in \mathbb{R}} \Phi_\beta(x, \eta)\right] \leq R(\epsilon)/2.$$

Then for any $x \in X$ with $d(x, \mathcal{G}) \geq \epsilon$,

$$\min_{\eta \in \mathbb{R}} \Phi_\beta(x, \eta) = \sup_{y \in Y} g(x, y) + \min_{\eta \in \mathbb{R}} \Phi_\beta(x, \eta) - \sup_{y \in Y} g(x, y) \geq R(\epsilon) - R(\epsilon)/2 > 0,$$

which implies $x \notin \mathcal{G}(\beta)$. This is equivalent to saying that $d(x, \mathcal{G}) < \epsilon$ for every $x \in \mathcal{G}(\beta)$, that is, $\mathbb{D}(\mathcal{G}(\beta), \mathcal{G}) \leq \epsilon$. The conclusion follows by noting that $\mathcal{G} \subset \mathcal{G}(\beta)$.

Part (ii). Under the Slater constraint qualification and convexity of $g(\cdot, y)$, it follows by [46, Proposition 2.8], that (20) satisfies the metric regularity condition, that is, there exists a positive constant C such that

$$d(x, \mathcal{G}) \leq C \left(\sup_{y \in Y} g(x, y) \right)_+, \quad \forall x \in X. \tag{35}$$

Let $\hat{x} \in \mathcal{G}(\beta)$. Then $\min_{\eta \in \mathbb{R}} \Phi_\beta(\hat{x}, \eta) \leq 0$ and

$$d(\hat{x}, \mathcal{G}) \leq C \left(\sup_{y \in Y} g(\hat{x}, y) \right)_+ - C \min_{\eta \in \mathbb{R}} \Phi_\beta(\hat{x}, \eta)$$

$$\leq C \sup_{x \in X} \left[\sup_{y \in Y} g(x, y) - \min_{\eta \in \mathbb{R}} \Phi_\beta(x, \eta) \right].$$

The rest follows from (10) under the consistent tail condition (4). $\qquad \square$

Proof of Theorem 4 Parts (i) and (ii). Write the constraints of MnCnMx as $\max_{y \in Y} g(x, y) \leq 0$. Since the two problems have identical objective functions, it suffices to look into the impact of the difference of the constraints on the optimal value and optimal solutions. Under conditions (a) and (b), it follows by Theorem 3 that the feasible set of (19) is closed in β at $\beta = 1$. By classical stability results (see e.g. [47, Theorem 4.2.1]), $\hat{\vartheta}(\beta)$ converges to $\hat{\vartheta}$ as $\beta \to 1$ and $\overline{\lim}_{\beta \to 1} \hat{X}(\beta) \subset \hat{X}$. Part (iii). By Theorem 3 (ii), the feasible set of (19) is upper pseudo-Lipschitz continuous ([4]) in β at $\beta = 1$. By applying Klatte's stability result [26, Theorem 1], we obtain

$$|\hat{\vartheta}(\beta) - \hat{\vartheta}| \leq L \mathbb{H}(\mathcal{G}(\beta), \mathcal{G}) \leq \frac{LC}{K^{1/\tau}} \frac{\tau}{1 + \tau} (1 - \beta)^{1/\tau},$$

which is (24). $\qquad \square$

Proof of Theorem 5 Under condition (b), $\bar{g}(x)$ is convex and $\mathcal{F}$ is a compact and convex set of X. Together with the Slater condition (22) under condition (a), we have by Robinson's error bound,

$$d(x, \mathcal{F}) \leq \frac{D}{\delta} \|(\overline{g}(x))_+\|. \tag{36}$$

Since x^* is feasible for MnCnCV(β) we have $h(x_\beta^*) \leq h(x^*)$. Then by the second order growth condition (c)

$$h(\Pi_{\mathcal{F}} x_\beta^*) - h(x_\beta^*) \geq h(\Pi_{\mathcal{F}} x_\beta^*) - h(x^*) \geq r \|\Pi_{\mathcal{F}} x_\beta^* - x^*\|^2, \tag{37}$$

where we write $\Pi_{\mathcal{F}}$ for the orthogonal projection onto $\mathcal{F}$. On the other hand, the Lipchitz continuity of h implies that

$$h(\Pi_{\mathcal{F}}x_{\beta}^*) - h(x_{\beta}^*) \leq L\|\Pi_{\mathcal{F}}x_{\beta}^* - x_{\beta}^*\|.$$

Combining the two inequalities above, we obtain

$$r\|\Pi_{\mathcal{F}}x_{\beta}^* - x^*\|^2 \leq L\|\Pi_{\mathcal{F}}x_{\beta}^* - x_{\beta}^*\|$$

Thus

$$\|x_{\beta}^* - x^*\| \leq \|x_{\beta}^* - \Pi_{\mathcal{F}}x_{\beta}^*\| + \|\Pi_{\mathcal{F}}x_{\beta}^* - x^*\|$$

$$\leq d(x_{\beta}^*, \mathcal{F}) + \sqrt{\frac{L}{r}}\sqrt{d(x_{\beta}^*, \mathcal{F})}$$

$$\leq \frac{D}{\delta}\left(\overline{g}(x_{\beta}^*)\right)_+ + \sqrt{\frac{LD}{r\delta}}\sqrt{(\overline{g}(x_{\beta}^*))_+}.$$

Using condition (d), we have

$$|g_{\beta}(x_{\beta}^*) - \overline{g}(x_{\beta}^*)| \leq C(1-\beta)^{1/\tau}.$$

Moreover as $\overline{g}(x^*) \leq 0$, condition (d) also implies $g_{\beta}(x^*) \leq C(1-\beta)^{1/\tau}$. Hence using the fact that x_{β}^* is feasible for MnCnCV(β),

$$\overline{g}(x_{\beta}^*) \leq g_{\beta}(x_{\beta}^*) + C(1-\beta)^{1/\tau} \leq g_{\beta}(x^*) + C(1-\beta)^{1/\tau} \leq 2C(1-\beta)^{1/\tau}(\beta),$$

giving the bound we require.   □

Proof of Theorem 6 Since the constraint in MnCnMx is active at x^*, we have from the Karush-Kuhn-Tucker conditions that

$$\frac{\partial h}{\partial x_i}(x^*) + \lambda^*\frac{\partial \overline{g}}{\partial x_i}(x^*) = 0, \quad \text{for} \ i = 1, 2, \ldots n \tag{38}$$

for some non-negative Lagrange multiplier λ^*. In the case when $\lambda^* = 0$, x^* is the unconstrained optimal solution for MnCnMx and hence also the optimal solution for MnCnCV(β) and the bound trivially holds. In what follows, we consider the case $\lambda^* > 0$. Let $F(y)$ denote the cumulative distribution function of ξ. Then by definition, $\text{VaR}_{\beta}(\xi) = F^{-1}(\beta)$. Moreover, for any $u \in (0, 1)$,

$$u = \Pr\left(g(x, \xi) \leq \text{VaR}_u(g(x, \xi))\right).$$

Since $g(x, \xi)$ is increasing in ξ, we also have

$$u = \Pr(\xi < \text{VaR}_u(\xi)) = \Pr\left(g(x, \xi) \leq g(x, \text{VaR}_u(\xi))\right).$$

Thus $\mathrm{VaR}_u(g(x,\xi)) = g(x, \mathrm{VaR}_u(\xi))$ and

$$g_\beta(x) = \frac{1}{1-\beta} \int_\beta^1 \mathrm{VaR}_u(g(x,\xi)) du = \frac{1}{1-\beta} \int_{\mathrm{VaR}_\beta(\xi)}^{\sup(\xi)} g(x,y)\rho(y)dy$$

by changing variable with $y = F^{-1}(u)$. Let κ be the supremum of the mixed second derivative of g at x^* that appears in the bound, i.e.

$$\kappa = \max_{i=1,\ldots,n} \sup_{\xi \in Y} \left| \frac{\partial g}{\partial x_i \partial \xi}(x^*,\xi) \right|.$$

We use $\overline{\xi}$ to denote the supreme of ξ. Then

$$\left| \frac{\partial g_\beta}{\partial x_i}(x^*) - \frac{\partial \overline{g}}{\partial x_i}(x^*) \right|$$

$$= \left| \frac{1}{1-\beta} \int_{\mathrm{VaR}_\beta(\xi)}^{\overline{\xi}} \frac{\partial g}{\partial x_i}(x^*,y)\rho(y)dy - \frac{\partial \overline{g}}{\partial x_i}(x^*) \right|$$

$$= \left| \frac{1}{1-\beta} \int_{\mathrm{VaR}_\beta(\xi)}^{\overline{\xi}} \left(\frac{\partial g}{\partial x_i}(x^*,\overline{\xi}) - \int_y^{\overline{\xi}} \frac{\partial^2 g}{\partial x_i \partial \xi}(x^*,t)dt \right) \rho(y)dy - \frac{\partial \overline{g}}{\partial x_i}(x^*) \right|$$

$$\leq \left| \frac{1}{1-\beta} \int_{\mathrm{VaR}_\beta(\xi)}^{\overline{\xi}} (\overline{\xi} - y)\kappa\rho(y)dy \right|$$

$$= \kappa \left(\overline{\xi} - \mathrm{CVaR}_\beta(\xi) \right)$$

where we have used the fact that monotonicity of g implies $\overline{g}(x) = g(x, \overline{\xi})$. Through (38), the inequality above implies

$$\left| \frac{\partial h}{\partial x_i}(x^*) + \lambda^* \frac{\partial g_\beta}{\partial x_i}(x^*) \right| \leq \lambda^* \kappa \left(\overline{\xi} - \mathrm{CVaR}_\beta(\xi) \right)$$

and consequently

$$(x_\beta^* - x^*)_i \frac{\partial h}{\partial x_i}(x^*) \geq -\lambda^*(x_\beta^* - x^*)_i \frac{\partial g_\beta}{\partial x_i}(x^*) - \lambda^* \kappa \left(\overline{\xi} - \mathrm{CVaR}_\beta(\xi) \right) \left| (x_\beta^* - x^*)_i \right|,$$

where $(x_\beta^* - x^*)_i$ is the ith component of $(x_\beta^* - x^*)$ for $i = 1, 2, \ldots, n$. By exploiting the convexity of h, the inequality above enables us to obtain

$$h(x_\beta^*) \geq h(x^*) + (x_\beta^* - x^*)^T \nabla h(x^*)$$

$$\geq h(x^*) - \lambda^*(x_\beta^* - x^*)^T \nabla g_\beta(x^*) - \lambda^* \kappa \left(\overline{\xi} - \mathrm{CVaR}_\beta(\xi) \right) \sum \left| (x_\beta^* - x^*)_i \right|$$

$$\geq h(x^*) - \lambda^*(x_\beta^* - x^*)^T \nabla g_\beta(x^*) - \lambda^* \kappa \left(\overline{\xi} - \mathrm{CVaR}_\beta(\xi) \right) \sqrt{n} \left\| x_\beta^* - x^* \right\|,$$

$$\tag{39}$$

where we have used the fact that $\sqrt{n}\,\|x\| \geq x^T e$ for $x \in \mathbb{R}^n$ in the last inequality and e is a vector with unit components.

On the other hand, by (25),

$$g_\beta(x_0) \geq g_\beta(x^*) + (x_\beta^* - x^*)^T \nabla g_\beta(x^*) + K_c \left\| x_\beta^* - x^* \right\|^2$$

and, because the constraint is active, $g_\beta(x_0) = g_\beta(x^*)$, we deduce from the inequality

$$-(x_\beta^* - x^*)^T \nabla g_\beta(x^*) \geq K_c \left\| x_\beta^* - x^* \right\|^2. \tag{40}$$

A combination of (39) and (40) gives rise to

$$h(x_\beta^*) \geq h(x^*) + \left(\lambda^* K_c\right) \left\| x_\beta^* - x^* \right\|^2 - \lambda^* \kappa \left(\overline{\xi} - \mathrm{CVaR}_\beta(\xi)\right) \sqrt{n} \left\| x_\beta^* - x^* \right\|.$$

Since x^* is a feasible solution to MnCnMx(β) whereas x_β^* is optimal, then $h(x_\beta^*) \leq h(x^*)$ and the inequality above implies

$$(\lambda^* K_c) \left\| x_\beta^* - x^* \right\|^2 < \lambda^* \kappa \left(\overline{\xi} - \mathrm{CVaR}_\beta(\xi)\right) \sqrt{n} \left\| x_\beta^* - x^* \right\|$$

from which we derive

$$\left\| x_\beta^* - x^* \right\| < \frac{\kappa}{K_c} \sqrt{n} \left(\overline{\xi} - \mathrm{CVaR}_\beta(\xi)\right)$$

as required. $\square$

Proof of Theorem 7 We can write MnCnCVSA(β) in the following equivalent form

$$\begin{aligned} &\min_x \; h(x) \\ &\text{s.t.} \; \min_\eta \Phi_\beta^N(x, \eta) \leq 0, \\ &\qquad x \in X, \end{aligned} \tag{41}$$

where

$$\Phi_\beta^N(x, \eta) := \eta + \frac{1}{(1-\beta)N} \sum_{j=1}^{N} (g(x, \xi^j) - \eta)_+.$$

Thus x^N is an optimal solution of (41). Let $\mathcal{G}(\beta)$ be defined as in Sect. 3, that is, the set of solutions to inequality (21). Then $\mathcal{G}(\beta)$ is the feasible set of minimization (19). Let $\mathcal{G}_N(\beta)$ denote the feasible set of (41). We give the proof in 6 steps.

Step 1 Since MnCnMx is feasible, then problem (19) satisfies the Slater condition which means there exists a positive constant δ and a point $\hat{x} \in \mathcal{G} \subset \mathcal{G}(\beta)$ such that

$$\min_{\eta \in \mathbb{R}} \Phi_\beta(\hat{x}, \eta) < -\delta < 0.$$

Let $x_\beta \in \hat{X}(\beta)$ be a solution of MnCnCV(β). Let $x_t = (1-t)x_\beta + t\hat{x}$. Since $g(x, \xi)$ is convex as a function of x, the function $\min_{\eta \in \mathbb{R}} \Phi_\beta(x, \eta)$ is also convex (see [17]) and

$$\min_{\eta \in \mathbb{R}} \Phi_\beta(x_t, \eta) \le -t\delta. \tag{42}$$

Let ϵ be a positive number and $R(\cdot)$ be defined as in (28). For $x \in \mathcal{G}(\beta)$, $d(x, \hat{X}(\beta)) \ge \epsilon$ implies $h(x) - \hat{\vartheta}(\beta) \ge R(\epsilon) > 0$. Let $\bar{t}$ be small enough that

$$L\bar{t}\|x_\beta - \hat{x}\| < R(\epsilon) - L\gamma, \tag{43}$$

where L is the Lipschitz modulus of $h(x)$ and $\gamma < R(\epsilon)/L$ is some small positive number.

Step 2 Let σ be a small positive number. Using conditions (c) and (d) on bounded moment generating functions we may apply the result of [25, Theorem 5.1] to show that there exist positive constants $C(\sigma, \beta)$, $\alpha(\sigma, \beta)$ and $N(\sigma, \beta)$ such that

$$\Pr\left(\sup_{x \in X}\left|\min_\eta \Phi_\beta^N(x, \eta) - \min_\eta \Phi_\beta(x, \eta)\right| \ge \sigma\right) \le C(\sigma, \beta)e^{-\alpha(\sigma,\beta)N} \tag{44}$$

for $N \ge N(\sigma, \beta)$.

Step 3 Let δ and $\bar{t}$ be given as in Step 1. We estimate $\Pr(x_{\bar{t}} \notin \mathcal{G}_N(\beta))$.

$$
\begin{aligned}
\Pr(x_{\bar{t}} \notin \mathcal{G}_N(\beta)) &= \Pr\left(\min_{\eta \in \mathbb{R}} \Phi_\beta^N(x_{\bar{t}}, \eta) > 0\right) \\
&= \Pr\left(\min_{\eta \in \mathbb{R}} \Phi_\beta^N(x_{\bar{t}}, \eta) - \min_{\eta \in \mathbb{R}} \Phi_\beta(x_{\bar{t}}, \eta) > -\min_{\eta \in \mathbb{R}} \Phi_\beta(x_{\bar{t}}, \eta)\right) \\
&\le \Pr\left(\min_{\eta \in \mathbb{R}} \Phi_\beta^N(x_{\bar{t}}, \eta) - \min_{\eta \in \mathbb{R}} \Phi_\beta(x_{\bar{t}}, \eta) > \delta\bar{t}\right) \quad \text{(by (42))} \\
&< C(\sigma, \beta)e^{-\alpha(\sigma,\beta)N} \tag{45}
\end{aligned}
$$

for $N \ge N(\sigma, \beta)$. The last inequality is due to (44) by setting $\sigma < \bar{t}\delta$. This shows that $x_{\bar{t}} \in \mathcal{G}_N(\beta)$ with probability $1 - C(\sigma, \beta)e^{-\alpha(\sigma,\beta)N}$.

Step 4 Suppose that $x^N \notin \mathcal{G}(\beta)$ then $\min_{\eta \in \mathbb{R}} \Phi_\beta(x^N, \eta) > 0$ and we define the point $y^N = (1-s)x^N + s\hat{x}$ by choosing s so that $\min_{\eta \in \mathbb{R}} \Phi_\beta(y^N, \eta) = 0$ (where s depends on N). Then from the convexity of $\min_{\eta \in \mathbb{R}} \Phi_\beta(y^N, \eta)$ we have

$$\min_\eta \Phi_\beta(x^N, \eta) > \frac{s}{1-s}\delta.$$

Letting D be the diameter of $\mathcal{G}(\beta)$, we have $d(y^N, \hat{x}) \le D$ so $d(y^N, x^N) \le \frac{s}{1-s}D$. Hence if $d(y^N, x^N) \ge \gamma$, where γ is given in Step 1, then

$$\min_\eta \Phi_\beta(x^N, \eta) \ge \frac{\delta}{D}\gamma.$$

Thus, setting $\sigma \leq \delta\gamma/D$, and since $\min_\eta \Phi_\beta^N(x^N, \eta) \leq 0$, we can deduce that

$$\Pr(\min_\eta \Phi_\beta(x^N, \eta) \geq \sigma) \leq \Pr\left(\sup_{x \in X}\left|\min_\eta \Phi_\beta^N(x, \eta) - \min_\eta \Phi_\beta(x, \eta)\right| \geq \sigma\right)$$
$$\leq C(\sigma, \beta)e^{-\alpha(\sigma,\beta)N}.$$

Therefore

$$\Pr(d(x^N, y^N) \geq \gamma) \leq C(\sigma, \beta)e^{-\alpha(\sigma,\beta)N}.$$

Step 5 We will establish that if $d(x^N, \hat{X}(\beta)) \geq \epsilon + \gamma$ then either $x_{\bar{i}} \notin \mathcal{G}_N(\beta)$; or $x^N \notin \mathcal{G}(\beta)$ and $d(x^N, y^N) \geq \gamma$ (these are the two cases dealt with in steps 3 and 4). Hence suppose that $x_{\bar{i}} \in \mathcal{G}_N(\beta)$ and $x^N \notin \mathcal{G}(\beta)$. If $d(y^N, \hat{X}(\beta)) \geq \epsilon$. Then $h(y^N) - h(x_\beta) \geq R(\epsilon)$. So

$$h(x^N) - h(y^N) = (h(x^N) - h(x_\beta)) - (h(y^N) - h(x_\beta))$$
$$\leq h(x_{\bar{i}}) - h(x_\beta) - R(\epsilon)$$
$$\leq L\bar{t}\|x_\beta - \hat{x}\| - R(\epsilon) < -L\gamma,$$

which implies that $d(y^N, x^N) > \gamma$ through the Lipschitness of h. On the other hand if $d(y^N, \hat{X}(\beta)) < \epsilon$ then (using the triangle inequality)

$$d(y^N, x^N) > d(x^N, \hat{X}(\beta)) - d(y^N, \hat{X}(\beta)) > \gamma.$$

The only case that remains is when $x_{\bar{i}} \in \mathcal{G}_N(\beta)$ and $x^N \in \mathcal{G}(\beta)$. But if $d(x^N, \hat{X}(\beta)) \geq \epsilon + \gamma$, then

$$h(x^N) - h(x_{\bar{i}}) = (h(x^N) - h(x_\beta)) - (h(x_{\bar{i}}) - h(x_\beta))$$
$$> R(\epsilon) - (R(\epsilon) - L\gamma) > 0$$

using the growth condition on h and inequality (43). However this inequality contradicts the optimality of x^N. Summarizing the discussions above, we can conclude that

$$\Pr(d(x^N, \hat{X}(\beta)) \geq \epsilon + \gamma) \leq 2C(\sigma, \beta)e^{-\alpha(\sigma,\beta)N}.$$

for $N \geq N(\sigma, \beta)$. Since ϵ and γ were chosen arbitrarily we can make $\varepsilon = \epsilon + \gamma$ to conclude the proof of Part (i).

Step 6 Finally, we show Part (ii). We estimate $\Pr(d(x^N, \hat{X})$. By the properties of $\mathbb{D}$, we have

$$\Pr(d(x^N, \hat{X}) \geq \varepsilon) \leq \Pr(d(x^N, \hat{X}(\beta)) + \mathbb{D}(\hat{X}(\beta), \hat{X})) \geq \varepsilon). \tag{46}$$

Let $\varepsilon = 2\epsilon$. Under the consistent tail condition of g, we know from Theorem 4 that by setting β sufficiently close to 1 so that $\mathbb{D}(\hat{X}(\beta), \hat{X}) \leq \epsilon$. By (46)

$$\begin{aligned}
\Pr(d(x^N, \hat{X}) \geq 2\epsilon) &\leq \Pr(d(x^N, \hat{X}(\beta)) + \mathbb{D}(\hat{X}(\beta), \hat{X})) \geq 2\epsilon) \\
&\leq \Pr(d(x^N, \hat{X}(\beta)) \geq \epsilon) \\
&\leq C(\sigma, \beta) e^{-\alpha(\sigma, \beta) N}
\end{aligned}$$

and this gives the result of Part (ii). $\qquad\square$

Proof of Proposition 2 We define the random variable $Z = \left(g(x, \xi) - \text{VaR}_\beta(g(x, \xi))\right)_+$ and suppose that Z has variance $\sigma_0(\beta)^2$. Then [7] Theorem 1, shows that the asymptotic variance of $\widehat{\text{CVaR}}_\beta$ as $N \to \infty$ is $\sigma_0(\beta)^2 / (N(1-\beta)^2)$.

We can calculate the variance of Z by noting that it is a mixture between the value 0 and a random variable Z_+ which has a distribution given by (29) with $u = \text{VaR}_\beta(g(x, \xi))$ (so the distribution function is $F_{Z_+}(t) = F_{> \text{VaR}_\beta(g(x,\xi))}(t)$). In fact for $t > 0$ we have

$$\begin{aligned}
F_Z(t) &= \Pr(\xi - \text{VaR}_\beta(\xi) < t) \\
&= \Pr(\xi \leq \text{VaR}_\beta(\xi)) \\
&\quad + \Pr(\xi > \text{VaR}_\beta(\xi)) \Pr(\xi - \text{VaR}_\beta(\xi) < t \mid \xi > \text{VaR}_\beta(\xi)) \\
&= \beta + (1 - \beta) F_{Z_+}(t),
\end{aligned}$$

which implies that Z has a probability mass of β at 0 and otherwise has a density $(1 - \beta) f_{Z_+}(t)$. Here we write $f_{Z_+}(t)$ for the density of Z_+. Then by the definition of $\sigma_0(\beta)^2$,

$$\begin{aligned}
\sigma_0(\beta)^2 &= \mathbb{E}(Z^2) - \mathbb{E}(Z)^2 \\
&= (1 - \beta) \int_0^\infty t^2 f_{Z_+}(t) dt - \left((1 - \beta) \int_0^\infty t f_{Z_+}(t) dt\right)^2 \\
&= (1 - \beta) \left(\int_0^\infty t^2 f_{Z_+}(t) dt - \left(\int_0^\infty t f_{Z_+}(t) dt\right)^2\right) \\
&\quad + \left((1 - \beta) - (1 - \beta)^2\right) \left(\int_0^\infty t f_{Z_+}(t) dt\right)^2 \\
&= (1 - \beta) \text{var}(Z_+) + (1 - \beta)\beta \mathbb{E}(Z_+)^2, \tag{47}
\end{aligned}$$

where $\text{var}(Z_+)$ denotes variance of Z_+.

We consider changing the value of β and we continue to let $u = \text{VaR}_\beta(g(x, \xi))$. The assumption we make is that the excess Z_+ can be approximated with a generalized Pareto distribution for u values greater than some threshold u_0. We make use of some properties of the generalized Pareto distribution (see [29]). In this case Z_+ will have a generalized Pareto distribution with parameter $\gamma = \gamma_0 + \kappa(u - u_0)$ where γ_0 is the parameter applying at u_0, and the shape parameter κ remains unchanged. The

generalized Pareto has mean $\gamma/(1 - \kappa)$ provided $\kappa < 1$, and variance $\gamma^2/((1 - \kappa)^2(1 - 2\kappa))$ provided $\kappa < 0.5$. Thus [substituting into Eq. (47)]

$$\sigma_0(\beta)^2 = \frac{(1-\beta)}{(1-\kappa)^2}\left(\frac{1}{(1-2\kappa)} + \beta\right)(\gamma_0 + \kappa(u - u_0))^2.$$

We can use the form of the generalized Pareto distribution to calculate the VaR (see [29]).

$$u = \mathrm{VaR}_\beta(g(x,\xi)) = u_0 + \frac{\gamma_0}{\kappa}\left(\left(\frac{1-\beta}{1 - F_x(u_0)}\right)^{-\kappa} - 1\right),$$

where $F_x(\cdot)$ is the cumulative distribution function for $g(x,\xi)$. Consequently

$$\sigma_0(\beta)^2 = \frac{(1-\beta)}{(1-\kappa)^2}\left(\frac{1}{(1-2\kappa)} + \beta\right)\gamma_0^2\left(\frac{1-\beta}{1 - F_x(u_0)}\right)^{-2\kappa}.$$

Thus the asymptotic variance of $\widehat{\mathrm{CVaR}}_\beta$ has β dependence given by

$$\sigma_0(\beta)^2/(N(1-\beta)^2) = \left(\frac{\gamma_0^2}{N(1-\kappa)^2(1 - F_x(u_0))}\right)\left(\frac{1}{(1-2\kappa)} + \beta\right)(1-\beta)^{-2\kappa-1}.$$

The proof is complete. □

References

1. Basel Committee on Banking Supervision, Fundamental review of the trading book: A revised market risk framework, Bank for International Settlements (2013)
2. Ben-Tal, A., El Ghaoui, L., Nemirovski, A.: Robust Optimization. Princeton University Press, Princeton (2009)
3. Polak, E.: On the mathematical foundations of nondifferentiable optimization in engineering design. SIAM Rev. **29**, 21–89 (1987)
4. Soyster, A.L.: Convex programming with set-inclusive constraints and applications to inexact linear programming. Oper. Res. **21**, 1154–1157 (1973)
5. Scaillet, O.: Nonparametric estimation and sensitivity analysis of expected shortfall. Math. Financ. **14**, 115–129 (2004)
6. Embrechts, P., Klueppelberg, C., Mikosch, T.: Modeling Extremal Events for Insurance and Finance. Springer, Berlin (1997)
7. Chen, S.X.: Nonparametric estimation of expected shortfall. J. Financ. Econ. **6**, 87–107 (2008)
8. Calafiore, G., Campi, M.C.: Uncertain convex programs: randomized solutions and confidence levels. Math. Program. **102**, 25–46 (2005)
9. Ben-Tal, A., Nemirovski, A.: Robust truss topology design via semidefinite programming. SIAM J. Optim. **7**, 991–1016 (1997)
10. Ben-Tal, A., Nemirovski, A.: Robust convex optimization. Math. Oper. Res. **23**, 769–805 (1998)
11. El Ghaoui, L., Lebret, H.: Robust solutions to uncertain semidefinite programs. SIAM J. Optim. **9**, 33–52 (1998)
12. Calafiore, G., Campi, M.C.: The scenario approach to robust control design. IEEE Trans. Autom. Control **51**, 742–753 (2006)

13. Calafiore, G.: Random convex programs. SIAM J. Optim. **20**, 3427–3464 (2010)
14. Campi, M.C., Garatti, S.: The exact feasibility of randomized solutions of robust convex programs. SIAM J. Optim. **19**, 1211–1230 (2008)
15. Campi, M.C., Garatti, S.: A sampling-and-discarding approach to chance constrained optimization: feasibility and optimality. J. Optim. Theory Appl. **148**, 257–280 (2010)
16. Vayanos, P., Kuhn, D., Rustem, B.: A constraint sampling approach for multi-stage robust optimization. Automatica **48**, 459–471 (2012)
17. Rockafellar, R.T., Uryasev, S.: Optimization of conditional value-at-risk. J. Risk **2**, 21–41 (2000)
18. Pflug, G., Römisch, W.: Modeling, Measuring and Managing Risk. World Scientific, Singapore (2007)
19. Zhu, L., Li, H.: Asymptotic analysis of multivariate tail conditional expectations. N. Am. Actuar. J. **16**, 350–363 (2012)
20. Rockafellar, R.T., Wets, R.J.-B.: Variational Analysis. Springer, Berlin (1998)
21. Rockafellar, R.T., Uryasev, S.: Conditional value-at-risk for general loss distributions. J. Bank. Financ. **26**, 1443–1471 (2002)
22. Ogryczak, W., Ruszczynski, A.: Dual stochastic dominance and related mean-risk models. SIAM J. Optim. **13**, 60–78 (2002)
23. Guo, S., Xu, H., Zhang, L.: Convergence analysis for mathematical programs with distributionally robust chance constraint. SIAM J. Optim. **27**, 784–816 (2017)
24. Robinson, S.M.: An application of error bounds for convex programming in a linear space. SIAM J. Control Optim. **13**, 271–273 (1975)
25. Shapiro, A., Xu, H.: Stochastic mathematical programs with equilibrium constraints, modeling and sample average approximation. Optimization **57**, 395–418 (2008)
26. Klatte, D.: A note on quantitative stability results in nonlinear optimization. Seminarbericht Nr. 90, Sektion Mathematik, Humboldt-Universität zu, pp. 77–86. Berlin, Berlin (1987)
27. Heyde, C.C., Kou, S.G.: On the controversy over tailweight of distributions. Oper. Res. Lett. **32**, 399–408 (2004)
28. Lim, A.E., Shanthikumar, J.G., Vahn, G.: Conditional value-at-risk in portfolio optimization: coherent but fragile. Oper. Res. Lett. **39**, 163–171 (2011)
29. McNeil, A.J., Frey, R., Embrechts, P.: Quantitative Risk Management. Princeton University Press, Princeton (2005)
30. Brazauskas, V., Jones, B.L., Puri, M.L., Zitikis, R.: Estimating conditional tail expectation with actuarial applications in view. J. Stat. Plan. Inference **138**, 3590–3604 (2008)
31. Deme, E.H., Girard, S., Guillou, A.: Reduced-bias estimator of the conditional tail expectation of heavy-tailed distributions. In: Hallin, M., Mason, D., Pfeifer, D., Steinebach, J. (eds.) Mathematical Statistics and Limit Theorems. Springer, New York (2015)
32. Necir, A., Rassoul, A., Zitikis, R.: Estimating the conditional tail expectation in the case of heavy-tailed losses. J. Probab. Stat. **2010**, 1–17 (2010)
33. Ballestin, F., Leus, R.: Resource-constrained project scheduling for timely project completion with stochastic activity durations. Prod. Oper. Manag. **18**, 459–474 (2009)
34. Laslo, Z.: Activity time-cost tradeoffs under time and cost chance constraints. Comput. Ind. Eng. **44**, 365–384 (2003)
35. AbouRizk, S.M., Halpin, D.W.: Statistical properties of construction duration data. J. Constr. Eng. Manag. **118**, 525–543 (1992)
36. Johnson, D.: The triangular distribution as a proxy for the beta distribution in risk analisis. J. R. Stat. Soc. Ser. Stat. **46**, 387–398 (1997)
37. Zhu, S., Fukushima, M.: Worst-case conditional value-at-risk with application to robust portfolio management. Oper. Res. **57**, 1155–1168 (2009)
38. Wozabal, D.: Robustifying convex risk measures for linear portfolios: a nonparametric approach. Oper. Res. **62**, 1302–1315 (2014)
39. Karoui, N.E., Lim, A.E., Vahn, G.Y.: Performance-based regularization in mean-CVaR portfolio optimization. arXiv preprint arXiv:1111.2091 (2011)
40. Markowitz, H., Usmen, N.: The likelihood of various stock market return distributions, Part 1: principles of inference. J. Risk Uncertain. **13**, 207–219 (1996a)
41. Markowitz, H., Usmen, N.: The likelihood of various stock market return distributions, Part 2: empirical results. J. Risk Uncertain. **13**, 221–247 (1996b)
42. Hu, W., Kercheval, A.N.: Portfolio optimization for student t and skewed t returns. Quant. Financ. **10**, 91–105 (2010)

43. Platen, E., Sidorowicz, R.: Empirical evidence on student t log-returns of diversified world stock indices. J. Stat. Theory Pract. **2**, 233–251 (2008)
44. Krokhmal, P., Palmquist, J., Uryasev, S.: Portfolio optimization with conditional value-at-risk objective and constraints. J. Risk **4**, 43–68 (2002)
45. Xu, H.: Uniform exponential convergence of sample average random functions under general sampling with applications in stochastic programming. J. Math. Anal. Appl. **368**, 692–710 (2010)
46. Tiba, D., Zalinescu, C.: On the necessity of some constraint qualification conditions in convex programming. J. Convex Anal. **11**, 95–110 (2004)
47. Bank, B., Guddat, J., Klatte, D., Kummer, B., Tammer, K.: Nonlinear Parametric Optimization. Akademine, Berlin (1982)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.