Confidence Levels for CVaR Risk Measures and Minimax Limits*

Edward Anderson

University of Sydney Business School, University of Sydney, NSW 2006, Australia, edward.anderson@sydney.edu.au

Huifu Xu

School of Engineering and Mathematical Sciences, City University London, EC1V 0HB, London, UK, Huifu.Xu.1@city.ac.uk

Dali Zhang

Sino-US Global Logistics Institute, Shanghai Jiao Tong University, 200030, Shanghai, China, zhangdl@sjtu.edu.cn

Conditional value at risk (CVaR) has been widely used as a risk measure in finance. When the confidence level of CVaR is set close to 1, the CVaR risk measure approximates the extreme (worst scenario) risk measure. In this paper, we present a quantitative analysis of the relationship between the two risk measures and its impact on optimal decision making when we wish to minimize the respective risks measures. We also investigate the difference between the optimal solutions to the two optimization problems with identical objective function but under constraints on the two risk measures. We discuss the benefits of a sample average approximation scheme for the CVaR constraints and investigate the convergence of the optimal solution obtained from this scheme as the sample size increases. We use some portfolio optimization problems to investigate the performance of the CVaR approximation approach. Our numerical results demonstrate how reducing the confidence level can lead to better overall performance.

Key words: CVaR approximation; robust optimization; minimax; semi-infinite programming; distributional robust optimization; sample average approximation

History:

1. Introduction

A fundamental problem in financial optimization is to choose decision variables x from a compact set X in order to minimize the risk of loss, where losses are given by a continuous function $f(x,\xi)$ with ξ a random variable having a known distribution. There are many options for risk measures to use in this framework, but here we will be concerned with Conditional Value at Risk (CVaR) where CVaR_β is defined as the average value of the highest $1-\beta$ proportion of the distribution.

* The research is supported by EPSRC grant EP/J014427/1.

The parameter β here is usually referred to as the confidence level. If we use the CVaR risk measure then we obtain the problem

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

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. Essentially confidence levels β that are very close to 1 will correspond to more conservative behavior, in which we focus on more and more unlikely events. In this paper we will investigate what happens to this problem as β is increased towards 1.

It is not too hard to see that, under certain circumstances, in the limit of $\beta \to 1$ the problem $MnCV(\beta)$ approaches the minimax problem

$$MnMx: \quad \min_{x\in X}\sup_{y\in Y} f(x,y),$$

where the set $Y \subset \mathbb{R}^m$ simply corresponds to the range of the random variable ξ.

Minimax problems in this form occur in enormous numbers of applications in economics and engineering (for examples see the review on this topic by Polak (1987)). 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 (1973). 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 by Ben-Tal et al. (2009). 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 the minimum risk problem $MnCV(\beta)$ as an approximation of the minimax problem *MnMx*.

In applications we may not have access to the complete distribution of ξ, and so we will consider a situation in which the solution to $MnCV(\beta)$ must be carried out through a sampling procedure. Our particular interest is in circumstances where we cannot control the samples directly, but must accept samples drawn from the original distribution for ξ. This may happen for example when the risk optimization model generates ξ from a complex simulation. The model also applies when we have access only to a historical set of values $\xi_1, \xi_2, ..., \xi_N$ without the opportunity to sample again. In the same way for *MnMx* we suppose that, for any given value of x, the possible values $f(x,y)$ for $y \in Y$ 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 $\text{CVaR}_\beta(f(x,\xi))$ simply from looking at the average of the top $1-\beta$ of the sample values $f(x,\xi_i)$, $i = 1,2,...,N$. For values of β greater than $1-1/N$ this will mean simply looking at $\max f(x,\xi_i)$. We note that there are other more complex approaches to this problem. An attractive option with a large number of samples is to assume that we have enough samples to make the assumptions of extreme value theory hold. In that case we can assume that threshold exceedences are distributed as a Generalized Pareto Distribution and estimate the parameters of this from the sample as a route to the estimation of the CVaR value (or the absolute maximum). However our simpler approach of a sample average approximation to CVaR will be required when the sample size is not large enough for extreme value theory to be applied, and our approach is also useful in establishing some specific bounds on convergence.

It is important to remember that we aim to provide good values of the decision variables x rather than simply estimate risk. We should note that even where there is agreement that risk should be measured at a particular value of the confidence level there may be an advantage, in the sample framework, from using a lower confidence level in order to gain from the information provided by using more data points. In particular even if we are concerned about extreme risk and wish to solve the minimax problem *MnMx*, there may be an advantage in looking at a value of β low enough to include multiple samples for each value of x. To illustrate this we now consider a simple motivating example.

1.1. A simple example

We consider a small example of a minimax problem where we want to choose a scalar x to minimize $\max_{y\in Y} f(x,y)$ where the set Y is unavailable to us, but we have some sort of simulation mechanism that, for a given value of x, allows us to sample $\xi \in Y$ and evaluate $f(x,\xi)$. Thus in order to deal with the inner maximization we need to take a sample $\xi_1,\xi_2,...,\xi_N$ from Y and evaluate $\max_i f(x,\xi_i)$ in order to find an estimate of the largest $f(x,y)$ value for a given x. So we can carry out a numerical optimization process that iterates the value of x and for each trial value of x uses a large sample $\{f(x,\xi_i) : i = 1,...N\}$ in order to estimate $\max_{y\in Y} f(x,y)$.

As an alternative we can consider estimating the CVaR_β value using the sample of values $\xi_1,\xi_2,...,\xi_N$ and using these estimates to solve the minimum risk problem: $MnCV(\beta)$. This may seem odd since we know that the CVaR value will always underestimate the maximum. But the problem with the direct approach is that sampling to estimate $\max_{y\in Y} f(x,y)$ will produce erratic values that then make it hard for the optimization over x to proceed smoothly. It turns out to be better to use a CVaR approximation to the supremum.

To illustrate this we suppose that we are dealing with an underlying function $f(x,y) = 6x^2 + y$ where $y \in Y = [-5,5]$ and samples from Y are taken according to a (truncated) $N(0,1)$ distribution. Hence $\max_{y\in Y} f(x,y) = 6x^2 + 5$ and the optimal solution occurs at $x^* = 0$. However if we are forced to solve this problem numerically purely on the basis of sample evaluations $f(x,\xi_i)$ with ξ_i drawn from Y, but without direct knowledge of the set Y, then the problem is harder. Suppose that we look at 100 different x values spread from -0.5 to $+0.5$ and at each x value take 2000 ξ_i samples, then we obtain the results shown in Figure 1. It is easy to see that the erratic behavior of the maximum sampled value leads to a poor estimate of the minimax point, x^*, while the better behavior of both VaR and CVaR at the 0.95 level gives a better estimate of the right value for x^*. When we repeat this multiple times we get the results given in Figure 2 which shows the spread of x^* values found over 200 separate runs: from this it is clear that the CVaR approach produces superior estimates.

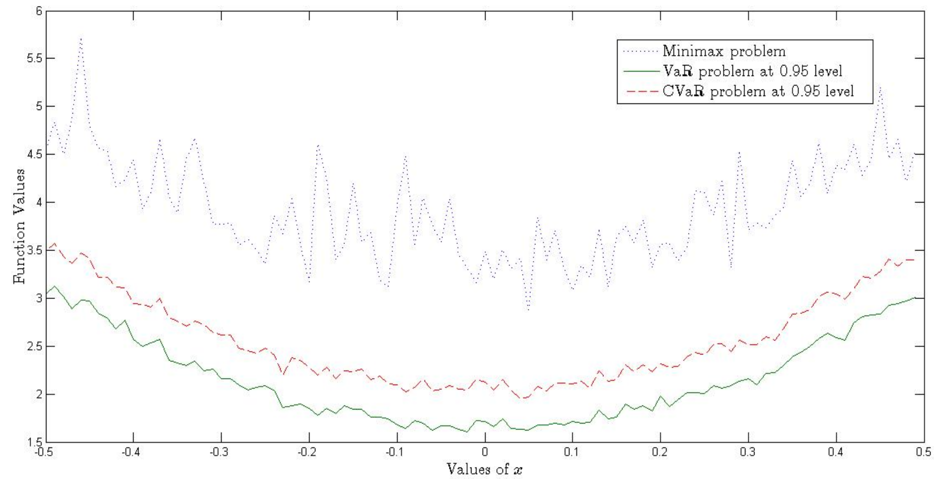


Figure 1 The minimax values, VaR and CVaR approximations of $\max_{y \in Y} f(x, y) = 6x^2 + y$.

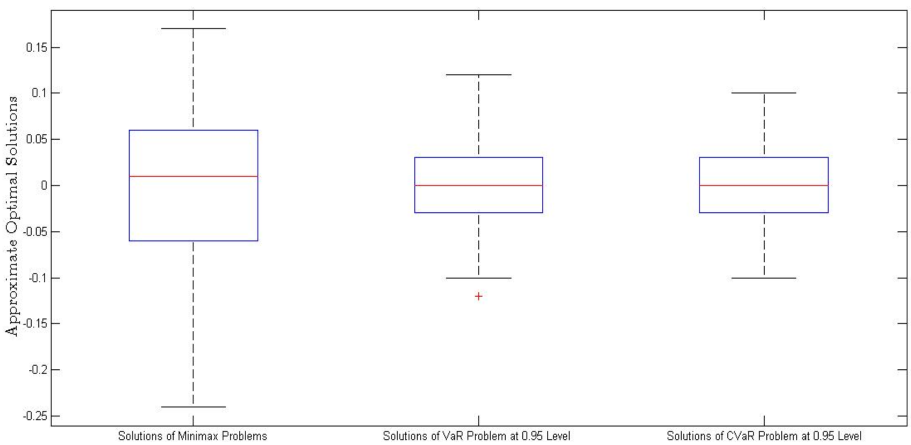


Figure 2 The distribution of solutions to $\min_{x \in X} \max_{y \in Y} f(x, y) = 6x^2 + y$ and its VaR/CVaR approximations.

1.2. The constrained risk problem

The second type of problem we consider is closely related. Alongside our study of the minimum risk problem $MnCV(\beta)$ and its minimax limit, we will also consider a constrained risk problem $MnCnCV(\beta)$, in which we seek to minimize some objective function h over $x \in X$, subject to a constraint on the maximum value U that the CVaR risk measure may take. We write this as

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

Again we can consider what happens to this problem as we vary β , reflecting a more and more conservative decision maker. Now we find that as β approaches 1 the problem approaches the following optimization problem with a semi-infinite constraint

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

where as before we define Y to be the range of the random variable ξ.

We may (redefining f 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. In other words $MnCnMx$ is an optimization problem with robust feasibility. Unless we specify otherwise we will assume that $U = 0$ in what follows. This problem has been considered by Calafiore and Campi (2005) and special cases with applications in engineering design and portfolio optimization can be found in Apkarian and Tuan (2000), Ben-Tal and Nemirovski (1997, 1998, 1999), Ghaoui and Lebret (1998).

For the constrained risk problem $MnCnCV(\beta)$ 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 99% CVaR being less than \$10,000 or minimizing h subject to the 99.5% CVaR being less than \$20,000. We will investigate the case where the solution of $MnCnCV(\beta)$ is estimated through a sample $\xi_1, \xi_2, ...\xi_N$ and this means that the choice of β value will affect the degree of inaccuracy in the estimation of the feasible region of the problem, with lower values of β giving more accurate estimations. The idea is that it may be better to opt for lower values of β in order to avoid the optimization being dependent on too small a number of samples. The need to make these kinds of decisions provides a motivation for an analysis of the convergence behavior of the solutions to $MnCnCV(\beta)$ as $\beta \to 1$.

The use of a sampling approach is also found in Calafiore and Campi (2005, 2006) and Calafiore (2010). These authors have investigated the problem $MnCnMx$ when $h(x)$ is a linear function and $f(x, \xi)$ 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 been extensively applied to various stochastic and robust programs including chance constrained stochastic programs and multistage stochastic programs, see Calafiore and Campi (2006), Calafiore (2010), Calafiore and Garatti (2008, 2010), Vayanos et al. (2012) and references therein.

1.3. Plan of the paper

Since the numerical approaches to be discussed in this paper are fundamentally based on the properties of CVaR for different values of β, we need to look more closely at the properties of the CVaR approximation. In particular, we will give a comprehensive analysis of upper and lower bounds for the value of CVaR in Theorem 1.

When the constraints for an optimization problem with robust feasibility are replaced with a CVaR approximation then there will be a change in the feasible set. Section 3 is a kind of counterpart to section 2, but instead of looking at bounds on the CVaR values we look at bounds on the sets involved.

Having completed these preliminaries in sections 2 and 3 we return to our underlying problems of robust minimax and optimization with robust constraints. In section 4 we give a discussion of how these two problems can be solved using a sample average approximation scheme and the CVaR approximation. We also explore the connection between these two problems, through considering a distributionally robust form of the minimax problem and showing how a dualization converts this to a problem of optimization with robust feasibility. Finally in section 5 we illustrate our work through applications to simple portfolio optimization problems.

2. Mathematical foundations and CVaR bounds

In this section, we will set up the problem more carefully and establish bounds for CVaR. Let $f : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ be a continuous loss function. Let X and Y be closed subsets of $\mathbb{R}^n$ and $\mathbb{R}^m$ respectively. Let $\xi : \Omega \to Y \subset \mathbb{R}^m$ be a vector of random variables defined on the probability space $(\Omega, \mathcal{F}, P)$ with support set Y. We suppose that $\xi(\omega) \in Y$ is a continuous random variable with density function $\rho(y)$, and we treat $f(x, \xi)$ for $x \in X \subset \mathbb{R}^n$ as a random variable induced by $\xi(\omega)$.

Since we are interested in the problem *MnMx* as a limit when $\beta \to 1$ and this involves the supremum of f, we need to be careful to link the set of all values that $\xi(\omega)$ can take and the density function. We make the additional assumption that the density function ρ has a non-zero integral on any ball around a point $\xi(\omega)$. This condition does no more than rule out the possibility of isolated points, since these can have an impact on $\sup_{y \in Y} f(x, y)$ but will have no effect on the value of CVaR. The formulation of the assumption is simpler when ξ is real valued: in this case we can set ρ to be the derivative of the cumulative distribution function and simply ask for ρ to be non-zero at all points $\xi(\omega)$.

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

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

Following Rockafellar and Uryasev (2000), we may define the value at risk (VaR) of $f(x,\xi)$ at a level $\beta\in(0,1)$ as

$$\mathrm{VaR}_\beta(f(x,\xi)):=\inf\{\alpha: F(x,\alpha)\ge\beta\}.$$

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

$$\mathrm{VaR}_\beta(f(x,\xi))=F_x^{-1}(\beta).$$

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

$$\mathrm{CVaR}_\beta(f(x,\xi)):=\frac{1}{1-\beta}\int_{\{y\,:\,f(x,y)\ge\mathrm{VaR}_\beta(f(x,\xi))\}}f(x,y)\rho(y)dy.$$

Since the probability that $f(x,y)\ge\mathrm{VaR}_\beta(f(x,\xi))$ 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 $\mathrm{VaR}_\beta(f(x,\xi))$ or greater. If there is an inverse function for F_x then we can write

$$\mathrm{CVaR}_\beta(f(x,\xi))=\frac{1}{1-\beta}\int_\beta^1 F_x^{-1}(t)dt,$$

see for instance (Pflug and Römisch 2007, Theorem 2.34).

The theorem below summarizes the properties of $\mathrm{VaR}_\beta(f(x,\xi))$, $\mathrm{CVaR}_\beta(f(x,\xi))$ and their relationships with $\sup_{y\in Y}f(x,y)$. Proofs of the theorems throughout the paper are given in the appendix.

Theorem 1. *Let x be some fixed element of X. Then the following assertions hold.*

(i)

$$\mathrm{VaR}_\beta(f(x,\xi))\le\mathrm{CVaR}_\beta(f(x,\xi))\le\sup_{\xi\in Y}f(x,\xi)$$

and

$$\lim_{\beta\to1}\mathrm{VaR}_\beta(f(x,\xi))=\lim_{\beta\to1}\mathrm{CVaR}_\beta(f(x,\xi))=\sup_{\xi\in Y}f(x,\xi). \tag{2}$$

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

$$\mathrm{VaR}_{(1+\beta)/2}(f(x,\xi)) \leq \mathrm{CVaR}_\beta(f(x,\xi)) \leq \frac{1}{2}\left[\mathrm{VaR}_\beta(f(x,\xi)) + \sup_{\xi\in Y} f(x,\xi)\right]. \tag{3}$$

(iii) If $\sup_{y\in Y} f(x,y) = f^*(x) < \infty$ *and there exist positive constants α_0 (depending on x), K and τ such that*

$$1 - F_x(\alpha) \geq K\left(f^*(x) - \alpha\right)^\tau, \ \textit{for all } \alpha \in (\alpha_0, f^*(x)), \tag{4}$$

holds, then for $\beta > F_x(\alpha_0)$,

$$f^*(x) - \mathrm{CVaR}_\beta(f(x,\xi)) \leq \frac{1}{K^{1/\tau}}\frac{\tau}{1+\tau}(1-\beta)^{1/\tau}. \tag{5}$$

Notice that parts (i) and (ii) of this theorem do not assume that $f(x,\xi)$ is bounded for fixed x, though this assumption is needed for part (iii). Part (i) of the theorem is enough to establish that *MnMx* is the limit of $MnCV(\beta)$ 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. However the requirements are quite minimal and will almost always be met in applications.

The condition for part (ii) that the derivative of $F_x(\alpha)$ is decreasing (not necessarily strictly) for α large enough amounts to saying that the density function of $f(x,\xi)$ has a largest mode (beyond which it is automatically decreasing). This is a very mild condition.

Condition (4) requires the cumulative distribution function to approach 1 faster than some power of the distance to $f^*(x)$. This is a natural constraint and will be available whenever the corresponding density function is bounded as $\alpha \to f^*(x)$, (even less of a restriction than part(ii) requires). We can think of τ as related to the way in which the density function of $f(x,\xi)$ approaches zero when α approaches its limit. If the density function approaches zero like an n'th power then we can set $\tau = n+1$. Cases where the density function is bounded away from zero in this region (for example, when $f(x,\xi)$ follows a uniform distribution) correspond to $\tau = 1$.

A great advantage of the CVaR risk measure is that it can be reformulated as the result of a minimization. We define

$$\Phi_\beta(x,\eta) := \eta + \frac{1}{1-\beta}\int_{y\in Y}(f(x,y)-\eta)_+\rho(y)dy \tag{6}$$

where $(t)_+ = \max(0,t)$. Rockafellar and Uryasev (2000) proved that

$$\mathrm{CVaR}_\beta(f(x,\xi)) = \min_{\eta\in\mathbb{R}} \Phi_\beta(x,\eta), \tag{7}$$

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

$$\min_{x \in X} \text{CVaR}_{\beta}(f(x,\xi)) = \min_{(x,\eta)\in X\times \mathbb{R}} \Phi_{\beta}(x,\eta). \tag{8}$$

We want to consider the problem $MnCV(\beta)$ as a route to the solution of the minimax problem *MnMx*. Traditional deterministic methods for the minimax problem would typically require $f(x,\cdot)$ to be concave for global convergence: this condition is not required.

Let $X^*(\beta)$ denote the set of optimal solutions of problem $MnCV(\beta)$ 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 $\mathbb{H}(\mathcal{C},\mathcal{D}) := \max\left(\mathbb{D}(\mathcal{C},\mathcal{D}), \mathbb{D}(\mathcal{D},\mathcal{C})\right)$ denotes the Hausdorff distance between $\mathcal{C}$ and $\mathcal{D}$. We investigate the relationship between the two sets, specifically, we estimate $\mathbb{D}(X^*(\beta), X^*)$. We do so by making use of a result from (Liu and Xu 2013, Lemma 3.8).

We will make use of the condition (4) that was introduced in Theorem 1.

Definition 1. We will say that $f(x,y)$ has *consistent tail behavior on* X if there are constants K and τ independent of x such that for each $x \in X$ we can find $\alpha_0(x) < f^*(x) = \sup_{y\in Y} f(x,y) < \infty$ with

$$1 - F_x(\alpha) \geq K\left(f^*(x) - \alpha\right)^{\tau}, \text{ for all } \alpha \in (\alpha_0(x), f^*(x)). \tag{9}$$

Let us give a simple example to explain condition (9).

Example 1. Consider $f(x,\xi) = x\xi$, where $x \in [0,1] \subset \mathbb{R}$ and ξ follows a uniform distribution over interval $[-1,1]$. Then $f^*(x) = x$ for $x \in [0,1]$. When $x = 0$, $f(0,\xi) = 0$ which is deterministic. In what follows, we consider the case when $x \in (0,1]$. It is easy to derive that for $\alpha < x$,

$$F_x(\alpha) = \frac{x+\alpha}{2x}$$

Let $K \in (0, 0.5)$ be a fixed constant and $\tau = 1$. Then (9) holds uniformly for $x \in (0,1]$ and $\alpha \in (0,x)$.

It is also convenient in this case (where $\sup_{y\in Y} f(x,y) < \infty$) to define $\Delta_{\beta}(x)$ as the error when the CVaR value is used as an approximation to the supremum:

$$\Delta_{\beta}(x) := \sup_{y\in Y} f(x,y) - \text{CVaR}_{\beta}(f(x,\xi)). \tag{10}$$

THEOREM 2. *Assume that X is a compact set and $f(x,\xi)$ has consistent tail behavior on X. Then*

$$\overline{\lim_{\beta\to 0}}\, X^*(\beta) \subset X^*. \tag{11}$$

Moreover, if the minimax robust optimization problem MnMx satisfies the second order growth condition at X^*, i.e., there exists a positive constant $K>0$ such that

$$\sup_{y\in Y} f(\hat{x},y) \geq \min_{x\in X}\left(\sup_{y\in Y} f(x,y)\right) + Kd(\hat{x},X^*)^2, \text{ for all } \hat{x}\in X, \tag{12}$$

then

$$\mathbb{D}(X^*(\beta),X^*) \leq \sqrt{\frac{3}{K}\sup_{x\in X}\Delta_\beta(x)}, \tag{13}$$

where $\Delta_\beta(x)$ is defined as in (10).

The theorem says that the optimal solution to $MnCV(\beta)$ is consistent with that of *MnMx* when β is close to 1 and under the second order growth condition the deviation of the former from the latter is bounded by the maximal difference of the objective function values of the two problems. When $f(x,\xi)$ has consistent tail behavior, it means that a small change of the confidence level from 1 will have a marginal impact on the optimal decision.

From the numerical point of view, an obvious advantage of the formulation of $MnCV(\beta)$ given by (8) is that, when f is convex in x, then it is a convex program whatever dependence f has on y. On the other hand, since Φ_β contains a nonsmooth term, it might be difficult or numerically expensive to compute the multidimensional integral when ξ has several components.

A well-known method to tackle the estimation of an expectation is sample average approximation (SAA) which is also known as the Monte Carlo method. The basic idea of SAA can be applied in this case as follows. Let $\xi^1,\cdots,\xi^N$ be an iid sampling, that is, these are independent random variables all having the same distribution to that of ξ. We construct the following SAA to $MnCV(\beta)$:

$$MnCVSA(\beta): \begin{cases} \min\limits_{x,\eta}\ \eta + \dfrac{1}{(1-\beta)N}\displaystyle\sum_{j=1}^{N}(f(x,\xi^j)-\eta)_+ \\ \text{s.t. } x\in X, \eta\in \mathbb{R}. \end{cases} \tag{14}$$

We refer to $MnCV(\beta)$ as the *true* problem and $MnCVSA(\beta)$ as its *sample average approximation.* Existence of an optimal solution of $MnCVSA(\beta)$ and asymptotic convergence of the optimal solutions as $N\to\infty$ has been well documented, see Section 5 in Xu and Zhang (2009).

3. Approximation of a semi-infinite convex inequality system

An interest in the connection between the constrained risk problem $MnCnCV(\beta)$ and the optimization with robust feasibility problem *MnCnMx* leads us to a comparison of the solutions to the following semi-infinite convex system of inequalities:

$$f(x,y) \leq 0, \text{ for all } y\in Y, \tag{15}$$

where $x \in X$, and its CVaR approximation, $\text{CVaR}_\beta(f(x,\xi)) \le 0$ which we can rewrite using the equivalence (7) as follows:

$$\min_{\eta \in \mathbb{R}} \left(\eta + \frac{1}{1-\beta} \mathbb{E}[(f(x,\xi) - \eta)_+] \right) \le 0. \tag{16}$$

In this section, we continue to assume that X is a closed convex set in $\mathbb{R}^n$, $f : \mathbb{R}^n \to \mathbb{R}$ is a continuous function and Y is a closed subset of $\mathbb{R}^m$. For each fixed $y \in Y$, $f(\cdot, y)$ is convex in $\mathbb{R}^n$. Throughout this section, we assume that the set of solutions of (15) is nonempty. Problem (15) is said to satisfy *Slater constraint qualification* if there exists a positive number $\bar{\delta}$ and a point $\bar{x} \in X$ such that

$$\max_{y \in Y} f(\bar{x}, y) \le -\bar{\delta}.$$

Our aim in this section is to understand the difference between the solution sets of the two systems, and to make some estimate of the error bound. So we will discuss the difference between $\mathcal{G}$, which we define as the solution set of (15), and $\mathcal{G}(\beta)$ which is defined as the solution set of (16) within X. It is easy to observe that $\mathcal{G}(\beta)$ provides an outer approximation of $\mathcal{G}$ (i.e. $\mathcal{G} \subset \mathcal{G}(\beta)$). 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.

THEOREM 3. *Assume that X is a compact set and for each $y \in Y$, $f(\cdot, y)$ is convex on X and $f(x,y)$ has consistent tail behavior on X. Then*

(i) for any $\epsilon > 0$, there exists a $\beta_0 \in (0,1) > 0$ such that when $\beta \in (\beta_0, 1)$,

$$\mathbb{H}(\mathcal{G}(\beta), \mathcal{G}) \le \epsilon;$$

(ii) in the case when (15) satisfies the Slater constraint qualification, then for any $\beta \in (0,1)$ there exists a positive constant C such that

$$\mathbb{H}(\mathcal{G}(\beta), \mathcal{G}) \le C \sup_{x \in X} \Delta_\beta(x)$$

where $\Delta_\beta(x)$ is given by (10).

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. Since $\Delta_\beta(x)$ can be estimated when $f(x,\xi)$ has consistent behavior, the latter result provides information on the impact on the solutions when β is reduced from 1.

Next we give an error bound for the approximating system (16) in the case that the original system (15) has a nonempty solution set.

Proposition 1. *Assume that $\mathcal{G} \neq \emptyset$, then there exists positive number β_0 and C (dependent on β) such that*

$$d(x, \mathcal{G}(\beta)) \leq C \left(\min_{\eta \in \mathbb{R}} \Phi_\beta(x, \eta) \right)_+ \tag{17}$$

for all $x \in X$ and $\beta \in (\beta_0, 1]$.

Note that the positive constant C may be estimated through Robinson's theorem in Robinson (1975). Indeed, if we let D be the diameter of $\mathcal{G}(\beta)$ and

$$\delta := -\min_{x \in \mathcal{G}} \mathrm{CVaR}_\beta(f(x, \xi)). \tag{18}$$

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 (18) may be taken over X. In the case when (15) satisfies the Slater constraint qualification, the δ value estimated from (18) 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 for (17).

4. Optimization with robust feasibility

In this section, we return to consider the constrained risk optimization problem *MnCnCV*(β) with $U = 0$. Using the characterization in terms of Φ_β this problem can be written as

$$\begin{aligned} &\min_{x \in X} \ h(x) \\ &\text{s.t.} \ \min_{\eta \in \mathbb{R}} \ \Phi_\beta(x, \eta) \leq 0, \end{aligned} \tag{19}$$

where $\Phi_\beta(x, \eta)$ is defined as in (6). For the simplicity of discussion, we assume throughout this section that X is a compact set.

Note that the constraint $\min_{\eta \in \mathbb{R}} \Phi_\beta(x, \eta) \leq 0$, of (19) can also be written $\sup_{y \in Y} f(x, y) \leq \Delta_\beta(x)$. Thus we can see that the formulations (19) and *MnCnCV*(β) are equivalent to a relaxation of the problem *MnCnMx*, so that the constraints become

$$f(x, \xi) \leq \Delta_\beta(x), \ \forall \xi \in Y.$$

Theorem 1 gives a route to some bounds on $\Delta_\beta(x)$ which may be used to explore the extent of this relaxation of the constraints.

Our focus in this section will be on the numerical solution of problem *MnCnMx* through an approximation by *MnCnCV*(β). To do this we need to carry out some quantitative analysis of the approximation in terms of optimal value and optimal solutions, and so we will look at the way that solutions to the problem (19) 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 that X is a compact set,* MnCnMx *satisfies the Slater constraint qualification and $f(x,y)$ has consistent tail behavior on X. 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) there exists a positive constant $\hat{C}$ such that

$$|\hat{\vartheta}(\beta) - \hat{\vartheta}| \leq \hat{C} \sup_{x\in X} \Delta_\beta(x), \tag{20}$$

where $\Delta_\beta(x)$ is defined in (18).

Parts (i) and (ii) ensure consistency of optimal value and optimal solution when β converges to 1 while Part (iii) quantifies the difference of the two optimal values.

We now return to our discussion of numerical methods for solving problem *MnCnCV*(β). We begin with the formulation (19) and observe that we can rewrite this in a way that treats η as a variable:

$$\begin{aligned} \min_{x\in X, \eta\in \mathbb{R}} \;& h(x) \\ \text{s.t.} \;& \eta + \frac{1}{1-\beta}\mathbb{E}[(f(x,\xi)-\eta)_+] \leq 0, \end{aligned} \tag{21}$$

To see the equivalence, note that the constraint in (21) is simply $\Phi_\beta(x,\eta) \leq 0$. Let x^* be an optimal solution to problem (19) and v^* be the optimal value, let $(\hat{x},\hat{\eta})$ be an optimal solution to problem (21) and $\hat{v}$ be the optimal value. Then there exists a finite η^* (see Rockafellar and Uryasev (2000)) such that

$$\Phi_\beta(x^*,\eta^*) = \min_{\eta\in\mathbb{R}} \Phi_\beta(x^*,\eta) \leq 0,$$

which means (x^*,η^*) is a feasible solution to (21). This shows $v^* \geq \hat{v}$. Conversely, since

$$\min_{\eta\in\mathbb{R}} \Phi_\beta(\hat{x},\eta) \leq \Phi(\hat{x},\hat{\eta}) \leq 0,$$

it means $\hat{x}$ is a feasible solution to (19). This shows $\hat{v} = h(\hat{x}) \geq v^*$. The equivalence follows.

Problem (21) is a nonlinear stochastic optimization problem with deterministic objective and a stochastic constraint. The main challenge here is to handle the expected value $\mathbb{E}[(f(x,\xi)-\eta)_+]$. This motivates us to consider sample average approximation as in the previous subsection. Let $\xi^1,\cdots,\xi^N$ be an independent and identically distributed (i.i.d.) sampling of ξ. We consider the following sample average approximation for *MnCnCV*(β) (using formulation (21)):

$$\mathit{MnCnCVSA}(\beta): \begin{cases} \min\limits_{x,\eta} \; h(x) \\ \text{s.t. } \eta + \dfrac{1}{(1-\beta)N}\displaystyle\sum_{j=1}^N (f(x,\xi^j)-\eta)_+ \leq 0, \\ \qquad x\in X, \eta\in\mathbb{R}. \end{cases}$$

For a fixed sample, *MnCnCVSA*(β) is a deterministic nonlinear programming (NLP) and therefore any appropriate NLP code can be applied to solve the problem.

At this point, it is helpful to link our SAA approach to Calafiore and Campi's randomization approach in Calafiore and Campi (2005, 2006) in which the following problem is taken as a sample-based approximation to *MnCnMx*:

$$\begin{array}{ll} \min_{x\in X} & h(x) \\ \text{s.t.} & f(x,\xi^j)\le 0, \text{ for } j=1,\cdots,N, \end{array} \tag{22}$$

where $\xi^1,\cdots,\xi^N$ 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, the latter provides an outer approximation to the feasible set of the true problem, and hence the optimal value of (22) gives rise to a lower bound for the optimal value of *MnCnMx*. From a practical point of view, an important issue of this kind of approximation scheme concerns feasibility of the optimal solution of (22) 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 Calafiore and Campi (2005) and Calafiore and Garatti (2008). Moreover, Calafiore and Garatti (2008) demonstrated an exact universal bound on the number of samples required to ensure that only a small portion of the constraints of the original problem is violated.

Analogous to (22), our approximation scheme *MnCnCVSA*(β) 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. The main differences can be summarized as follows.

1. Writing the constraints of (22) as $\max_{j=1}^N f(x,\xi^j)\le 0$, we can see that the difference between $\eta+\dfrac{1}{(1-\beta)N}\sum_{j=1}^N (f(x,\xi^j)-\eta)_+$ and $\max_{j=1}^N f(x,\xi^j)$ is that the former uses a few samples at the tail of the distribution of $f(x,\xi)$ whereas the latter only uses the largest sample value of $f(x,\xi)$; the averaging effect will make the former change less drastically as the sample changes.

2. Problem *MnCnCVSA*(β) always satisfies the Slater constraint qualification as long as *MnCnMx* is feasible, whereas (22) might not in the case when *MnCnMx* fails the Slater condition.

3. There is a single nonsmooth convex constraint in *MnCnCVSA*(β) whereas (22) has N smooth constraints.

Let (x^N,η^N) be an optimal solution which is obtained from solving the sample average approximation *MnCnCVSA*(β) with sample size N. In the next result we estimate the probability of a significant violation, that is the probability of x^N deviating from the feasible set $\mathcal{G}$ for problem *MnCnMx* by more than an amount ϵ as we take more and more samples ($N\to\infty$).

THEOREM 5. *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 nonempty, (b) $f(x,\xi)$ is convex* *as a function of x, (c) we may choose a measurable function $\kappa(\xi)$ as the Lipschitz constant for f, so that*

$$|f(x',\xi) - f(x,\xi)| \le \kappa(\xi)\|x' - x\|$$

for all $\xi \in Y$ and all $x', x \in X$, and the moment generating function of $\kappa(\xi)$ (as a function of t) is finite for t in a neighborhood of zero, (d) for every x the moment generating function for $f(x,\xi)$, 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

$$R(\delta) := \inf_{x\in\mathcal{G}(\beta), d(x,\hat{X}(\beta))\ge\delta} h(x) - \hat{\vartheta}(\beta) > 0 \tag{23}$$

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

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

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 and $f(x,\xi)$ has consistent tail behavior on X, then for any $\varepsilon > 0$ there exist positive constants β_0, $C(\epsilon)$ and $\alpha(\epsilon)$ (independent of N) such that for $\beta > \beta_0$*

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

for N sufficiently large.

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 $\kappa(\xi)$ and $f(x,\xi)$ 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 $f(x,\xi)$ as a function of x. These conditions are standard for deriving exponential rate of convergence, see for example Shapiro and Xu (2008). 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 Klatte (1987).

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 is r_N-feasible where

$$r_N := \max_{y\in Y} f(x^N, y) - \max_{\eta}\left(\eta + \frac{1}{(1-\beta)N}\sum_{j=1}^{N}(f(x^N,\xi^j) - \eta)_+\right).$$

4.1. Application in distributional robust optimization

The uncertain convex program models and numerical schemes presented in the preceding discussions may be applied to distributional robust optimization when the distributional set is specified through moment conditions. Here we outline how this may be done but we will not go into much detail as it is not the main focus of this paper. To simplify the discussion, let us consider the following distributional robust optimization problem:

$$\min_{x\in X}\max_{P\in\mathcal{P}} \mathbb{E}_P[f(x,\xi(\omega))] \tag{24}$$

where X is a nonempty closed convex set of $\mathbb{R}^n$, $f:\mathbb{R}^n\times\mathbb{R}^k\to\mathbb{R}$ is a continuous function, $\xi(\omega)$ is a random vector defined on a probability space $(\Omega,\mathcal{G},P)$ with support set $\Xi\subset\mathbb{R}^d$, and $\mathcal{P}$ is a set of probability distributions. In this model we are assuming that the true probability distribution P of ξ is unknown. The only information available is that the distribution P is located in set $\mathcal{P}$.

Assume that $\mathcal{P}$ is specified through first order moment conditions:

$$\mathcal{P}:=\left\{P: \begin{array}{l}\mathbb{E}_P[\psi_s(\xi(\omega))]=\mu_s, \text{ for } s=1,\cdots,p\\ \mathbb{E}_P[\psi_s(\xi(\omega))]\le\mu_s, \text{ for } s=p+1,\cdots,q\end{array}\right\}, \tag{25}$$

where $\psi_i:\mathbb{R}^k\to\mathbb{R}$, $i=1,\cdots,q$, are measurable functions (In general, ψ does not have to be a function of ξ.) In other words, we know the random variable ξ satisfies certain moment conditions albeit its true distribution is unknown.

This is a classical moment problem. To simplify our discussion, we assume ψ_i, $i=1,\cdots,q$, are continuous functions, Ω is a compact subset of a finite dimensional space and f is convex in x for every ξ. Under these circumstance, we can reformulate (24), by taking the dual of its inner maximization through Shapiro's duality theorem (Shapiro 2001, Proposition 3.1), as:

$$\begin{array}{cl}\min\limits_{x\in X,\lambda_0,\cdots,\lambda_q} & \lambda_0+\sum\limits_{s=1}^q\lambda_s\mu_s\\ \text{s.t.} & \lambda_s\ge 0, \text{ for } s=p+1,\cdots,q,\\ & f(x,\xi(\omega))\le\lambda_0+\sum\limits_{s=1}^q\lambda_s\psi_s(\xi(\omega)), \text{ a.e. } \omega\in\Omega,\end{array} \tag{26}$$

which is a semi-infinite programming problem. To simplify the exposition, let

$$g(x,\lambda_0,\cdots,\lambda_q,\xi(\omega)):=f(x,\xi(\omega))-\lambda_0-\sum_{s=1}^q\lambda_s\psi_s(\xi(\omega)), \text{ a.e. } \omega\in\Omega.$$

Thus (26) can be written as

$$\begin{array}{cl}\min\limits_{x\in X,\lambda_0,\cdots,\lambda_q} & \lambda_0+\sum\limits_{s=1}^q\lambda_s\mu_s\\ \text{s.t.} & \lambda_s\ge 0, \text{ for } s=p+1,\cdots,q,\\ & \sup\limits_{\xi\in\Xi} g(x,\lambda_0,\cdots,\lambda_q,\xi)\le 0.\end{array} \tag{27}$$

This is a convex optimization problem with robust feasibility. The CVaR approximation of the problem is:

$$\begin{array}{ll} \min\limits_{x\in X,\lambda_0,\cdots,\lambda_q} & \lambda_0+\sum\limits_{s=1}^{q}\lambda_s\mu_s \\ \text{s.t.} & \lambda_s\geq 0,\ \ \text{for } s=p+1,\cdots,q, \\ & \text{CVaR}_\beta(g(x,\lambda_0,\cdots,\lambda_q,\xi))\leq 0. \end{array} \tag{28}$$

Let $\xi^1,\cdots,\xi^N$ be an independent and identically distributed (i.i.d.) sampling of ξ. We may construct a SAA scheme for the problem:

$$\begin{array}{ll} \min\limits_{x\in X,\eta\in\mathbb{R},\lambda_0,\cdots,\lambda_q} & \lambda_0+\sum\limits_{s=1}^{q}\lambda_s\mu_s \\ \text{s.t.} & \lambda_s\geq 0,\ \ \text{for } s=p+1,\cdots,q, \\ & \eta+\dfrac{1}{(1-\beta)N}\sum\limits_{j=1}^{N}(g(x,\lambda_0,\cdots,\lambda_q,\xi^j)-\eta)_+\leq 0. \end{array} \tag{29}$$

The underlying argument of this formulation can be explained as follows: we don't know the distribution of random variable ξ but know it satisfies certain moment conditions and it is possible to obtain some empirical data of ξ. Using the moment condition and the empirical data, we may construct a SAA scheme for (24). Thus (29) provides an approximation scheme which combines moments and Monte Carlo sampling for the classical distributional robust optimization problem.

Note that the same approach can also be applied to an optimization problem with a distributional form for the robust constraint

$$\begin{array}{l} \min\limits_{x\in X}\ h(x) \\ \text{s.t. } \max\limits_{P\in\mathcal{P}}\mathbb{E}_P[f(x,\xi(\omega))]\leq 0, \end{array}$$

for which we can construct a SAA of the CVaR approximation through dualization:

$$\begin{array}{ll} \min\limits_{x\in X,\eta\in\mathbb{R},\lambda_0,\cdots,\lambda_q} & h(x) \\ \text{s.t.} & \lambda_0+\sum\limits_{s=1}^{q}\lambda_s\mu_s\leq 0 \\ & \lambda_s\geq 0,\ \ \text{for } s=p+1,\cdots,q, \\ & \eta+\dfrac{1}{(1-\beta)N}\sum\limits_{j=1}^{N}(g(x,\lambda_0,\cdots,\lambda_q,\xi^j)-\eta)_+\leq 0. \end{array}$$

5. Numerical Experiments

In this section, we consider a robust portfolio optimization problem and formulate it as a *MnMx* problem where the maximum is taken as occurring at the worst scenario of the underlying random variables. Using the approach of section 3, we approximate the maximum by CVaR and then solve the minimization as *MnCV*(β), through SAA. We compare the results of this approximation scheme with a direct approach to the *MnMx* problem. We also consider a risk constrained portfolio optimization problem using the same numerical example. So we start with a *MnCnMx* formulation of the portfolio optimization problem under (semi-infinite) risk constraints and consider reformulating the constraint through a CVaR approximation, as we have discussed in section 4. This leads to a *MnCnCV*(β) problem. Again we will compare the performance of portfolios generated through this approximation scheme using different values of the confidence level β.

5.1. Minimax formulation

5.1.1. Problem Setting Consider a market consisting of n stocks. The investment period starts at $t=0$ and finishes at time $t=T$, which is the end of the investment horizon. In the investment portfolio problem, an investor decides on a vector of weights $\mathbf{w} \in \mathbb{R}^n$, whose elements add up to 1 with the i'th component w_i being the proportion of total wealth to be invested in the i'th stock at time $t=0$ for $i=1,2,\cdots,n$. Let $\tilde{r}$ denote the random vector of the stock returns over the investment horizon, where the components $\tilde{r}_i$, $i=1,2,\cdots,n$, take nonnegative values. By definition, the investor will receive $\tilde{r}_i$ dollars at time T for every dollar invested in stock i at time 0. Here, we use the notation $\mathcal{R} \subset \mathbb{R}^n$ to denote the support set of the random variable $\tilde{r}$. We do not need to specify the overall capital to be invested since this just acts to scale up the returns.

In the robust framework we consider, the return vector $\tilde{r}$ over the investment horizon remains unknown but is believed to lie within support set $\mathcal{R}$. To immunize the portfolio against the inherent uncertainty in the investment horizon, we maximize the worst-case portfolio return, where the worst-case is calculated with respect to all asset returns in $\mathcal{R}$, and can be formulated as a max-min problem

$$\begin{array}{ll} \max\limits_{\mathbf{w}\in\mathbb{R}^n} \min\limits_{\tilde{r}\in\mathcal{R}} & \mathbf{w}'\tilde{r} \\ \text{s.t.} & \mathbf{w}'\mathbf{1} = 1, \\ & \mathbf{w}^{\mathbf{L}} \leq \mathbf{w} \leq \mathbf{w}^{\mathbf{U}}. \end{array} \tag{30}$$

where the components in the n-dimensional vectors $\mathbf{w}^{\mathbf{L}} := (0,0,\cdots,0)'$, $\mathbf{w}^{\mathbf{U}} := (1,1,\cdots,1)'$ are used to denote the lower and upper bounds of investment weight for the corresponding stock w_i for $i=1,2,\cdots,n$, and $\mathbf{1} \in \mathbb{R}^n$ denotes a vector of ones with an appropriate dimension. Here, we define the feasible set $W_n \subset \mathbb{R}^n$ for problem (30) as $W_n := \{\mathbf{w} \,|\, \mathbf{w}'\mathbf{1} = 1, \mathbf{w}^{\mathbf{L}} \leq \mathbf{w} \leq \mathbf{w}^{\mathbf{U}}\}$.

We can reverse the objective function to $\min_{\mathbf{w}\in\mathbb{R}^n} \max_{\tilde{r}\in\mathcal{R}} -\mathbf{w}^T\tilde{r}$ so that this is a *MnMx* problem and then use *MnCV*(β) as an approximation to (30) with parameter β, that is,

$$\min_{\mathbf{w}\in W_n} \text{CVaR}_\beta(-\mathbf{w}^T\tilde{r}).$$

In the numerical tests, we study the portfolio optimization problems over the stocks in the FTSE-100 index. Due to the limited availability of historical prices, we removed five of these stocks from our investigation giving a total of 95 stocks that are available in the portfolio optimization problems, i.e. $n=95$.

We denote the return by $\tilde{r} = (\tilde{r}_1, \tilde{r}_2, \cdots, \tilde{r}_n)'$ with components corresponding to the 95 stocks and we use the vector $\mathbf{w} := (w_1, w_2, \cdots, w_n)'$ to denote the portfolio split between these 95 stocks. The corresponding *MnMx* problem can be written as

$$\min_{\mathbf{w} \in W_{95}} \max_{\tilde{r} \in \mathcal{R}} -\sum_{i=1}^{95} w_i \tilde{r}_i. \tag{31}$$

The *MnCV(β)* version of this problem can be written as

$$\min_{\mathbf{w} \in W_{95}} \text{CVaR}_\beta \left(-\sum_{i=1}^{95} w_i \tilde{r}_i \right). \tag{32}$$

We will look at a case where the investor is concerned with the portfolio return over a five day time horizon and where the only information available to the investor about possible returns $\mathcal{R}$ arises from historical data. In this model, we use a set of data over a four-year period spanning from Dec 2008 to Nov 2013 including a total of 1200 records on the historical stock returns (these are obtained from http://finance.google.com with adjustment for stock splitting).

5.1.2. Test on real data In the first set of numerical tests, we consider a case where the investor makes an optimal decision based on the *MnMx* problem using historical data between Dec 2008 and Jul 2011 (with sample size being 895). Let $\tilde{r}^j := \left(\tilde{r}_1^j, \tilde{r}_2^j, \cdots, \tilde{r}_{95}^j\right)'$ and $\tilde{\mathcal{R}} := \{\tilde{r}^j, j = 1, 2, \cdots, 895\}$ denote the data on returns (based on a five day rolling time horizon). By replacing $\mathcal{R}$ with the data set, we may rewrite (31) as

$$\min_{\mathbf{w} \in W_{95}} \max_{\tilde{r}^j \in \tilde{\mathcal{R}}} -\sum_{i=1}^{95} w_i \tilde{r}_i^j.$$

Solving the problem, we obtain an optimal solution to (31) in Figure EC.1.

We have also applied the SAA of *MnCV(β)* to solve (31), that is,

$$\min_{\mathbf{w} \in W_{95}, \eta \in \mathbb{R}} \eta + \frac{1}{(1-\beta)N} \sum_{j=1}^{N} \left(\left(-\sum_{i=1}^{95} w_i \tilde{r}_i^j \right) - \eta \right)_+ \tag{33}$$

where $N = 895$ and parameter β ranges from 0.95 to 0.995.

The tests are carried out in MATLAB 7.2 installed in a PC with Windows XP where the optimization problems in *MnMx* and its *MnCV(β)* approximation are solved by the built-in optimization solver *Linprog* with 'interior-point' algorithm by transferring *MnCV(β)* into linear programming in Krokhmal et al. (2002). The numerical results for the optimal values of (33) and (31) are displayed in Figure 3. The detailed values of the weights w_i for different values of β are given in Figure EC.1 in the electronic companion. It is easy to observe the convergence of the optimal solutions and optimal values of $MnCV(\beta)$ problems to $\mathbf{w}^*$ and the optimal value of $MnMx$ denoted by f^* as

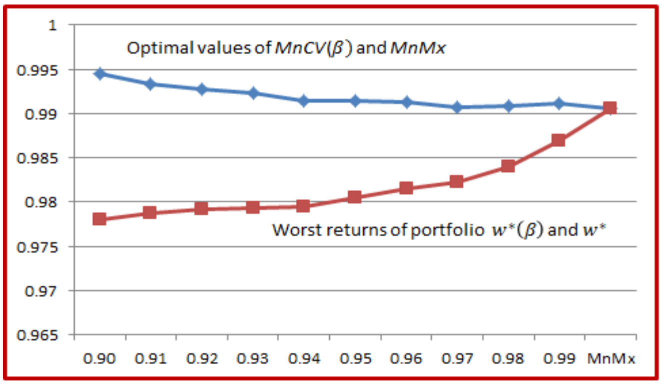


Figure 3 Optimal values and worst returns of $MnCV(\beta)$ and $MnMx$ portfolios with real data.

β tends to 1. Note that there was no significant difference in the computation times required to solve *MnMx* and *MnCV*(β).

Next we want to test the quality of the *MnCV*(β) approximation to the *MnMx* solution on 'out of sample' data. The idea parallels the motivating example given in the introduction. In that example we found that using a *MnCV*(β) approximation provided a better solution x^* ($\mathbf{w}^*$ in this example) for the underlying (unobserved) *MnMx* problem than simply taking the maximum of the sampled set at different x ($\mathbf{w}$ in this example) values. Here we ask how a set of portfolio weights based on a historical data window of two years (more precisely 500 days) will perform if the aim is to maximize the worst case result over the next quarter (more precisely 50 days). The idea is that, by using more information from the sample of different results over 500 days, the *MnCV*(β) approximation may do a better job than the straightforward *MnMx* solution in selecting a portfolio which will avoid a bad performance over the next 50 days. However in deliberately choosing the "wrong" objective we may also expect some degradation of performance: the lower line in Figure 3 shows this for the in-sample data where optimizing with lower beta values leads to lower values for the worst case return.

To reduce the effects of chance factors we repeat the test with a rolling window, always using 500 days to determine the weights and then checking the worst performance over the next 50 days. More than 1050 data points then allow for a total of 500 experiments, as we let the start date move from day 1 to day 500, and the start of the test window of 50 days moves from 501 to 1000. In each experiment we compare the performance of the solution to *MnMx* problem (31) and its *MnCV*(β) approximation of (33).

The results of these experiments are shown in Figure 4, where the graphs give the worst return rate over a 50 day period for the investment of the *MnMx* portfolio of (31) and its *MnCV*(β) approximate portfolio of (33). The comparison is not clearcut with periods where the *MnCV*(β) approximate portfolio achieves a better worst case return during the 50 day window, and periods

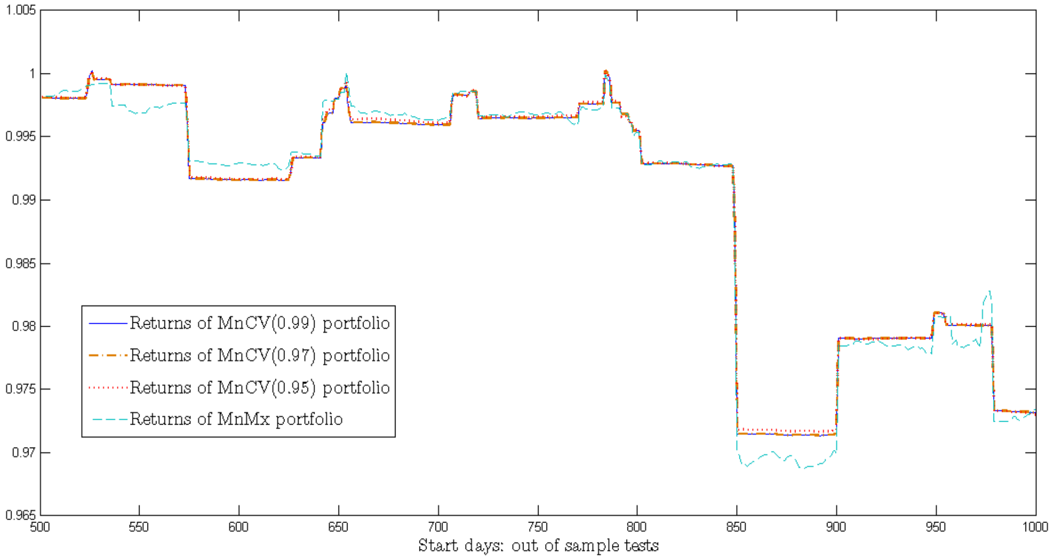


Figure 4 The worst return scenario of $MnMx$ portfolio and $MnCV(\beta)$ portfolios.

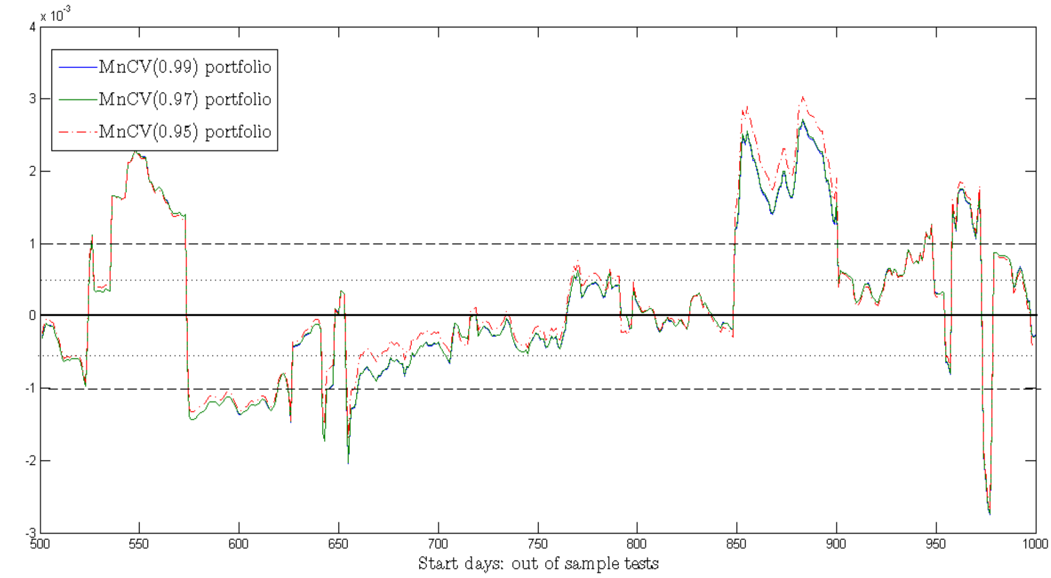


Figure 5 The differences between worst return rates of $MnMx$ portfolio and $MnCV(\beta)$ portfolios.

where the reverse is true. Notice that the rolling procedure through which these experiments are generated often gives rise to a portfolio choice for $MnCV(\beta)$ that is constant for a period of time and a worst scenario that occurs in a whole set of adjacent 50-day windows. This explains the horizontal sections that appear in the graphs of Figure 4. Figure 5 gives a plot of the difference between the $MnMx$ portfolio and its approximation solved from $MnCV(\beta)$. From this we can see that there is a very slight advantage in using a β value of 0.95 over against 0.97 or 0.99. The average worst case returns are as follows: 0.9895 for the $MnMx$ portfolio, and 0.9898, 0.9897, and 0.9897 for the $MnCV(\beta)$ portfolios with $\beta = 0.95, 0.97$ and 0.99 respectively. This single example, with relatively small differences observed, shows that the advantage of using a lower value of β can more than compensate for the change in the objective function and gives some support for the approach we have proposed, but leaves open the question of whether the $MnCV(\beta)$ approach will be superior in general.

5.1.3. Test on modified data The experiment reported above on real data is not very conclusive. We have also carried out the numerical tests with synthetic data. We consider a situation where actual historical data $\tilde{\mathcal{R}}$ is used to construct a set of distributions and we then check the performance of different approaches using data drawn from these distributions.

We will model stock returns using log-normal distributions with parameters set to match the historical data. Specifically we suppose that the components of the return vector are independent and each follows a log-normal distribution, denoted by LogNormal(m_i, σ_i) for $i = 1, 2, \cdots, 95$. Here m_i and σ_i are the mean and standard deviation of the stock i for $i = 1, 2, \cdots, 95$ respectively, where the values of m_i and σ_i are estimated from the historical data used in Section 5.1.2. The assumption of independence is clearly very different from the real data, but we do not think that in the present context independence or lack of it will favour portfolios generated either from the $MnMx$ problem of from the $MnCV(\beta)$ approximation scheme.

We will use synthetic data sampled from the log-normal distributions and this will help us in exploring the behavior of the $MnCV(\beta)$ approximation schemes. Notice though that the log-normal distribution implies a minimum return of 0, so the underlying minimax problem implied by the $MnMx$ problem (31) is not very meaningful.

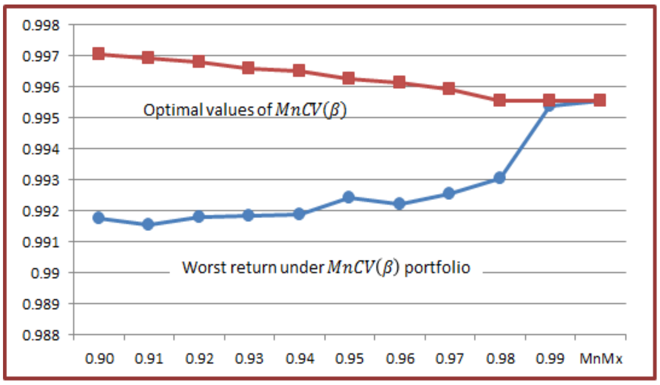


Figure 6 Optimal values and worst returns of $MnCV(\beta)$ and $MnMx$ portfolios with modified data.

Using random samples of size 5000, we obtain optimal solution and optimal return of *MnMx* portfolio problem (31). Similar tests were implemented to *MnCV(β)* optimization problem (33) and the results are displayed in Figure 6 (the details of these solutions are shown in Figure EC.2 in the electronic companion). We can see similar convergence behavior with this synthetic data to that we observed for the real data.

5.1.4. The distribution of solutions When we solve (31) and (33), the solutions change for different samples (even though these samples are generated from an identical distribution). We expect that lower values of the confidence level β will lead to more consistent solutions, as a greater part of the distribution of returns is used. Here, we present some numerical results to test this conjecture through looking at the distributions of solutions to (31) and (33) with the samples from the same log-normal distribution in Section 5.1.3.

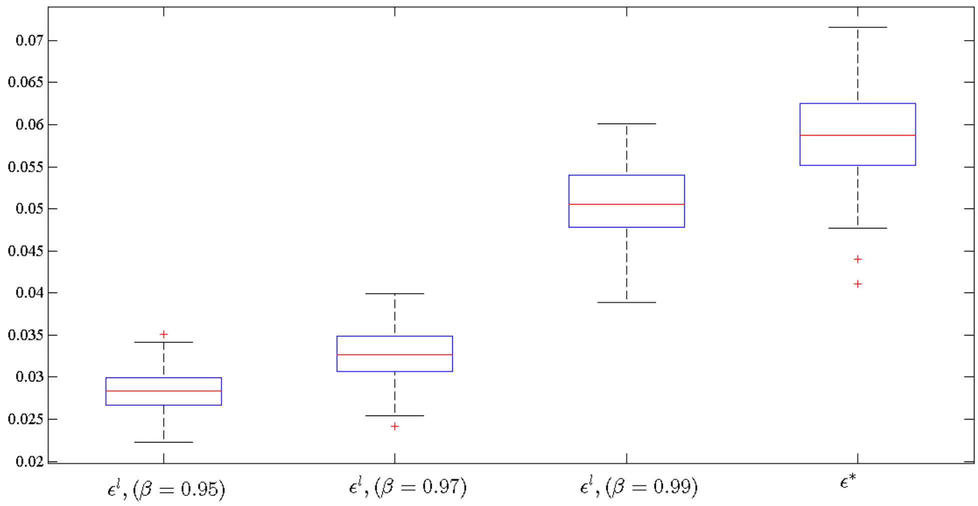


Figure 7 The box diagram for the deviations of ϵ^l and $\epsilon^{*,l}$.

To this end, we generate L samples each of size N, denoted by $\tilde{R}_l := \{\tilde{r}^{j,l}, j = 1, 2, \cdots, N\}$, $l = 1, 2, \cdots, L$. By using these samples, we will obtain L different solutions to the portfolio optimization problem, denoted by $w^{*,l} := (w_1^{*,l}, w_2^{*,l}, \cdots, w_{95}^{*,l})'$ to (31) and L solutions to the $MnCV(\beta)$ problems (33), denoted by $w^l := (w_1^l, w_2^l, \cdots, w_{95}^l)'$ for $l = 1, 2, \cdots, L$. Let $\bar{w}_i := (1/L)\sum_{l=1}^{L} w_i^l$ and $\bar{w}_i^* := (1/L)\sum_{l=1}^{L} w_i^{*,l}$ denote the mean value of each set of solutions. We consider the distribution of deviations (measured as the sum of the absolute values of the component differences) from a single solution w^l to the mean $\bar{w}_i$, that is,

$$\varepsilon^l := \sum_{i=1}^{95} |w_i^l - \bar{w}_i|$$

and the corresponding distribution of deviations from a single solution $w^{*,l}$ to the mean $\bar{w}_i^*$ given by

$$\varepsilon^{*,l} := \sum_{i=1}^{95} |w_i^{*,l} - \bar{w}_i^*|.$$

We look at the case with $L = 200$ and $N = 5000$. Figure 7 gives the distributions of the sets $\{\varepsilon^l, l = 1, 2, \cdots, L\}$ and $\{\varepsilon^{*,l}, l = 1, 2, \cdots, L\}$ in a box diagram. We can observe from Figure 7 that, as we expected, the $MnCV(\beta)$ problems yield a set of approximate optimal portfolios with less variance induced by the sampling as β is reduced. The greatest variance occurs when we consider optimal portfolios arising from the $MnMx$ problems.

The more consistent behavior we observe from the solution to $MnCV(\beta)$ in comparison to the solution to the $MnMx$ problem is similar to the behavior we see in the simple example of Figure 1. We can expect this to result in better performance out of sample when we are trying to maximize the worst portfolio return.

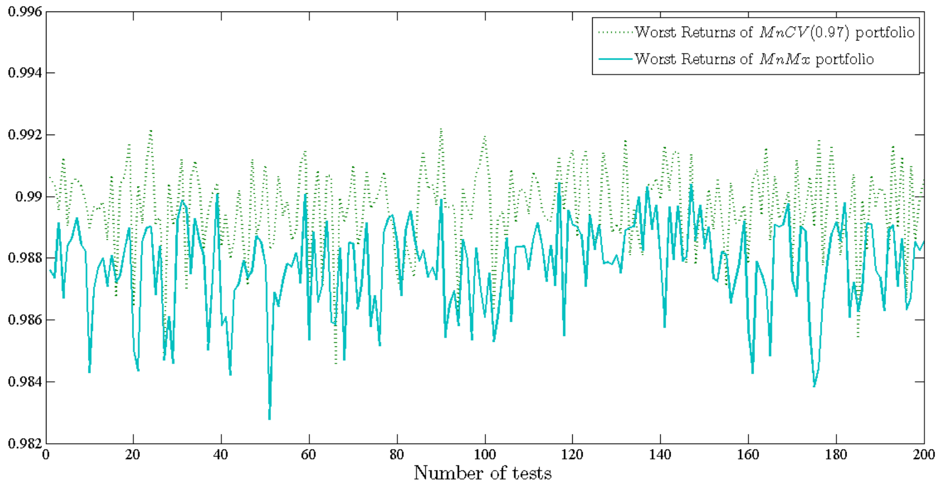


Figure 8 The worst return scenario of $MnMx$ portfolio and $MnCV(0.97)$ portfolios.

With this simulated data we no longer need to use the rolling window approach of section 5.1.2. Figure 8 shows the result of finding an optimal portfolio on one set of 5000 days of data and then applying this portfolio to another set of 5000 data points. We do this sequentially, generating 200 different sets of 5000 days of data. We can see that the worst returns observed out of sample for the portfolios generated from $MnMx$ are not as good as the worst returns observed out of sample for the $MnCV(0.97)$ portfolios.

In fact the average of the worst returns for the $MnMx$ portfolios is 0.9877 and this is improved to 0.9896 using the $MnCV(0.97)$ portfolios. A value of $\beta = 0.97$ is not critical here; we get very similar results with $\beta = 0.95$ and $\beta = 0.99$.

5.2. Constrained risk problem

In this section, we consider a *MnCnMx* formulation of the portfolio optimization problem, corresponding to the problem of optimizing the expected return over a set of portfolios subject to a constraint on the minimum value U that the worst return scenario may take.

The constrained robust counterpart of (30) can be written as

$$\begin{aligned} \max_{\mathbf{w}\in W_{95}} \ & \mathbb{E}[\mathbf{w}'\tilde{r}] \\ \text{s.t.} \ & \mathbf{w}'\tilde{r} \geq U, \quad \text{for all } \tilde{r}\in\mathcal{R}, \end{aligned} \tag{34}$$

where random variable $\tilde{r}$ represents the return vector of the stocks with support set $\mathcal{R}$. In (34), $\mathbf{w^L}$ and $\mathbf{w^U}$ are defined the same as their counterparts in Section 5.1. We can reverse the objective function to $\min_{w\in\mathbb{R}^n} -\mathbf{w}'\tilde{r}$ and the first constraints to $-\mathbf{w}'\tilde{r} \leq -U$ so that this is a *MnCnMx* problem. Then the *MnCnCV*(β) approximation to (34) with parameter β, is given by

$$\begin{aligned} \min_{\mathbf{w}\in W_{95}} \ & -\mathbb{E}[\mathbf{w}'\tilde{r}] \\ \text{s.t.} \ & \text{CVaR}_\beta(-\mathbf{w}'\tilde{r}) \leq -U. \end{aligned} \tag{35}$$

In the numerical tests, we study the portfolio optimization problems over the same stocks and their historical prices investigated in Section 5.1. We denote the return by $\tilde{r} = (\tilde{r}_1, \tilde{r}_2, \cdots, \tilde{r}_n)'$ and use the vector $\mathbf{w} := (w_1, w_2, \cdots, w_n)'$ to denote the split between these 95 stocks. The corresponding *MnCnMx* problem can be written as

$$\begin{aligned} \min_{\mathbf{w}\in W_{95}} \ & -\mathbb{E}\left[\sum_{i=1}^{95} w_i\tilde{r}_i\right] \\ \text{s.t.} \ & -\sum_{i=1}^{95} w_i\tilde{r}_i \leq -U, \quad \text{for all } \tilde{r}\in\mathcal{R}. \end{aligned} \tag{36}$$

The corresponding *MnCnCV*(β) can be written as

$$\begin{aligned} \min_{\mathbf{w}\in W_{95}} \ & -\mathbb{E}\left[\sum_{i=1}^{95} w_i\tilde{r}_i\right] \\ \text{s.t.} \ & \text{CVaR}_\beta\left(-\sum_{i=1}^{95} w_i\tilde{r}_i\right) \leq -U, \quad \text{for all } \tilde{r}\in\mathcal{R}. \end{aligned} \tag{37}$$

In the same way as for Section 5.1 we will look at a case where the investor is concerned with the portfolio return over a five day time horizon and uses information from historical data to generate the required portfolio.

5.2.1. Test on real data We first consider a set of numerical tests where the investor makes an optimal decision based on the *MnCnMx* problem using historical data between Dec 2008 and Jul 2011 (with sample size being 895). Let $\tilde{r}^j := \left(\tilde{r}_1^j, \tilde{r}_2^j, \cdots, \tilde{r}_{95}\right)'$ and $\tilde{\mathcal{R}}$ denote the data on returns (based on a five day rolling time horizon). We may rewrite (36) as

$$\begin{aligned} \min_{\mathbf{w}\in W_{95}} \ & -\frac{1}{N}\sum_{j=1}^{N}\sum_{i=1}^{95} w_i\tilde{r}_i^j \\ \text{s.t.} \ & -\sum_{i=1}^{95} w_i\tilde{r}_i^j \leq -U, \quad \text{for all } \tilde{r}^j\in\tilde{\mathcal{R}}. \end{aligned}$$

We have also applied the SAA of $MnCnCV(\beta)$ to solve (36), that is,

$$\begin{aligned}
\min_{\mathbf{w}\in W_{95},\eta\in\mathbb{R}} \quad & -\frac{1}{N}\sum_{j=1}^{N}\sum_{i=1}^{95} w_i\tilde{r}_i^j \\
\text{s.t.} \quad & \eta + \frac{1}{(1-\beta)N}\sum_{j=1}^{N}\left(\left(-\sum_{i=1}^{95} w_i\tilde{r}_i^j\right)-\eta\right)_+ \le -U.
\end{aligned}$$

Solving the problem with different values of U, we obtain optimal solutions to (36) and (5.2.1) with β ranging between 0.97 to 0.99 in Figure EC.3 to Figure EC.5 (the details of these solutions are shown in the electronic companion).

The tests are carried out by using the same built-in solver 'Linprog' and 'interior-point' algorithm as in Section 5.1. For each value of parameter U ranging between 0.97 to 0.99 we can observe that there is some convergence of the optimal solutions and optimal values to $MnCnCV(\beta)$ problems (5.2.1), denoted respectively by $\tilde{\mathbf{w}}^*(\beta,U)$ and $\tilde{f}^*(\beta,U)$ to optimal solution $\tilde{\mathbf{w}}^*(U)$ and optimal value $\tilde{f}^*(U)$ of problem (36).

In Figure 9 we show the average portfolio performance for the risk constrained problem with different values of β and U. We can see the way that the optimal value converges to the $MnCnMx$ value as β approaches 1. This Figure also demonstrates how the more conservative behavior associated with trying to avoid a loss in value gives rise to a lower average portfolio performance. Lower values of β and lower values of U correspond to a higher risk and allow portfolios to be chosen with better average performance.

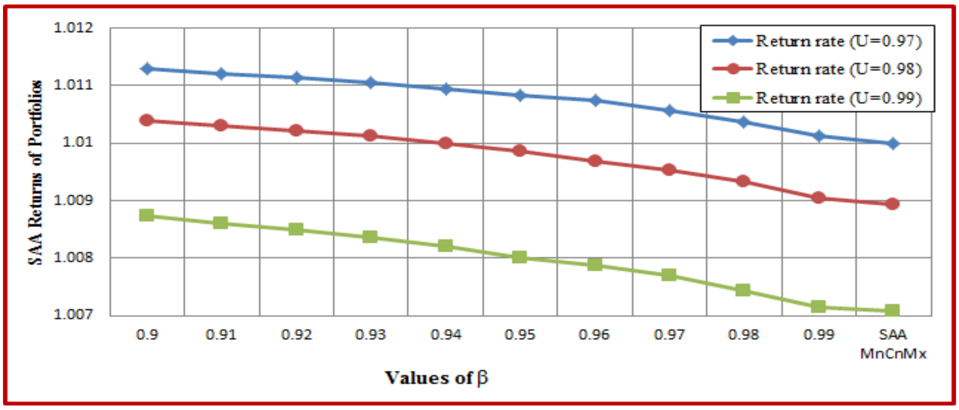


Figure 9 The return rate of $MnCnCV(\beta)$ and $MnCnMx$ portfolios with real data.

5.2.2. The performance of solutions We carry out a set of numerical experiments in this section to test the performances of different optimal portfolios solved from $MnCnCV(\beta_0)$ problems where we denote the optimal portfolio by $\mathbf{w}^*(\beta_0,U_0)$ for different U_0.

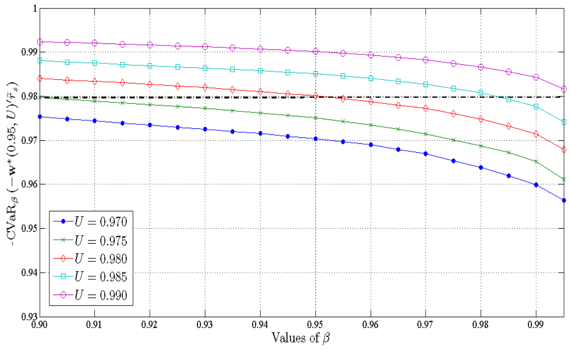
We first consider an $MnCnCV(\beta)$ problem with a parameter pair of U_0 and β_0. By solving the problem, we obtain the optimal portfolio $\mathbf{w}^*(\beta_0, U_0)$ from $MnCnCVSA(\beta_0)$ for U_0, and implement the portfolio into the sample averaged $MnCnCV(\beta)$ problem and also the out of sample $MnCnCV(\beta)$ problem with different values of U and β.

Our aim is to show that if we wish to find a portfolio having good performance while meeting a constraint on return U_0 at a certain β level, β_0, and we need to do this out of sample (i.e. relying on historical data on returns) then we are better off using a portfolio $\mathbf{w}^*(\beta, U)$ with a lower confidence level and a higher level of U (so $\beta < \beta_0$ and $U > U_0$).

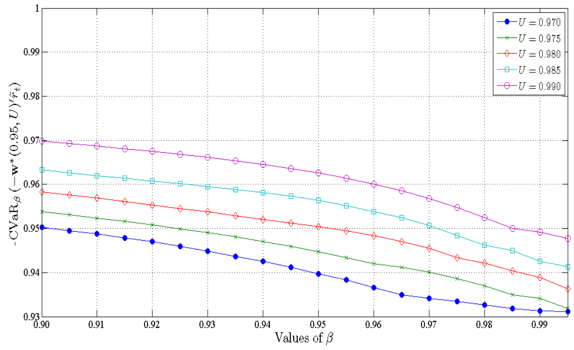
To carry out the tests, we use the historical data from Dec 2008 with sample size being 1000. We divide the data into two sets: a sample set (the first 500 points in the sample) used for solving the optimal portfolio of $MnCnCVSA(\beta)$ and an out-of-sample set (the second part of the sample with size being 500) to test the performances of the optimal portfolios. The two sets of samples are denoted by $\tilde{r}_s \in \mathbb{R}^{s\times 95}$ being a matrix with its (j,i)th element being $\tilde{r}_i^j$ for $j = 1, 2, \cdots, 500$, and $\tilde{r}_t \in \mathbb{R}^{t\times 95}$ being a matrix with its (j,i)th element being $\tilde{r}_i^j$ for $j = 501, \cdots, 1000$. Figure 10 presents a set of test results, where the optimal portfolios $\mathbf{w}^*(\beta_0, U_0)$ for the values of β_0 ranging between 0.95 to 0.99 and the values of U_0 ranging between 0.97 to 0.99 for sample set $\tilde{r}_s$ and out of sample set $\tilde{r}_t$. The figure shows how the value of -CVaR$_\beta(-\mathbf{w}^*(\beta_0, U_0)^\top \tilde{r}_s)$ varies on the sample data (day 1 to day 500) for different values of β (in the left column of Figure 10 (a), (c) and (e)) and how the value of -CVaR$_\beta(-\mathbf{w}^*(\beta_0, U_0)^\top \tilde{r}_t)$ varies on the out of sample data (day 500 to day 1000) for different values of β (in the right column of Figure 10 (b), (d) and (f)) .

Suppose that our aim is to find a portfolio w to maximize average return subject to a risk constraint that $CVaR_{0.97}(w'\tilde{r}_s) > 0.98$. So $U = 0.98$ and $\beta = 0.97$ and the optimal solution is $\mathbf{w}^*(0.97, 0.98)$. If we look at Figure 10 (a) and (c) we can see that portfolios $\mathbf{w}^*(0.95, 0.985)$ and $\mathbf{w}^*(0.99, 0.975)$ are close to achieving the risk constraint required. This makes sense: moving to a lower value of β and a higher value of U gives roughly the same result. Naturally the average return from these portfolios is less since they are not optimal.

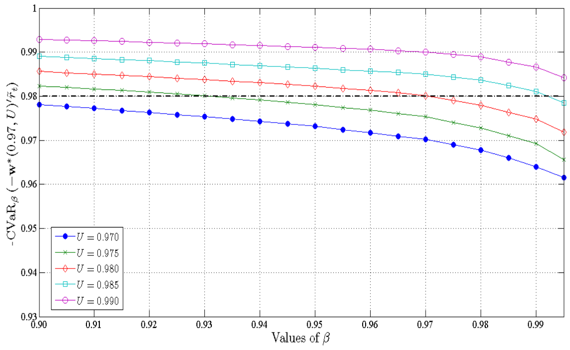
As we can see from Figure 10 (d) in the out-of-sample period the portfolio $\mathbf{w}^*(0.97, 0.98)$ no longer meets the risk constraint at level $U = 0.98$; in fact we have $CVaR_{0.97}(\mathbf{w}^{*\prime}\tilde{r}_s) = 0.951$. Of course this is to be expected; the risk of poor performance is bound to increase out-of-sample. The degradation in performance out-of-sample is also shared by the alternative portfolios $\mathbf{w}^*(0.95, 0.985)$ and $\mathbf{w}^*(0.99, 0.975)$ which achieve $CVaR_{0.97}$ values of 0.953 and 0.946 respectively when applied to out-of-sample data (as can be seen in Figure 10 (b) and (f)). So the question arises: if we wish to satisfy a risk constraint $CVaR_{0.97}(\mathbf{w}^{*\prime}\tilde{r}_s) \geq 0.95$ on out of sample data would we be better off just boosting the U value to 0.98 and optimizing with a β value of 0.97 or would it be better to change both β and U?



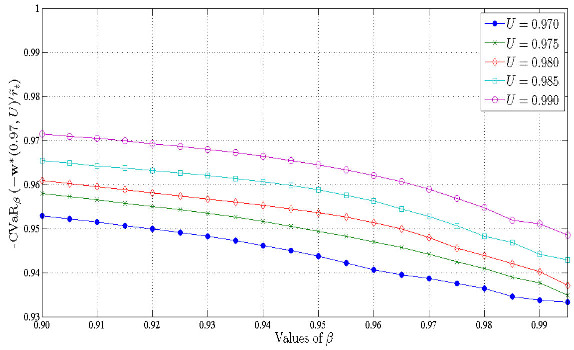
(a) -CVaR$_\beta(-\mathbf{w}^*(0.95, U)'\tilde{r}_s)$



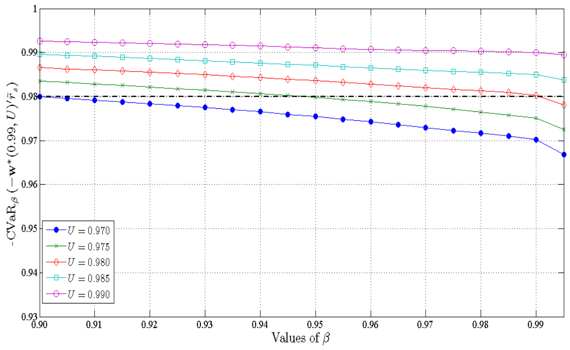
(b) -CVaR$_\beta(-\mathbf{w}^*(0.95, U)'\tilde{r}_t)$



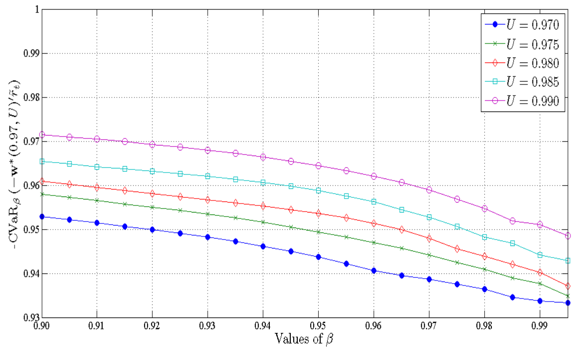
(c) -CVaR$_\beta(-\mathbf{w}^*(0.97, U)'\tilde{r}_s)$



(d) -CVaR$_\beta(-\mathbf{w}^*(0.97, U)'\tilde{r}_t)$



(e) -CVaR$_\beta(-\mathbf{w}^*(0.99, U)'\tilde{r}_s)$



(f) -CVaR$_\beta(-\mathbf{w}^*(0.99, U)'\tilde{r}_t)$

Figure 10 The performance of MnCnCV(β) portfolios.

To investigate this question we carry out an experiment on a rolling basis. For each of 500 sets of data from period t to period $t + 499$ where $t = 1, 2, ..., 500$ we calculate the three portfolios $\mathbf{w}^*(0.95, 0.985)$, $\mathbf{w}^*(0.97, 0.98)$ and $\mathbf{w}^*(0.99, 0.975)$ and then check their performance over the next 50 days. In each case we look at both the average out-of-sample portfolio return and an estimate for the out-of-sample $CVaR_{0.97}$ value obtained by taking a point half way between (a) the average of the lowest two returns (corresponding to an estimate of $CVaR_{0.96}$ given 50 data points) and (b) the lowest return (corresponding to an estimate of $CVaR_{0.98}$).

Figure 11 shows the out-of-sample $CVaR_{0.97}$ estimates over the 500 rolling-window tests. Because we are looking less far ahead, the reduction in CVaR values by moving out of sample is much less than we see in Figure 10. We can see that there is substantial variation over the 500 days, but that the portfolio based on $\beta = 0.95$, $U = 0.985$ has uniformly better CVaR values than that based on $\beta = 0.97$, $U = 0.98$. The mean values for $CVaR_{0.97}$ for $\mathbf{w}^*(0.95, 0.985)$, $\mathbf{w}^*(0.97, 0.98)$ and $\mathbf{w}^*(0.99, 0.975)$ are 0.9858, 0.9836 and 0.9821 respectively. So $\mathbf{w}^*(0.95, 0.985)$ is significantly less risky, but what about its average performance? Figure 12 shows the mean out-of-sample performance over the 500 rolling-window tests. We find that a lower β value gives a slightly better performance, with $\mathbf{w}^*(0.95, 0.985)$ achieving higher average returns after day 610 except for a short period around day 775. The average portfolio returns for $\mathbf{w}^*(0.95, 0.985)$, $\mathbf{w}^*(0.97, 0.98)$ and $\mathbf{w}^*(0.99, 0.975)$ are 1.0114, 1.0112 and 1.0106 respectively. Though it would be wrong to read too much into a single example, this shows how an approach based on deliberately choosing lower values of the confidence level β can lead to significantly better performance.

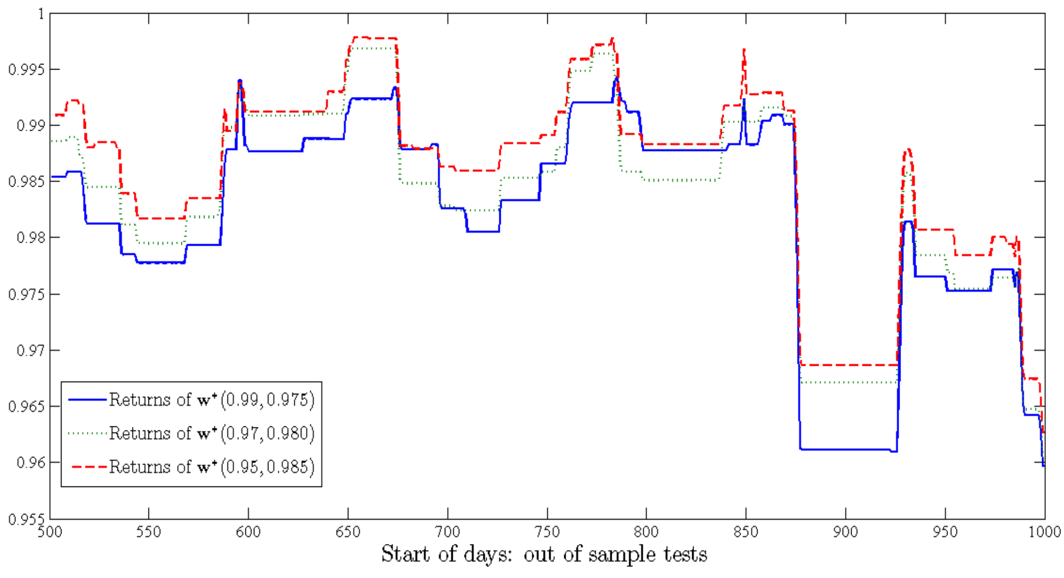


Figure 11 Out-of-sample $CVaR_{0.97}$ estimates over the 500 rolling-window tests.

6. Conclusion

This paper explores the effect of changing the confidence level β when using a CVaR risk measure within an optimization problem. In particular we are interested in the convergence behavior as β

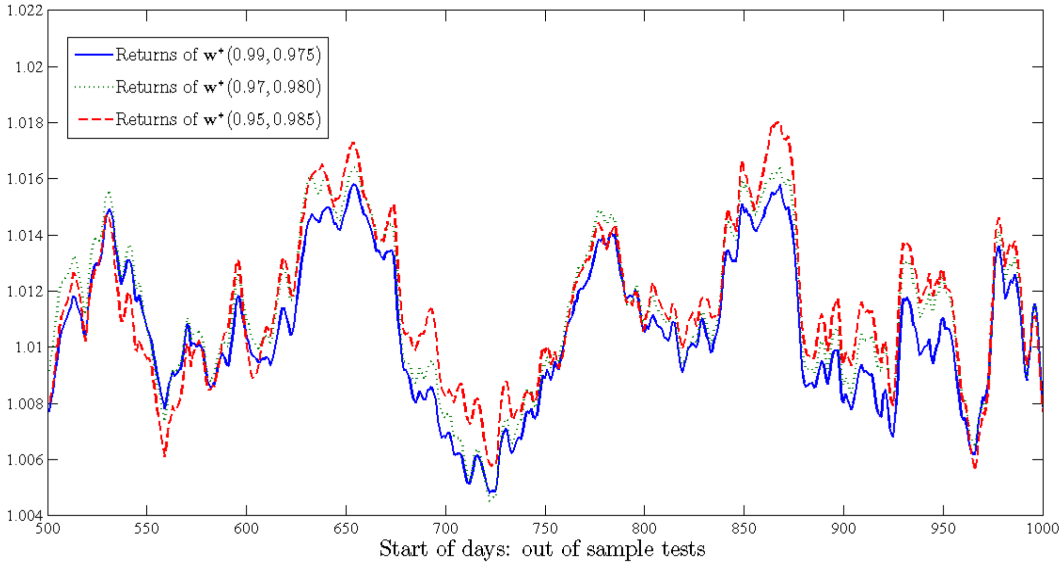


Figure 12 Average out-of-sample performance over the 500 rolling-window tests.

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.

One reason for looking carefully at this question is the way that sample approximations will be better behaved as the confidence level is reduced from 1. This gives a potential advantage to the use of a CVaR-based approximation even when the problem of interest has a minimax form. We explore this possibility through a numerical investigation of certain risk-based portfolio optimization problems. For these examples we show that there can be an advantage in moving to a CVaR approximation when we wish to optimize worst case performance. Moreover when the aim is to optimize average portfolio performance with a risk constraint, then it is best to adjust the risk constraint in a way that makes use of lower β levels.

An underlying theme in this paper relates to the use of CVaR approximations in a context where the original problem is one of robust optimization. We can replace the CVaR approximation with any risk measure which approximates the extreme value, e.g., an entropic risk measure Föllmer and Knispel (2011), for our approximation schemes discussed in this paper. Indeed, the relationship between an uncertain set and a risk measure has been well investigated by Natarajan et al. (2009). One interesting avenue for exploration is the development of similar approximation schemes for distributional robust optimization problems where the dual formulation cannot be obtained explicitly as in Section 4.1. In this context we could consider randomizing over the distributional set, in order to approximate the worst probability distribution with CVaR of the random variable. Consequently the minimax operation in the distributional robust optimization is reduced to minimization of a risk function (e.g., CVaR or entropic risk measure) of the random variable defined over the distributional set.

References

Apkarian P, Tuan HD, (2000) Parameterized LMIs in control theory. *SIAM Journal on Control and Optimization*, 38:1241–1264.

Bank B, Guddat J, Klatte D, Kummer B, Tammer K, (1982) *Nonlinear Parametric Optimization*, Akademine-Verlag, Berlin.

Ben-Tal A, Nemirovski A, (1997) Robust truss topology design via semidefinite programming. *SIAM Journal on Optimization*, 7:991–1016.

Ben-Tal A, Nemirovski A, (1998) Robust convex optimization. *Mathematics of Operations Research*, 23, 769–805.

Ben-Tal A, Nemirovski A, (1999) Robust solutions of uncertain linear programs. *Operations Research Letters*, 25:1–13.

Ben-Tal, A, El Ghaoui L, Nemirovski A (2009) *Robust Optimization*, Princeton University Press, Princeton, NJ, 2009.

Calafiore G (2010), Random convex programs. *SIAM J. Optimization*, 20:3427–3464.

Calafiore G, Campi MC (2005) Uncertain convex programs: randomized solutions and confidence levels. *Mathematical Programming*, 102: 25–46.

Calafiore G, Campi MC (2006) The scenario approach to robust control desig. *IEEE Transactions on Automatic Control*, 51: 742–753.

Campi MC and Garatti S (2008) The exact feasibility of randomized solutions of robust convex programs. *SIAM J. Optimization*, 19:1211–1230.

Campi MC and Garatti S (2010) A sampling-and-discarding approach to chance constrained optimization: feasibility and optimality. *Journal of Optimization Theory and Applications*, 148:257–280.

Föllmer H, Knispel T (2011) Entropic risk measures: coherence vs. convexity, model ambiguity, and robust large deviations. *Stochastics and Dynamics*, 11:333–351.

El Ghaoui L, Lebret H (1998) Robust solutions to uncertain semidefinite programs. *SIAM Journal on Optimization*, 9:33–52.

Klatte D (1987) A note on quantitative stability results in nonlinear optimization. *Seminarbericht Nr. 90*, Sektion Mathematik, Humboldt-Universität zu Berlin, Berlin, 77–86.

Krokhmal P, Palmquist J, Uryasev S (2002) Portfolio optimization with Conditional Value-at-Risk objective and constraints. *Journal of Risk*, 4:43–68.

Liu Y, Xu H (2012) Stability and sensitivity analysis of stochastic programs with second order dominance constraints. *Mathematical Programming Series A*, 142:435-460.

Natarajan K, Pachamanova D, Sim M (2009) Constructing risk measures from uncertainty sets. *Operations Research*, 57:1129–1147.

Pflug G, Römisch W (2007) Modeling, Measuring and Managing Risk, World Scientific, Singapore.

Polak E (1987) On the mathematical foundations of nondifferentiable optimization in engineering design. *SIAM Review*, 29:21–89.

Robinson SM (1975) An application of error bounds for convex programming in a linear space. *SIAM Journal on Control and Optimization*, 13:271–273.

Rockafellar RT and Uryasev S (2000) Optimization of conditional value-at-risk. *The Journal of Risk*, 2:21–41.

Shapiro A (2001) On duality theory of conic linear problems, Miguel A. Goberna and Marco A. López, eds., *Semi-Infinite Programming: Recent Advances, Kluwer Academic Publishers*, 135–165.

Shapiro A (2008) Asymptotics of minimax stochastic programs. *Statistics and Probability Letters*, 78:150–157.

Shapiro A, Xu H (2008) Stochastic mathematical programs with equilibrium constraints, modeling and sample average approximation. *Optimization*, 57:395–418.

Soyster AL (1973) Convex programming with set-inclusive constraints and applications to inexact linear programming. *Operations Research*, 21:1154–1157.

Sun H, Xu H, Wang Y (2012) Asymptotic analysis of sample average approximation for stochastic optimization problems with joint chance constraints via CVaR/DC approximations. *Journal of Optimization Theory and Applications*, in press, 2013.

Tiba D, Zalinescu C (2004) On the necessity of some constraint qualification conditions in convex programming. *Journal of Convex Analysis*, 11:95–110.

Vayanos P, Kuhn D, Rustem B (2012) A constraint sampling approach for multi-stage robust optimization. *Automatica*, 48:459–471.

Xu H (2010) Uniform exponential convergence of sample average random functions under general sampling with applications in stochastic programming. *Journal of Mathematical Analysis and Applications*, 368:692–710.

Xu H, Zhang D (2009) Smooth sample average approximation of stationary points in nonsmooth stochastic optimization and applications. *Mathematical Programming Series A*, 119:371-401.

7. Appendix

Proof of Theorem 1. Part (i) The inequalities follow immediately from the definition of $\mathrm{CVaR}_\beta(f(x,\xi))$ since the values of f in the integral all lie between $\mathrm{VaR}_\beta(f(x,\xi))$ and $\sup_{\xi\in Y} f(x,\xi)$. To establish the statement about limits we suppose first that $\sup_{\xi\in Y} f(x,\xi) = f^*(x) < \infty$. Then it is enough to show that $\lim_{\beta\to 1}\mathrm{VaR}_\beta(f(x,\xi)) = f^*(x)$. Observe that for any $\varepsilon > 0$ we can find an ω with $f(x,y_0) > f^*(x) - \varepsilon$ where $y_0 = \xi(\omega)$. Since f is a continuous function we can deduce that $\rho(y)$ has a positive integral over the open set $f^{-1}(A)$ where $A = (f(x,y_0) - \varepsilon/2, f(x,y_0) + \varepsilon/2)$. Thus

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

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

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

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

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 $\mathrm{VaR}_\beta(f(x,\xi))$ is convex w.r.t. β). By the definition of CVaR

$$\begin{aligned}
\mathrm{CVaR}_\beta(f(x,\xi)) &= \frac{1}{1-\beta}\left(\int_\beta^{(1+\beta)/2} F_x^{-1}(t)dt + \int_{(1+\beta)/2}^1 F_x^{-1}(t)dt\right) \\
&= \frac{1}{1-\beta}\int_0^{(1-\beta)/2}\left[F_x^{-1}\left(\frac{1+\beta}{2} - s\right) + F_x^{-1}\left(\frac{1+\beta}{2} + s\right)\right] ds \\
&\geq \frac{1}{1-\beta}\int_0^{(1-\beta)/2} 2F_x^{-1}\left(\frac{1+\beta}{2}\right) ds = F_x^{-1}\left(\frac{1+\beta}{2}\right).
\end{aligned}$$

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 $\sup_{\xi\in Y} f(x,\xi) = \infty$. So we can assume that f achieves its maximum value, and this is given by $F_x^{-1}(1)$. Now let $t = \beta + (1-\beta)s$ and then, using the convexity of $F_x^{-1}(t)$, we have

$$\begin{aligned}
\frac{1}{1-\beta}\int_\beta^1 F_x^{-1}(t)dt &\leq \int_0^1 \left[sF_x^{-1}(1) + (1-s)F_x^{-1}(\beta)\right] ds \leq \frac{1}{2}\left[F_x^{-1}(1) + F_x^{-1}(\beta)\right] \\
&= \frac{1}{2}\left[\mathrm{VaR}_\beta(f(x,\xi)) + \mathrm{VaR}_1(f(x,\xi))\right].
\end{aligned}$$

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

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

Hence we have

$$f^*(x) - \text{CVaR}_\beta(f(x,\xi) = \frac{1}{1-\beta}\int_\beta^1 (f^*(x) - F_x^{-1}(t))dt \le \frac{1}{1-\beta}\int_\beta^1 \left(\frac{1-t}{K}\right)^{1/\tau} dt$$
$$= \frac{1}{K^{1/\tau}}\frac{1}{1+(1/\tau)}(1-\beta)^{1/\tau}.$$

The proof is complete. □

Proof of Theorem 2. Applying (Liu and Xu 2013, Lemma 3.8) to $MnCV(\beta)$ (treating it as a perturbation of $MnMx$), we know that for any $\epsilon > 0$, there exists a $\delta > 0$ such that $\mathbb{D}(X^*(\beta), X^*) \le \epsilon$ when $\Delta_\beta(x) \le \delta$. This shows (11) because $\lim_{\beta\to 0}\sup_{x\in X}\Delta_\beta(x) = 0$ from Theorem 1. Inequality (13) is straightforward from the second order growth condition (12) again from (Liu and Xu 2013, Lemma 3.8). □

Proof of Theorem 3. Part (i). The conclusion follows from (Xu 2010, Lemma 4.2 (i)). Here we provide a proof for completeness. Let ϵ be a fixed small positive number. Define

$$R(\epsilon) := \inf_{\{x\in X, d(x,\mathcal{G})\ge\epsilon\}} \sup_{y\in Y} f(x,y). \tag{38}$$

Then $R(\epsilon) > 0$ as we take an infimum over x values outside of $\mathcal{G}$. Let $\delta := R(\epsilon)/2$ and $\Phi_\beta(x,\eta)$ be defined as in (6). Under the condition that $f(x,y)$ has consistent tail behavior on X, then Theorem 1 implies that $\min_{\eta\in\mathbb{R}}\Phi_\beta(x,\eta)$ approximates $\sup_{y\in Y} f(x,y)$ uniformly w.r.t. x over X as $\beta \to 1$, i.e. if we choose β sufficiently close to 1 then

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

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

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

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, $\mathbb{D}(\mathcal{G}(\beta),\mathcal{G}) \le \epsilon$. The conclusion follows by noting that $\mathcal{G} \subset \mathcal{G}(\beta)$.

Part (ii). Under the Slater constraint qualification and convexity of $f(\cdot,y)$, it follows by (Tiba and Zalinescu 2004, Proposition 2.8), that (15) satisfies the metric regularity condition, that is, there exists a positive constant C such that

$$d(x,\mathcal{G}) \le C\left(\sup_{y\in Y} f(x,y)\right)_+, \quad \forall x \in X. \tag{39}$$

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

$$d(\hat{x}, \mathcal{G}) \le C \left(\sup_{y \in Y} f(\hat{x}, y) \right)_+ - C \min_{\eta \in \mathbb{R}} \Phi_\beta(\hat{x}, \eta) \le C \sup_{x \in X} \left[\sup_{y \in Y} f(x, y) - \min_{\eta \in \mathbb{R}} \Phi_\beta(x, \eta) \right].$$

The proof is complete. □

Proof of Proposition 1. By Theorem 2 in Rockafellar and Uryasev (2000), $\min_{\eta \in \mathbb{R}} \Phi_\beta(x, \eta)$ is a convex function of x. Under the feasibility condition, it follows from our previous discussion that the system (16) satisfies the Slater constraint qualification. Through Robinson's error bound for a convex system (see Robinson (1975)), we obtain the error bound (17) for some constant C. □

Proof of Theorem 4. Parts (i) and (ii). Write the constraints of *MnCnMx* as $\max_{y \in Y} f(x, y) \le 0$. 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. (Bank et al. 1982, 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 pseudo-Lipschitz continuous in β at $\beta = 1$. By applying Klatte's stability result (Klatte 1987, Theorem 1), we obtain (20) with some positive constant $\hat{C}$. □

Proof of Theorem 5. We can write $MnCnCVSA(\beta)$ in the following equivalent form

$$\begin{aligned} \min_x \ & h(x) \\ \text{s.t.} \ & \min_\eta \Phi_\beta^N(x, \eta) \le 0, \\ & x \in X, \end{aligned} \tag{40}$$

where

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

Thus x^N is an optimal solution of (40). Let $\mathcal{G}(\beta)$ be defined as in Section 3, that is, the set of solutions to inequality (16). Then $\mathcal{G}(\beta)$ is the feasible set of minimization (19). Let $\mathcal{G}_N(\beta)$ denote the feasible set of (40). 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 $f(x,\xi)$ is convex as a function of x, the function $\min_{\eta\in\mathbb{R}} \Phi_\beta(x,\eta)$ is also convex (see Rockafellar and Uryasev (2000)) and

$$\min_{\eta\in\mathbb{R}} \Phi_\beta(x_t,\eta) \le -t\delta. \tag{41}$$

Let ϵ be a positive number and $R(\cdot)$ be defined as in (23). 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

$$L\bar{t}\|x_\beta-\hat{x}\| < R(\epsilon) - L\gamma \tag{42}$$

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 (Shapiro and Xu 2008, Theorem 5.1) to show that there exist positive constants $C(\sigma)$, $\alpha(\sigma)$ and $N(\sigma)$ such that

$$\text{Prob}\left(\sup_{x\in X}\left|\min_\eta \Phi_\beta^N(x,\eta) - \min_\eta \Phi_\beta(x,\eta)\right| \ge \sigma\right) \le C(\sigma)e^{-\alpha(\sigma)N} \tag{43}$$

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

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

$$\begin{aligned}
\text{Prob}\,(x_{\bar{t}}\notin\mathcal{G}_N(\beta)) &= \text{Prob}\left(\min_{\eta\in\mathbb{R}}\Phi_\beta^N(x_{\bar{t}},\eta)>0\right)\\
&= \text{Prob}\left(\min_{\eta\in\mathbb{R}}\Phi_\beta^N(x_{\bar{t}},\eta)-\min_{\eta\in\mathbb{R}}\Phi_\beta(x_{\bar{t}},\eta) > -\min_{\eta\in\mathbb{R}}\Phi_\beta(x_{\bar{t}},\eta)\right)\\
&\le \text{Prob}\left(\min_{\eta\in\mathbb{R}}\Phi_\beta^N(x_{\bar{t}},\eta)-\min_{\eta\in\mathbb{R}}\Phi_\beta(x_{\bar{t}},\eta) > \delta\bar{t}\right) \qquad \text{(by (41))}\\
&< C(\sigma)e^{-\alpha(\sigma)N}
\end{aligned} \tag{44}$$

for $N\ge N(\sigma)$. The last inequality is due to (43) by setting $\sigma<\bar{t}\delta$. This shows that $x_{\bar{t}}\in\mathcal{G}_N(\beta)$with probability $1-C(\sigma)e^{-\alpha(\sigma)N}$.

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

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

Therefore

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

Step 5. We will establish that if $d(x^N, \hat{X}(\beta)) \geq \epsilon + \gamma$ then either $x_{\bar{t}} \notin \mathcal{G}_N(\beta)$; 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 $x_{\bar{t}} \in \mathcal{G}_N(\beta)$ 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

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

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 $x_{\bar{t}} \in \mathcal{G}_N(\beta)$ and $x^N \in \mathcal{G}(\beta)$. But if $d(x^N, \hat{X}(\beta)) \geq \epsilon + \gamma$, then

$$\begin{aligned} h(x^N) - h(x_{\bar{t}}) &= (h(x^N) - h(x_\beta)) - (h(x_{\bar{t}}) - h(x_\beta)) \\ &> R(\epsilon) - (R(\epsilon) - L\gamma) > 0 \end{aligned}$$

using the growth condition on h and inequality (42). However this inequality contradicts the optimality of x^N.

Summarizing the discussion above we can conclude that

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

for $N \geq N(\sigma)$. 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 $\text{Prob}(d(x^N, \hat{X})$. By the properties of $\mathbb{D}$, we have

$$\text{Prob}(d(x^N, \hat{X}) \geq \varepsilon) \leq \text{Prob}(d(x^N, \hat{X}(\beta)) + \mathbb{D}(\hat{X}(\beta), \hat{X})) \geq \varepsilon) \tag{45}$$

Let $\varepsilon = 2\epsilon$. Under the consistent tail condition of f, we can set β sufficiently small by virtue of Theorem 4 such that

$$\mathbb{D}(\hat{X}(\beta), \hat{X})) \leq \epsilon.$$

By (45)

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

and this gives the result of Part (ii). □

E-Companion

EC.1. Figures in Section 5.1.2.

Figure EC.1 $MnCV(\beta)$ and $MnMx$ portfolios with real data.

EC.2. Figures in Section 5.1.3.

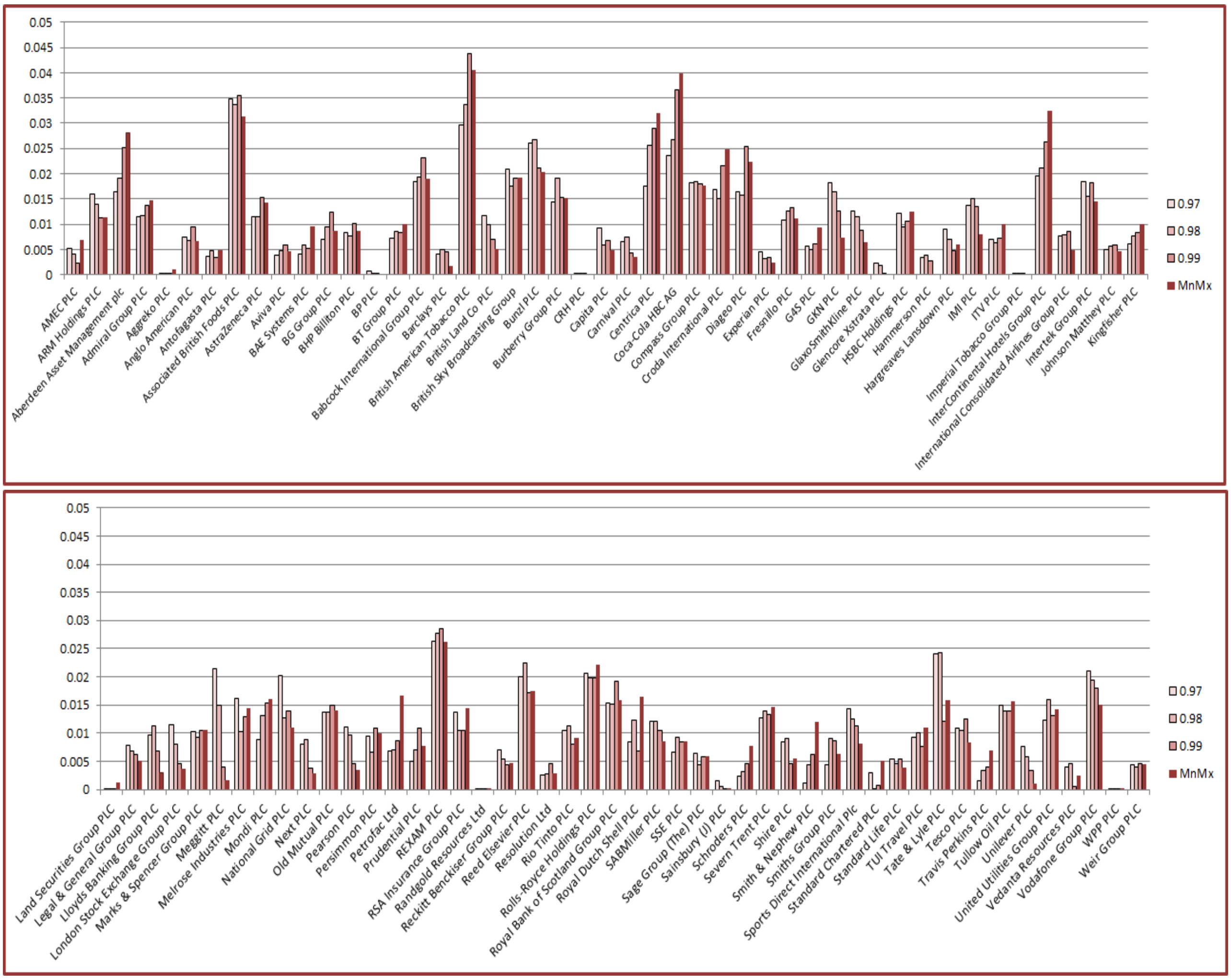


Figure EC.2 Optimal portfolio solved from $MnCV(\beta)$ and $MnMx$ with modified data.

EC.3. Figures in Section 5.2.1.

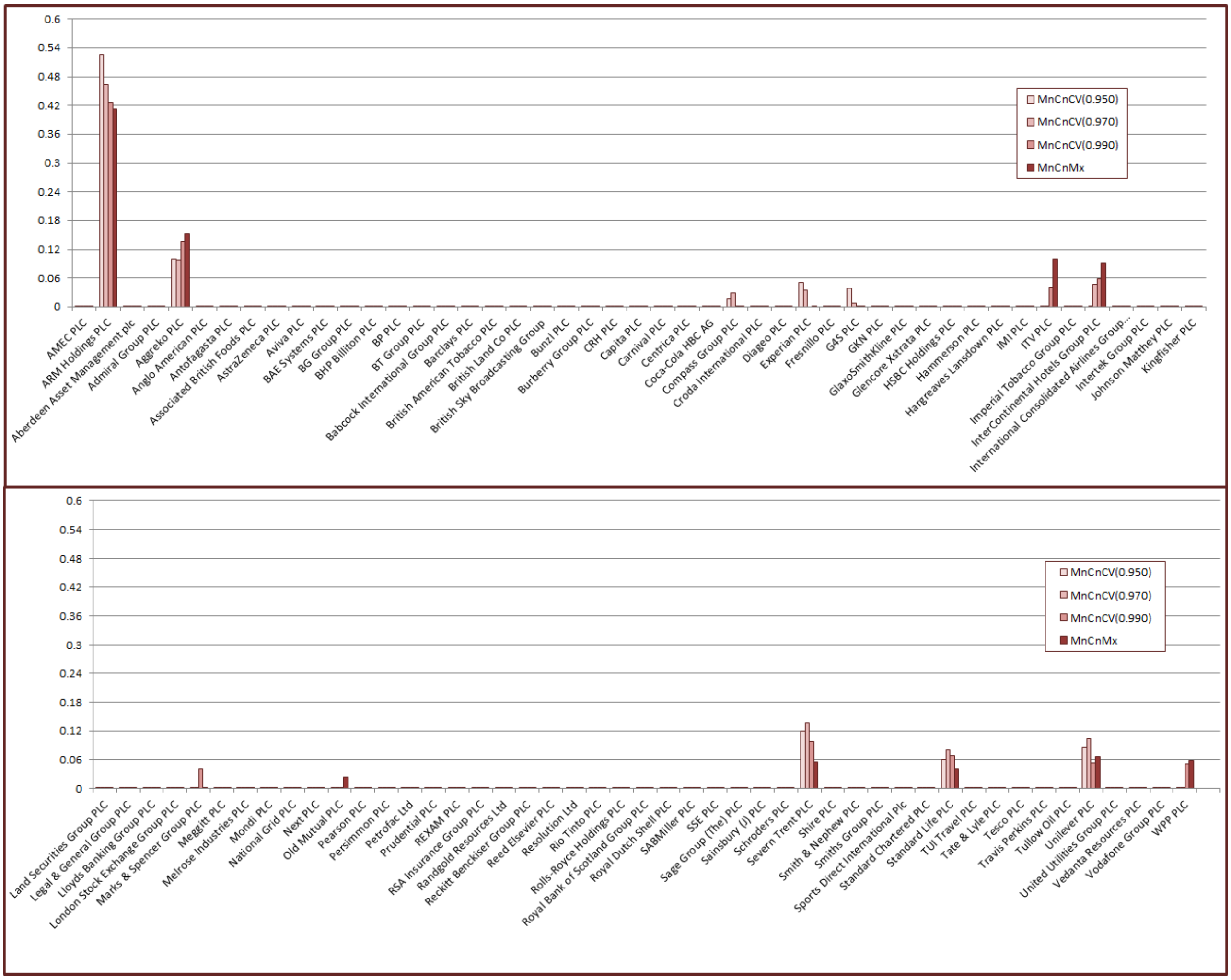


Figure EC.3 $MnCnCV(\beta)$ and $MnCnMx$ portfolios with real data and $U = 0.97$.

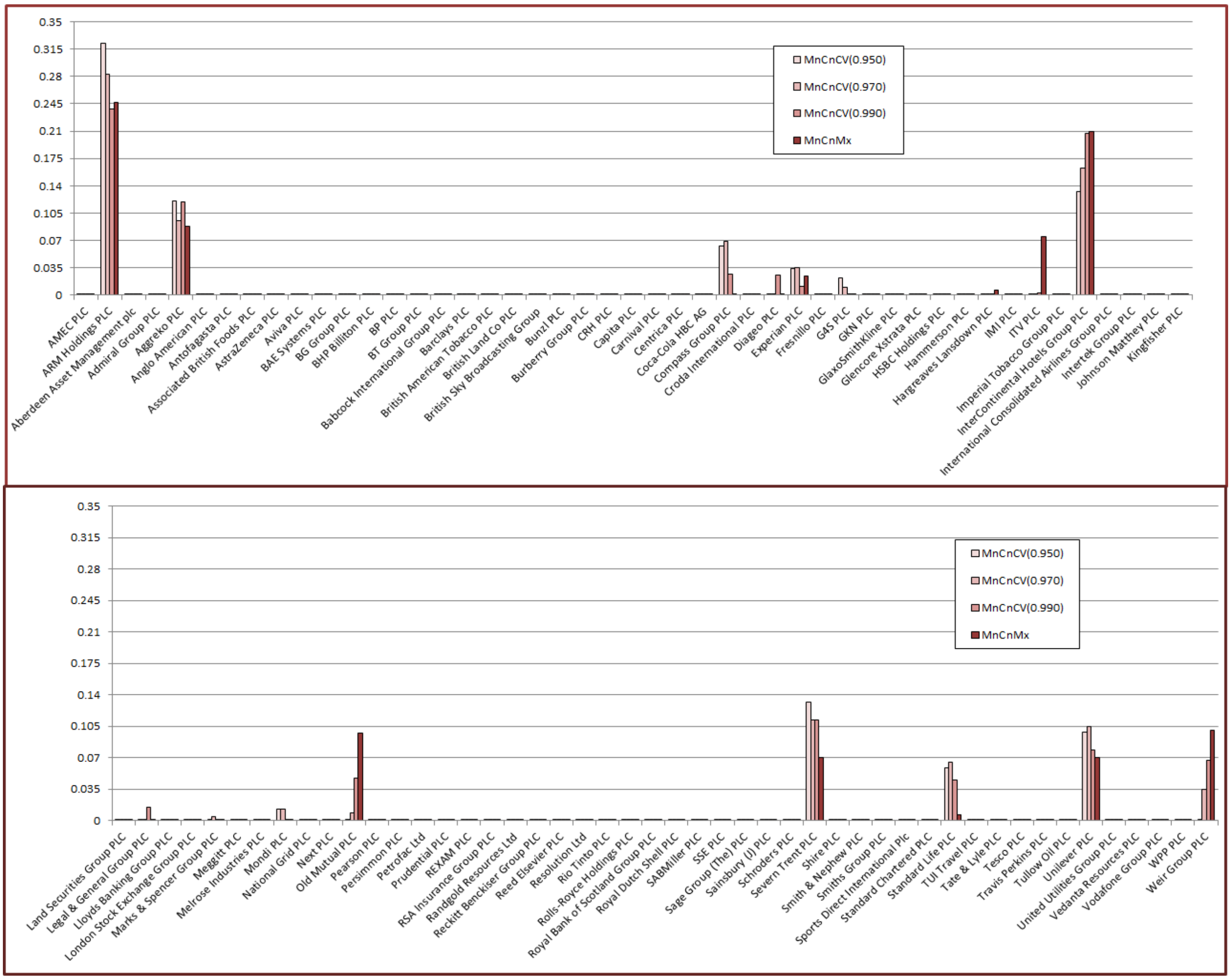


Figure EC.4 $MnCnCV(\beta)$ and $MnCnMx$ portfolios with real data and $U = 0.98$.

Note that when U is increased to a value of 0.99 (corresponding to the portfolio not losing more than 1% of its value over a five day period in the worst case) the number of stocks included in the portfolio increases and the proportion of the investment allocated to any stock is reduced to less than 20%, compared with maximum proportions of more than 53% when $U = 0.97$.

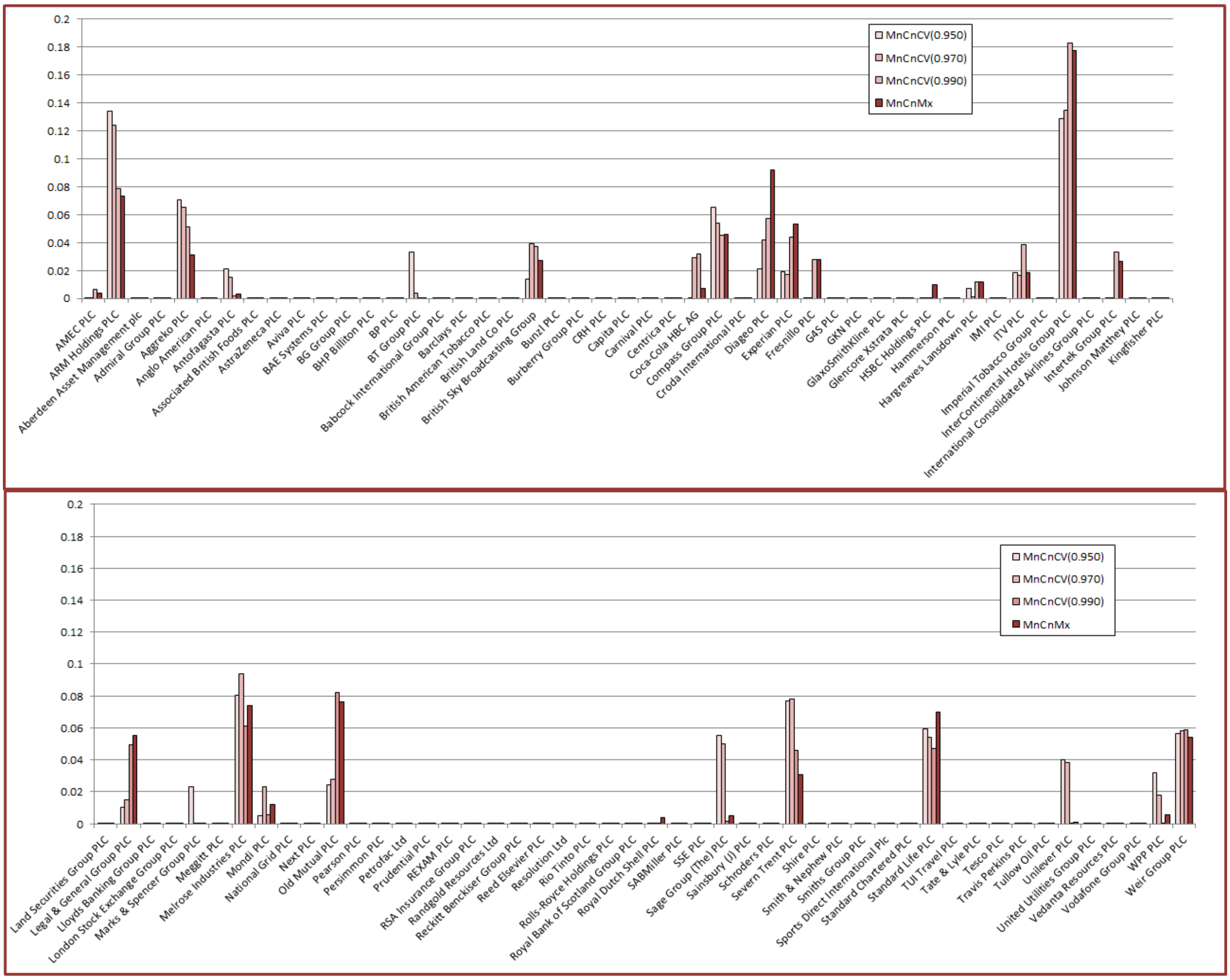


Figure EC.5 $MnCnCV(\beta)$ and $MnCnMx$ portfolios with real data and $U = 0.99$.