Worst-Case Conditional Value-at-Risk with Application to Robust Portfolio Management

Shu-Shang Zhu

Department of Management Science, School of Management, Fudan University, Shanghai 200433, China, sszhu@fudan.edu.cn

Masao Fukushima

Department of Applied Mathematics and Physics, Graduate School of Informatics, Kyoto University, Kyoto 606-8501, Japan, fuku@i.kyoto-u.ac.jp

July 5, 2005

This paper considers the worst-case CVaR in the case where only partial information on the underlying probability distribution is given. The minimization of worst-case CVaR under the mixture distribution uncertainty, componentwise bounded uncertainty and ellipsoidal uncertainty are investigated. The application of worst-case CVaR to robust portfolio optimization is proposed, and the corresponding problems are cast as linear programs and second-order cone programs which can be efficiently solved. Market data simulation and Monte Carlo simulation examples are presented to illustrate the methods. Our approaches can be applied in many situations, including those outside of financial risk management.

Subject Classifications: Finance, portfolio: conditional VaR, portfolio optimization; Programming, linear, nonlinear: robust optimization, second-order cone programming.

1 Introduction

Generally, two types of decision making frameworks are adopted in financial optimization. One is the return-risk trade-off, and the other is utility maximization. The former is widely applied both in practice and in the theoretical study, while the latter is mainly used in the theoretical study.

Markowitz (1952) paved the foundation for modern portfolio theory. His mean-variance analysis is a representative methodology in the framework of return-risk trade-off, where variance is adopted as the measure of risk. Since the middle of 1990s, Value-at-Risk (VaR, see RiskMetrics$^{\text{TM}}$ 1996), a new measure of downside risk, has become popular in financial risk management. It has even been recommended as a standard on banking supervision by Basel Committee. However, VaR is criticized in recent years, especially in three aspects. First, VaR is not sub-additive in the general distribution case, consequently it is not a coherent risk measure in the sense of Artzner et al. (1999). Next, as a function of the portfolio positions, VaR may exhibit multiple local extrema for discrete distributions. Therefore VaR is hard to be optimized in this case. Finally, VaR is just a percentile of a probability distribution, and it does not fully grasp the information of the uncertainty beyond itself. Philippe (1996) presents some details of risk management using VaR. One can find plenty of materials on the theory, modeling, algorithms, and applications related to VaR at http://www.gloriamundi.org which is updated on-line.

Conditional Value-at-Risk (CVaR), defined as the mean of the tail distribution exceeding VaR, has attracted much attention in recent years. As a measure of risk, CVaR exhibits some better properties than VaR. Rockafellar and Uryasev (2000, 2002) showed that minimizing CVaR can be achieved by minimizing a more tractable auxiliary function without predetermining the corresponding VaR first, and at the same time, VaR can be calculated as a by-product. The CVaR minimization formulation given by Rockafellar and Uryasev (2000, 2002) usually results in convex programs, and even linear programs. Thus, their work opened the door to applying CVaR to financial optimization and risk management in practice. Pflug (2000) and Acerbi and Tasche (2002) proved that CVaR is a coherent risk measure. Rockafellar and Uryasev (2002) further explained the coherence of CVaR, and showed that CVaR is stable in the sense of continuity with respect to the confidence level β. Pflug (2000) and Ogryczak and Ruszczyński (2002) showed that CVaR is in harmony with the stochastic dominance principles which are closely related to the utility theory. Konno, Waki and Yuuki (2002) illustrated the significance of using CVaR in reducing downside risk in portfolio optimization. All these stimulate the application of CVaR in practice. It is evidenced that CVaR is becoming more and more popular in financial management (Andersson et al. 2001, Bogentoft, Romeijn and Uryasev 2001, Topaloglou, Vladimirou and Zenios 2002).

As pointed out by Black and Litterman (1992), for the classical mean-variance model, the portfolio decision is very sensitive to the mean and the covariance matrix, especially to the mean. They showed that a small change in the mean can produce a large change in the optimal portfolio position. Thus the modeling risk arises due to the uncertainty of the underlying probability distribution. The uncertainty of the distribution can readily be observed in the case where enough data samples are not available, or the data samples are unstable. Moreover, it occurs in many other situations, such as portfolio selection with uncertain time of exit (Martellini and Uroševió 2001). Another typical example is the decentralized investment management system (Mulvey and Erkan 2003), where each decision maker might have personal views on the future markets and cannot agree with each other.

Recently, a few researchers have paid more attention to the issue of lack of robustness. Ben-Tal, Margalit and Nemirovski (1999) formulated a robust multi-stage portfolio problem using a robust linear programming approach. Lobo and Boyd (2000), Costa and Paiva (2002), Goldfarb and Iyengar (2003) studied the robust portfolio in the mean-variance framework. Instead of the precise information on the mean and the covariance matrix of asset returns, they introduced some types of uncertainties, such as polytopic uncertainty, box uncertainty and ellipsoidal uncertainty, in the parameters involved in the mean and the covariance matrix, and then translated the problem into semidefinite programs or second-order cone programs, which can efficiently be solved by interior-point algorithms developed in recent years. Halldórsson and Tütüncü (2003) applied their interior-point method for saddle-point problems to the robust mean-variance portfolio selection under the box uncertainty in the elements of the mean vector and the covariance matrix. El Ghaoui, Oks and Oustry (2003) investigated the robust portfolio optimization using worst-case VaR, where only partial information on the distribution is known. Several formulations corresponding to various partial information structures have extensively been exploited to formulate the problems as semidefinite programs. Goldfarb and Iyengar (2003) also considered the robust VaR portfolio selection problem by assuming a normal distribution.

Robust optimization is not a new field in operations research. However, it is the breakthrough of the research in conic programming that has greatly stimulated the state-of-the-art of the robust optimization. The reader interested in robust optimization is referred to Ben-Tal and Nemirovski (2002) and the references therein.

In this paper, we consider the the worst-case CVaR in the situation where the information on the underlying probability distribution is not exactly known. The paper is outlined as follows: In the next section, we introduce the concept of worst-case CVaR and discuss the minimization of worst-case CVaR in detail. Three types of uncertainties in the distributions, mixture distribution uncertainty, box uncertainty and ellipsoidal uncertainty are investigated. In Section 3, we present the application of worst-case CVaR to robust portfolio optimization, together with some illustrative numerical examples. Finally, we give some concluding remarks and discuss some future directions in this topic in Section 4.

2 Minimization of Worst-Case CVaR

Let $f(\boldsymbol{x}, \boldsymbol{y})$ denote the loss associated with decision vector $\boldsymbol{x} \in \mathcal{X} \subseteq \mathcal{R}^n$ and random vector $\boldsymbol{y} \in \mathcal{R}^m$ (We use boldface letters to denote vectors and capital letters to denote matrices). For the sake of simple formulation and clear understanding, we assume that $\boldsymbol{y}$ follows a continuous distribution in the first part of this section, and denote its density function as $p(\cdot)$. However all the results remain true for general distributions (see Remark 1). We also assume $E(|f(\boldsymbol{x}, \boldsymbol{y})|) < +\infty$ for each $\boldsymbol{x} \in \mathcal{X}$, so that CVaR and worst-case CVaR will be properly defined.

Given a decision $\boldsymbol{x} \in \mathcal{X}$, the probability of $f(\boldsymbol{x}, \boldsymbol{y})$ not exceeding a threshold α is represented as

$$\Psi(\boldsymbol{x}, \alpha) \triangleq \int_{f(\boldsymbol{x}, \boldsymbol{y}) \leq \alpha} p(\boldsymbol{y}) d\boldsymbol{y}.$$

Given a confidence level β (usually greater than 0.9) and a fixed $\boldsymbol{x} \in \mathcal{X}$, the value-at-risk is defined as

$$\mathrm{VaR}_\beta(\boldsymbol{x}) \triangleq \min\{\alpha \in \mathcal{R} : \Psi(\boldsymbol{x}, \alpha) \geq \beta\}.$$

The corresponding conditional value-at-risk, denoted by $\mathrm{CVaR}_\beta(\boldsymbol{x})$, is defined as the expected value of loss that exceeds $\mathrm{VaR}_\beta(\boldsymbol{x})$, that is,

$$\mathrm{CVaR}_\beta(\boldsymbol{x}) \triangleq \frac{1}{1-\beta} \int_{f(\boldsymbol{x},\boldsymbol{y}) \geq \mathrm{VaR}_\beta(\boldsymbol{x})} f(\boldsymbol{x}, \boldsymbol{y}) p(\boldsymbol{y}) d\boldsymbol{y}.$$

Rockafellar and Uryasev (2000, 2002) demonstrate that the calculation of CVaR can be achieved by minimizing the following auxiliary function with respect to variable $\alpha \in \mathcal{R}$:

$$F_\beta(\boldsymbol{x}, \alpha) \triangleq \alpha + \frac{1}{1-\beta} \int_{\boldsymbol{y} \in \mathcal{R}^m} [f(\boldsymbol{x}, \boldsymbol{y}) - \alpha]^+ p(\boldsymbol{y}) d\boldsymbol{y}, \tag{1}$$

where $[t]^+ = \max\{t, 0\}$. Thus we have the formula

$$\mathrm{CVaR}_\beta(\boldsymbol{x}) = \min_{\alpha \in \mathcal{R}} F_\beta(\boldsymbol{x}, \alpha). \tag{2}$$

Instead of assuming the precise knowledge of the distribution of random vector $\boldsymbol{y}$, we assume in this paper that the density function is only known to belong to a certain set $\mathscr{P}$ of distributions, i.e.,

$$p(\cdot) \in \mathscr{P}.$$

This is the case widely faced in practice.

Definition 1 *The worst-case* CVaR (WCVaR) *for fixed* $\boldsymbol{x} \in \mathcal{X}$ *with respect to* $\mathscr{P}$ *is defined as*

$$\mathrm{WCVaR}_\beta(\boldsymbol{x}) \triangleq \sup_{p(\cdot) \in \mathscr{P}} \mathrm{CVaR}_\beta(\boldsymbol{x}). \tag{3}$$

In the sequel, we will make further investigations on some special cases of $\mathscr{P}$ that meet practical requirements and, at the same time, can be efficiently solved. First we need to quote the following lemma which will serve as a key to transform the problem to a tractable one.

Lemma 1 *Suppose* $\mathcal{X}$ *and* $\mathcal{Y}$ *are nonempty compact convex sets in* $\mathcal{R}^n$ *and* $\mathcal{R}^m$*, respectively, and the function* $\phi(\boldsymbol{x}, \boldsymbol{y})$ *is convex in* $\boldsymbol{x}$ *for any given* $\boldsymbol{y}$*, and concave in* $\boldsymbol{y}$ *for any given* $\boldsymbol{x}$*. Then we have*

$$\min_{\boldsymbol{x} \in \mathcal{X}} \max_{\boldsymbol{y} \in \mathcal{Y}} \phi(\boldsymbol{x}, \boldsymbol{y}) = \max_{\boldsymbol{y} \in \mathcal{Y}} \min_{\boldsymbol{x} \in \mathcal{X}} \phi(\boldsymbol{x}, \boldsymbol{y}).$$

One can find the detail of this lemma in Fan (1953) and Bazaraa, Sherali and Shetty (1993, Chapter 6).

2.1 Mixture Distribution

We assume in this subsection that the density function of $\boldsymbol{y}$ is only known to belong to a set of distributions which consists of all the mixture distributions of some possible distribution scenarios, i.e.,

$$p(\cdot) \in \mathscr{P}_M \triangleq \left\{ \sum_{i=1}^{l} \lambda_i p^i(\cdot) : \sum_{i=1}^{l} \lambda_i = 1, \ \lambda_i \geq 0, \ i = 1, \cdots, l \right\}, \tag{4}$$

where $p^i(\cdot)$ denotes the i-th distribution scenario, and l denotes the number of possible scenarios. Mixture distribution has already been studied in robust statistics and used in modeling the distribution of financial data (Hall, Brorsen and Irwin 1989, Peel and McLachlan 2000). Denote

$$\Lambda \triangleq \left\{ \boldsymbol{\lambda} = (\lambda_1, \cdots, \lambda_l) : \sum_{i=1}^{l} \lambda_i = 1, \lambda_i \geq 0, i = 1, \cdots, l \right\}. \tag{5}$$

Define

$$F_\beta^i(\boldsymbol{x}, \alpha) \triangleq \alpha + \frac{1}{1-\beta} \int_{\boldsymbol{y} \in \mathcal{R}^m} [f(\boldsymbol{x}, \boldsymbol{y}) - \alpha]^+ p^i(\boldsymbol{y}) d\boldsymbol{y}, \quad i = 1, \cdots, l.$$

Theorem 1 *For each $\boldsymbol{x}$,* $\mathrm{WCVaR}_\beta(\boldsymbol{x})$ *with respect to $\mathscr{P}_M$ is given by*

$$\mathrm{WCVaR}_\beta(\boldsymbol{x}) = \min_{\alpha \in \mathcal{R}} \max_{i \in \mathcal{L}} F_\beta^i(\boldsymbol{x}, \alpha), \tag{6}$$

where $\mathcal{L} \triangleq \{1, 2, \cdots, l\}$.

Proof. For given $\boldsymbol{x} \in \mathcal{X}$, define

$$\begin{aligned} H_\beta(\boldsymbol{x}, \alpha, \boldsymbol{\lambda}) &\triangleq \alpha + \frac{1}{1-\beta} \int_{\boldsymbol{y} \in \mathcal{R}^m} [f(\boldsymbol{x}, \boldsymbol{y}) - \alpha]^+ \left[\sum_{i=1}^{l} \lambda_i p^i(\boldsymbol{y}) \right] d\boldsymbol{y} \\ &= \sum_{i=1}^{l} \lambda_i F_\beta^i(\boldsymbol{x}, \alpha), \end{aligned} \tag{7}$$

where $\boldsymbol{\lambda} \in \Lambda$. $H_\beta(\boldsymbol{x}, \alpha, \boldsymbol{\lambda})$ is convex in α (see Rockafellar and Uryasev 2000, 2002) and affine (concave) in $\boldsymbol{\lambda}$. It is easy to see that $\min_{\alpha \in \mathcal{R}} H_\beta(\boldsymbol{x}, \alpha, \boldsymbol{\lambda})$ is a continuous function with respect to $\boldsymbol{\lambda}$. By (2), (3), (4) and the fact that Λ is compact, we can write

$$\mathrm{WCVaR}_\beta(\boldsymbol{x}) = \max_{\boldsymbol{\lambda} \in \Lambda} \min_{\alpha \in \mathcal{R}} H_\beta(\boldsymbol{x}, \alpha, \boldsymbol{\lambda}) = \max_{\boldsymbol{\lambda} \in \Lambda} \min_{\alpha \in \mathcal{R}} \sum_{i=1}^{l} \lambda_i F_\beta^i(\boldsymbol{x}, \alpha). \tag{8}$$

For each i and fixed $\boldsymbol{x}$, the optimal solution set of $\min_{\alpha \in \mathcal{R}} F_\beta^i(\boldsymbol{x}, \alpha)$ is a nonempty, closed and bounded interval (see Rockafellar and Uryasev 2000, 2002). Thus we can denote

$$[\underline{\alpha}_i^*, \overline{\alpha}_i^*] \triangleq \operatorname*{argmin}_{\alpha \in \mathcal{R}} F_\beta^i(\boldsymbol{x}, \alpha), \quad i = 1, \cdots, l.$$

Suppose $g_1(t)$ and $g_2(t)$ are two convex functions defined on $\mathcal{R}$, and the nonempty, closed and bounded intervals $\left[\underline{t}_1^*, \overline{t}_1^*\right]$, $\left[\underline{t}_2^*, \overline{t}_2^*\right]$ are the sets of minima of these two functions, respectively. It can be easily verified that for any $\beta_1 \geq 0$ and $\beta_2 \geq 0$ such that $\beta_1+\beta_2 > 0$, $\beta_1 g_1(t)+\beta_2 g_2(t)$ is convex too, and the set of minima of $\beta_1 g_1(t) + \beta_2 g_2(t)$ must lie in the nonempty, closed and bounded interval $\left[\min\{\underline{t}_1^*, \underline{t}_2^*\}, \max\{\overline{t}_1^*, \overline{t}_2^*\}\right]$. From this fact and (7), we get

$$\operatorname*{argmin}_{\alpha\in\mathcal{R}} H_\beta(\boldsymbol{x}, \alpha, \boldsymbol{\lambda}) \subseteq \mathcal{A}, \quad \forall \boldsymbol{\lambda} \in \Lambda,$$

where $\mathcal{A}$ is the nonempty, closed and bounded interval given by

$$\mathcal{A} \triangleq \left[\min_{i\in\mathcal{L}} \underline{\alpha}_i^*, \max_{i\in\mathcal{L}} \overline{\alpha}_i^*\right].$$

This implies

$$\min_{\alpha\in\mathcal{R}} H_\beta(\boldsymbol{x}, \alpha, \boldsymbol{\lambda}) = \min_{\alpha\in\mathcal{A}} H_\beta(\boldsymbol{x}, \alpha, \boldsymbol{\lambda}).$$

Therefore, by Lemma 1, we have

$$\max_{\boldsymbol{\lambda}\in\Lambda} \min_{\alpha\in\mathcal{R}} H_\beta(\boldsymbol{x}, \alpha, \boldsymbol{\lambda}) = \max_{\boldsymbol{\lambda}\in\Lambda} \min_{\alpha\in\mathcal{A}} H_\beta(\boldsymbol{x}, \alpha, \boldsymbol{\lambda}) = \min_{\alpha\in\mathcal{A}} \max_{\boldsymbol{\lambda}\in\Lambda} H_\beta(\boldsymbol{x}, \alpha, \boldsymbol{\lambda}). \tag{9}$$

It is obvious that

$$\min_{\alpha\in\mathcal{A}} \max_{\boldsymbol{\lambda}\in\Lambda} H_\beta(\boldsymbol{x}, \alpha, \boldsymbol{\lambda}) \geq \inf_{\alpha\in\mathcal{R}} \max_{\boldsymbol{\lambda}\in\Lambda} H_\beta(\boldsymbol{x}, \alpha, \boldsymbol{\lambda}). \tag{10}$$

By (9), (10) and the well known result on the min-max inequality

$$\inf_{\alpha\in\mathcal{R}} \max_{\boldsymbol{\lambda}\in\Lambda} H_\beta(\boldsymbol{x}, \alpha, \boldsymbol{\lambda}) \geq \max_{\boldsymbol{\lambda}\in\Lambda} \min_{\alpha\in\mathcal{R}} H_\beta(\boldsymbol{x}, \alpha, \boldsymbol{\lambda}),$$

we immediately get

$$\max_{\boldsymbol{\lambda}\in\Lambda} \min_{\alpha\in\mathcal{R}} H_\beta(\boldsymbol{x}, \alpha, \boldsymbol{\lambda}) = \min_{\alpha\in\mathcal{R}} \max_{\boldsymbol{\lambda}\in\Lambda} H_\beta(\boldsymbol{x}, \alpha, \boldsymbol{\lambda}).$$

It then follows from (8), along with (7) that

$$\mathrm{WCVaR}_\beta(\boldsymbol{x}) = \min_{\alpha\in\mathcal{R}} \max_{\boldsymbol{\lambda}\in\Lambda} H_\beta(\boldsymbol{x}, \alpha, \boldsymbol{\lambda}) = \min_{\alpha\in\mathcal{R}} \max_{\boldsymbol{\lambda}\in\Lambda} \sum_{i=1}^{l} \lambda_i F_\beta^i(\boldsymbol{x}, \alpha). \tag{11}$$

Now we only need to verify the equivalence of the right-hand sides of (6) and (11). As an optimization problem, the right-hand side of (11) is equivalent to

$$\begin{aligned} &\min_{(\alpha,\theta)\in\mathcal{R}\times\mathcal{R}} \theta \\ &\text{s.t.} \quad \sum_{i=1}^{l} \lambda_i F_\beta^i(\boldsymbol{x}, \alpha) \leq \theta, \quad \forall \boldsymbol{\lambda} \in \Lambda. \end{aligned} \tag{12}$$

From (5), it is clear that any feasible solution of (12) satisfies

$$F_\beta^i(\boldsymbol{x}, \alpha) \leq \theta, \quad i = 1, \cdots, l. \tag{13}$$

On the other hand, if (13) holds, then for any $\boldsymbol{\lambda} \in \Lambda$, we have

$$\sum_{i=1}^{l} \lambda_i F_\beta^i(\boldsymbol{x}, \alpha) \leq \sum_{i=1}^{l} \lambda_i \theta = \theta.$$

Thus problem (12) is equivalent to

$$\begin{aligned} &\min_{(\alpha,\theta)\in\mathcal{R}\times\mathcal{R}} \theta \\ &\text{s.t.} \quad F_\beta^i(\boldsymbol{x}, \alpha) \leq \theta, \quad i = 1, \cdots, l, \end{aligned}$$

which is nothing but the right-hand side of (6). This completes the proof. □

Denote

$$F_\beta^{\mathcal{L}}(\boldsymbol{x}, \alpha) \triangleq \max_{i\in\mathcal{L}} F_\beta^i(\boldsymbol{x}, \alpha).$$

By Theorem 1, we get the following corollary immediately.

Corollary 1 *Minimizing* $\mathrm{WCVaR}_\beta(\boldsymbol{x})$ *over* $\mathcal{X}$ *can be achieved by minimizing* $F_\beta^{\mathcal{L}}(\boldsymbol{x}, \alpha)$ *over* $\mathcal{X} \times \mathcal{R}$*, i.e.,*

$$\min_{\boldsymbol{x}\in\mathcal{X}} \mathrm{WCVaR}_\beta(\boldsymbol{x}) = \min_{(\boldsymbol{x},\alpha)\in\mathcal{X}\times\mathcal{R}} F_\beta^{\mathcal{L}}(\boldsymbol{x}, \alpha).$$

More specifically, if $(\boldsymbol{x}^*, \alpha^*)$ *attains the right-hand side minimum, then* $\boldsymbol{x}^*$ *attains the left-hand side minimum and* α^* *attains the minimum of* $F_\beta^{\mathcal{L}}(\boldsymbol{x}^*, \alpha)$*, and vice versa.*

It is known that $F_\beta(\boldsymbol{x}, \alpha)$ defined by (1) is convex in $(\boldsymbol{x}, \alpha)$ if $f(\boldsymbol{x}, \boldsymbol{y})$ is convex in $\boldsymbol{x}$ (see Rockafellar and Uryasev 2000, 2002). By the fact that the function $g(\boldsymbol{t}) = \max\{g_1(\boldsymbol{t}), g_2(\boldsymbol{t})\}$ is convex if both $g_1(\boldsymbol{t})$ and $g_2(\boldsymbol{t})$ are convex, we get that if $f(\boldsymbol{x}, \boldsymbol{y})$ is convex in $\boldsymbol{x}$, then $F_\beta^{\mathcal{L}}(\boldsymbol{x}, \alpha)$ is convex in $(\boldsymbol{x}, \alpha)$. So, if $\mathcal{X}$ is a convex set and $f(\boldsymbol{x}, \boldsymbol{y})$ is a convex function of $\boldsymbol{x}$, then the WCVaR minimization problem is a convex program.

From now on, we discuss the computational aspect of minimization of WCVaR. Theorem 1 and Corollary 1 help us to translate the original problem to a more tractable one. It can be seen that the WCVaR minimization is equivalent to

$$\begin{aligned} &\min_{(\boldsymbol{x},\alpha,\theta)\in\mathcal{X}\times\mathcal{R}\times\mathcal{R}} \theta \\ &\text{s.t.} \quad \alpha + \frac{1}{1-\beta}\int_{\boldsymbol{y}\in\mathcal{R}^m} [f(\boldsymbol{x}, \boldsymbol{y}) - \alpha]^+ p^i(\boldsymbol{y}) d\boldsymbol{y} \leq \theta, \quad i = 1, \cdots, l. \end{aligned} \tag{14}$$

The most difficult part in the computation of (14) is the calculation of the integral of a multivariate and non-smooth function. However, an approximation method can be used to deal with this difficulty. Monte Carlo simulation is one of the most efficient methods for high dimensional integral computation. Rockafellar and Uryasev (2000) use this method to approximate $F_\beta(\boldsymbol{x}, \alpha)$ as

$$\tilde{F}_\beta(\boldsymbol{x}, \alpha) = \alpha + \frac{1}{S(1-\beta)} \sum_{k=1}^{S} [f(\boldsymbol{x}, \boldsymbol{y}_{[k]}) - \alpha]^+, \tag{15}$$

where $\boldsymbol{y}_{[k]}$ denotes the k-th sample (We use the subscript $[k]$ to distinguish a vector from a scalar) generated by simple random sampling with respect to $\boldsymbol{y}$ according to its density function $p(\cdot)$, and S denotes the number of samples. The Law of Large Numbers in probability theory guarantees the approximation accuracy (or convergence) when the number of samples becomes large enough. If $f(\boldsymbol{x}, \boldsymbol{y})$ is linear with respect to $\boldsymbol{x}$ and $\mathcal{X}$ is a convex polyhedron, then the Monte Carlo method produces linear programs, which can be efficiently solved.

Replacing the integral in (14) with (15) yields

$$\begin{aligned} &\min_{(\boldsymbol{x},\alpha,\theta)\in\mathcal{X}\times\mathcal{R}\times\mathcal{R}} \quad \theta \\ &\text{s.t.} \quad \alpha + \frac{1}{S^i(1-\beta)} \sum_{k=1}^{S^i} [f(\boldsymbol{x}, \boldsymbol{y}^i_{[k]}) - \alpha]^+ \le \theta, \quad i = 1, \cdots, l, \end{aligned} \tag{16}$$

where $\boldsymbol{y}^i_{[k]}$ is the k-th sample with respect to the i-th distribution scenario $p^i(\cdot)$, and S^i denotes the number of the corresponding samples. Instead of the simple random sampling method, some improved sampling approaches can be used to approximate the integral. Generally, the approximation of problem (14) can be formulated as

$$\begin{aligned} &\min_{(\boldsymbol{x},\alpha,\theta)\in\mathcal{X}\times\mathcal{R}\times\mathcal{R}} \quad \theta \\ &\text{s.t.} \quad \alpha + \frac{1}{1-\beta} \sum_{k=1}^{S^i} \pi^i_k [f(\boldsymbol{x}, \boldsymbol{y}^i_{[k]}) - \alpha]^+ \le \theta, \quad i = 1, \cdots, l, \end{aligned} \tag{17}$$

where π^i_k denotes the probability according to the k-th sample with respect to the i-th distribution scenario $p^i(\cdot)$. If π^i_k is equal to $\frac{1}{S^i}$ for all k, then (17) reduces to (16). In the following, we denote $\boldsymbol{\pi}^i = \left(\pi^i_1, \cdots, \pi^i_{S^i}\right)^T$.

Proposition 1 *Let $m = \sum_{i=1}^{l} S^i$. Then, by introducing an auxiliary vector $\boldsymbol{u} = (\boldsymbol{u}^1; \boldsymbol{u}^2; \cdots; \boldsymbol{u}^l) \in \mathcal{R}^m$, the optimization problem (17) can be rewritten as the following minimization problem with variables $(\boldsymbol{x}, \boldsymbol{u}, \alpha, \theta) \in \mathcal{R}^n \times \mathcal{R}^m \times \mathcal{R} \times \mathcal{R}$:*

$$\begin{aligned} &\min \quad \theta \\ &\text{s.t.} \quad \boldsymbol{x} \in \mathcal{X}, \\ &\qquad \alpha + \frac{1}{1-\beta} \left(\boldsymbol{\pi}^i\right)^T \boldsymbol{u}^i \le \theta, \\ &\qquad u^i_k \ge f(\boldsymbol{x}, \boldsymbol{y}^i_{[k]}) - \alpha, \\ &\qquad u^i_k \ge 0, \qquad k = 1, \cdots, S^i, \quad i = 1, \cdots, l. \end{aligned} \tag{18}$$

More specifically, if $(\boldsymbol{x}^*, \boldsymbol{u}^*, \alpha^*, \theta^*)$ solves (18), then $(\boldsymbol{x}^*, \alpha^*, \theta^*)$ solves (17), Conversely, if $(\boldsymbol{x}^*, \alpha^*, \theta^*)$ solves (17), then $(\boldsymbol{x}^*, \boldsymbol{u}^*, \alpha^*, \theta^*)$ solves (18), where $\boldsymbol{u}^*$ is constructed as

$$u_k^{i*} = [f(\boldsymbol{x}^*, \boldsymbol{y}_{[k]}^i) - \alpha^*]^+, \quad k = 1, \cdots, S^i, \quad i = 1, \cdots, l.$$

Proof. Suppose $(\boldsymbol{x}^*, \boldsymbol{u}^*, \alpha^*, \theta^*)$ is an optimal solution to (18) and $(\tilde{\boldsymbol{x}}^*, \tilde{\alpha}^*, \tilde{\theta}^*)$ is an optimal solution to (17).

Since $u_k^{i*} \geq f(\boldsymbol{x}^*, \boldsymbol{y}_{[k]}^i) - \alpha^*$ and $u_k^{i*} \geq 0$, it can be easily observed that

$$\alpha^* + \frac{1}{1-\beta} \sum_{k=1}^{S^i} \pi_k^i [f(\boldsymbol{x}^*, \boldsymbol{y}_{[k]}^i) - \alpha^*]^+ \leq \alpha^* + \frac{1}{1-\beta} \left(\boldsymbol{\pi}^i\right)^T \boldsymbol{u}^{i*} \leq \theta^*, \quad i = 1, \cdots, l,$$

which implies that $(\boldsymbol{x}^*, \alpha^*, \theta^*)$ is feasible to (17). Thus we have $\theta^* \geq \tilde{\theta}^*$, since $\tilde{\theta}^*$ is the optimal objective value of (17).

On the other hand, let

$$\tilde{u}_k^{i*} = [f(\tilde{\boldsymbol{x}}^*, \boldsymbol{y}_{[k]}^i) - \tilde{\alpha}^*]^+, \quad k = 1, \cdots, S^i, \quad i = 1, \cdots, l.$$

Then we immediately get

$$\begin{aligned} &\tilde{\alpha}^* + \frac{1}{1-\beta} \left(\boldsymbol{\pi}^i\right)^T \tilde{\boldsymbol{u}}^{i*} = \tilde{\alpha}^* + \frac{1}{1-\beta} \sum_{k=1}^{S^i} \pi_k^i [f(\tilde{\boldsymbol{x}}^*, \boldsymbol{y}_{[k]}^i) - \tilde{\alpha}^*]^+ \leq \tilde{\theta}^*, \\ &\tilde{u}_k^{i*} \geq f(\tilde{\boldsymbol{x}}^*, \boldsymbol{y}_{[k]}^i) - \tilde{\alpha}^*, \ \tilde{u}_k^{i*} \geq 0, \quad k = 1, \cdots, S^i, \quad i = 1, \cdots, l, \end{aligned}$$

which means that $(\tilde{\boldsymbol{x}}^*, \tilde{\boldsymbol{u}}^*, \tilde{\alpha}^*, \tilde{\theta}^*)$ is feasible to (18). Thus we have $\theta^* \leq \tilde{\theta}^*$, since θ^* is the optimal objective value of (18). Therefore, it must be ture that $\theta^* = \tilde{\theta}^*$, which means $(\boldsymbol{x}^*, \alpha^*, \theta^*)$ solves (17) and $(\tilde{\boldsymbol{x}}^*, \tilde{\boldsymbol{u}}^*, \tilde{\alpha}^*, \tilde{\theta}^*)$ solves (18). This completes the proof. □

Proposition 1 claims that solving the original non-smooth problem (17) can be substituted by solving a more tractable formulation (18). Furthermore, if $f(\boldsymbol{x}, \boldsymbol{y})$ is linear with respect to $\boldsymbol{x}$ and $\mathcal{X}$ is a convex polyhedron, then the problem can actually be solved via a linear programming approach.

Remark 1 *In the special case where $l = 1$, i.e., the distribution of $\boldsymbol{y}$ is confirmed to be only one scenario without any other possibility, problem (18) is exactly that of Rockafellar and Uryasev (2000) with $\pi_k^i = \frac{1}{S^1}$ for all k. Although in the above discussion we assume a continuous distribution, it is easy to see from Rockafellar and Uryasev (2002) that the results also hold in the general distribution case. For example, in the discrete distribution case, it is straightforward to interpret the integral as a summation. Moreover, it should be interpreted as a mixture of an integral and a summation in the case of mixed continuous and discrete distributions.*

2.2 Discrete Distribution

In this subsection we assume that $\boldsymbol{y}$ follows a discrete distribution and discuss the minimization of the worst-case CVaR under componentwise bounded uncertainty and ellipsoidal uncertainty. From a practical viewpoint, this consideration still makes sense for continuous distributions, since we usually use a discretization procedure to approximate the integral resulted form a continuous distribution.

Let the sample space of random vector $\boldsymbol{y}$ be given by $\left\{\boldsymbol{y}_{[1]}, \boldsymbol{y}_{[2]}, \cdots, \boldsymbol{y}_{[S]}\right\}$ with $\Pr\left\{\boldsymbol{y}_{[i]}\right\} = \pi_i$ and $\sum_{i=1}^{S} \pi_i = 1,\ \pi_i \geq 0,\ i = 1, \cdots, S$. Denote $\boldsymbol{\pi} = (\pi_1, \pi_2, \cdots, \pi_S)^T$ and define

$$G_\beta(\boldsymbol{x}, \alpha, \boldsymbol{\pi}) \triangleq \alpha + \frac{1}{1-\beta} \sum_{k=1}^{S} \pi_k [f(\boldsymbol{x}, \boldsymbol{y}_{[k]}) - \alpha]^+.$$

For given $\boldsymbol{x}$ and $\boldsymbol{\pi}$, the corresponding CVaR is then defined as (Rockafellar and Uryasev 2002)

$$\mathrm{CVaR}_\beta(\boldsymbol{x}, \boldsymbol{\pi}) \triangleq \min_{\alpha \in \mathcal{R}} G_\beta(\boldsymbol{x}, \alpha, \boldsymbol{\pi}).$$

Especially, we denote $\mathscr{P}$ as $\mathscr{P}_\pi$ in the case of discrete distribution. Then we may identify $\mathscr{P}_\pi$ as a subset of $\mathcal{R}^S$ and the worst-case CVaR for fixed $\boldsymbol{x} \in \mathcal{X}$ with respect to $\mathscr{P}_\pi$ is defined as

$$\mathrm{WCVaR}_\beta(\boldsymbol{x}) \triangleq \sup_{\boldsymbol{\pi} \in \mathscr{P}_\pi} \mathrm{CVaR}_\beta(\boldsymbol{x}, \boldsymbol{\pi}),$$

or equivalently,

$$\mathrm{WCVaR}_\beta(\boldsymbol{x}) \triangleq \sup_{\boldsymbol{\pi} \in \mathscr{P}_\pi} \min_{\alpha \in \mathcal{R}} G_\beta(\boldsymbol{x}, \alpha, \boldsymbol{\pi}).$$

Theorem 2 *Suppose $\mathscr{P}_\pi$ is a compact convex set. Then, for each $\boldsymbol{x}$, we have*

$$\mathrm{WCVaR}_\beta(\boldsymbol{x}) = \min_{\alpha \in \mathcal{R}} \max_{\boldsymbol{\pi} \in \mathscr{P}_\pi} G_\beta(\boldsymbol{x}, \alpha, \boldsymbol{\pi}).$$

Proof. Since $\mathscr{P}_\pi$ is bounded, it is contained in a polytope, i.e.,

$$\mathscr{P}_\pi \subseteq \left\{ \boldsymbol{\pi} \in \mathcal{R}^S :\ \boldsymbol{\pi} = \sum_{i=1}^{l} \lambda_i \boldsymbol{\pi}^i,\ \boldsymbol{\lambda} \in \Lambda \right\}$$

for some positive integer l and distribution scenarios $\{\boldsymbol{\pi}^i\}_{i=1}^{l}$, where Λ is given by (5). Therefore, for any given $\boldsymbol{\pi} \in \mathscr{P}_\pi$, using a similar argument to that in the first part of Theorem 1, we can show that

$$\min_{\alpha \in \mathcal{R}} G_\beta(\boldsymbol{x}, \alpha, \boldsymbol{\pi}) = \min_{\alpha \in \mathcal{A}} G_\beta(\boldsymbol{x}, \alpha, \boldsymbol{\pi}),$$

and hence

$$\max_{\boldsymbol{\pi}\in\mathscr{P}_\pi}\min_{\alpha\in\mathcal{R}} G_\beta(\boldsymbol{x},\alpha,\boldsymbol{\pi}) = \max_{\boldsymbol{\pi}\in\mathscr{P}_\pi}\min_{\alpha\in\mathcal{A}} G_\beta(\boldsymbol{x},\alpha,\boldsymbol{\pi}),$$

where $\mathcal{A}$ is a nonempty, closed and bounded interval.

For fixed $\boldsymbol{x}$, $G_\beta(\boldsymbol{x},\alpha,\boldsymbol{\pi})$ is convex in α (see Rockafellar and Uryasev 2002) and affine (concave) in $\boldsymbol{\pi}$. By Lemma 1, we get

$$\max_{\boldsymbol{\pi}\in\mathscr{P}_\pi}\min_{\alpha\in\mathcal{A}} G_\beta(\boldsymbol{x},\alpha,\boldsymbol{\pi}) = \min_{\alpha\in\mathcal{A}}\max_{\boldsymbol{\pi}\in\mathscr{P}_\pi} G_\beta(\boldsymbol{x},\alpha,\boldsymbol{\pi}).$$

Performing a further discussion similar to the proof of Theorem 1, we can show

$$\mathrm{WCVaR}_\beta(\boldsymbol{x}) = \min_{\alpha\in\mathcal{R}}\max_{\boldsymbol{\pi}\in\mathscr{P}_\pi} G_\beta(\boldsymbol{x},\alpha,\boldsymbol{\pi}).$$

This completes the proof. □

Theorem 2 indicates that, if $\mathscr{P}_\pi$ is a compact convex set, the problem of minimizing $\mathrm{WCVaR}_\beta(\boldsymbol{x})$ over $\mathcal{X}$ can be written as

$$\begin{aligned} &\min_{(\boldsymbol{x},\alpha,\theta)\in\mathcal{X}\times\mathcal{R}\times\mathcal{R}} \theta \\ &\text{s.t.} \quad \max_{\boldsymbol{\pi}\in\mathscr{P}_\pi} \alpha + \frac{1}{1-\beta}\sum_{k=1}^{S}\pi_k[f(\boldsymbol{x},\boldsymbol{y}_{[k]})-\alpha]^+ \le \theta. \end{aligned} \tag{19}$$

By introducing an auxiliary vector $\boldsymbol{u}\in\mathcal{R}^S$, we can show as in Proposition 1 that problem (19) is equivalent to the following minimization problem with variables $(\boldsymbol{x},\boldsymbol{u},\alpha,\theta)\in\mathcal{R}^n\times\mathcal{R}^S\times\mathcal{R}\times\mathcal{R}$:

$$\begin{aligned} \min\ & \theta \\ \text{s.t.}\ & \boldsymbol{x}\in\mathcal{X}, \\ & \max_{\boldsymbol{\pi}\in\mathscr{P}_\pi} \alpha + \frac{1}{1-\beta}\boldsymbol{\pi}^T\boldsymbol{u} \le \theta, \\ & u_k \ge f(\boldsymbol{x},\boldsymbol{y}_{[k]}) - \alpha, \\ & u_k \ge 0, \qquad k = 1,\cdots,S. \end{aligned} \tag{20}$$

Problem (20) is not ready for application because of the max operation involved in the constraints. In the following, we show that, under the componentwise bounded uncertainty and ellipsoidal uncertainty in the distributions, (20) can be cast as linear programs and second-order cone programs, respectively.

2.2.1 Componentwise Bounded Uncertainty

Suppose $\boldsymbol{\pi}$ belongs to a componentwise bounded set, i.e.,

$$\boldsymbol{\pi}\in\mathscr{P}_\pi^B \triangleq \left\{\boldsymbol{\pi}:\boldsymbol{\pi}=\boldsymbol{\pi}^0+\boldsymbol{\eta}, \boldsymbol{e}^T\boldsymbol{\eta}=0, \underline{\boldsymbol{\eta}}\le\boldsymbol{\eta}\le\overline{\boldsymbol{\eta}}\right\}, \tag{21}$$

where $\boldsymbol{\pi}^0$ is a nominal distribution which represents the most likely distribution, $\boldsymbol{e}$ denotes the vector of ones, and $\underline{\boldsymbol{\eta}}$ and $\overline{\boldsymbol{\eta}}$ are given constant vectors. The condition $\boldsymbol{e}^T\boldsymbol{\eta} = 0$ ensures $\boldsymbol{\pi}$ to be a probability distribution, and the non-negativity constraint $\boldsymbol{\pi} \geq 0$ is included in the bound constraints $\underline{\boldsymbol{\eta}} \leq \boldsymbol{\eta} \leq \overline{\boldsymbol{\eta}}$.

Since

$$\alpha + \frac{1}{1-\beta}\boldsymbol{\pi}^T\boldsymbol{u} = \alpha + \frac{1}{1-\beta}\left(\boldsymbol{\pi}^0\right)^T\boldsymbol{u} + \frac{1}{1-\beta}\boldsymbol{\eta}^T\boldsymbol{u},$$

we have

$$\max_{\boldsymbol{\pi}\in\mathscr{P}_\pi^B} \alpha + \frac{1}{1-\beta}\boldsymbol{\pi}^T\boldsymbol{u} = \alpha + \frac{1}{1-\beta}\left(\boldsymbol{\pi}^0\right)^T\boldsymbol{u} + \frac{\gamma^*(\boldsymbol{u})}{1-\beta},$$

where $\gamma^*(\boldsymbol{u})$ is the optimal value of the following linear program

$$\begin{aligned} &\max_{\boldsymbol{\eta}\in\mathcal{R}^S} \boldsymbol{u}^T\boldsymbol{\eta} \\ &\text{s.t.} \quad \boldsymbol{e}^T\boldsymbol{\eta} = 0, \\ &\qquad\ \ \underline{\boldsymbol{\eta}} \leq \boldsymbol{\eta} \leq \overline{\boldsymbol{\eta}}. \end{aligned} \tag{22}$$

The dual program of (22) is given by

$$\begin{aligned} &\min_{(z,\boldsymbol{\xi},\boldsymbol{\omega})\in\mathcal{R}\times\mathcal{R}^S\times\mathcal{R}^S} \overline{\boldsymbol{\eta}}^T\boldsymbol{\xi} + \underline{\boldsymbol{\eta}}^T\boldsymbol{\omega} \\ &\text{s.t.} \quad \boldsymbol{e}z + \boldsymbol{\xi} + \boldsymbol{\omega} = \boldsymbol{u}, \\ &\qquad\ \ \boldsymbol{\xi} \geq 0,\ \boldsymbol{\omega} \leq 0. \end{aligned} \tag{23}$$

Consider the following minimization problem over $(\boldsymbol{x}, \boldsymbol{u}, z, \boldsymbol{\xi}, \boldsymbol{\omega}, \alpha, \theta) \in \mathcal{R}^n \times \mathcal{R}^S \times \mathcal{R} \times \mathcal{R}^S \times \mathcal{R}^S \times \mathcal{R} \times \mathcal{R}$:

$$\begin{aligned} &\min\ \theta \\ &\text{s.t.} \quad \boldsymbol{x} \in \mathcal{X}, \\ &\qquad\ \ \alpha + \frac{1}{1-\beta}\left(\boldsymbol{\pi}^0\right)^T\boldsymbol{u} + \frac{1}{1-\beta}\left(\overline{\boldsymbol{\eta}}^T\boldsymbol{\xi} + \underline{\boldsymbol{\eta}}^T\boldsymbol{\omega}\right) \leq \theta, \\ &\qquad\ \ \boldsymbol{e}z + \boldsymbol{\xi} + \boldsymbol{\omega} = \boldsymbol{u}, \\ &\qquad\ \ \boldsymbol{\xi} \geq 0,\ \boldsymbol{\omega} \leq 0, \\ &\qquad\ \ u_k \geq f(\boldsymbol{x}, \boldsymbol{y}_{[k]}) - \alpha, \\ &\qquad\ \ u_k \geq 0, \qquad\qquad\qquad k = 1, \cdots, S. \end{aligned} \tag{24}$$

Proposition 2 *If* $(\boldsymbol{x}^*, \boldsymbol{u}^*, z^*, \boldsymbol{\xi}^*, \boldsymbol{\omega}^*, \alpha^*, \theta^*)$ *solves (24), then* $(\boldsymbol{x}^*, \boldsymbol{u}^*, \alpha^*, \theta^*)$ *solves (20) with* $\mathscr{P}_\pi = \mathscr{P}_\pi^B$*; Conversely, if* $(\tilde{\boldsymbol{x}}^*, \tilde{\boldsymbol{u}}^*, \tilde{\alpha}^*, \tilde{\theta}^*)$ *solves (20) with* $\mathscr{P}_\pi = \mathscr{P}_\pi^B$*, then* $(\tilde{\boldsymbol{x}}^*, \tilde{\boldsymbol{u}}^*, \tilde{z}^*, \tilde{\boldsymbol{\xi}}^*, \tilde{\boldsymbol{\omega}}^*, \tilde{\alpha}^*, \tilde{\theta}^*)$ *solves (24), where* $(\tilde{z}^*, \tilde{\boldsymbol{\xi}}^*, \tilde{\boldsymbol{\omega}}^*)$ *is an optimal solution to (23) with* $\boldsymbol{u} = \tilde{\boldsymbol{u}}^*$*.*

Proof. Let $(\boldsymbol{x}^*, \boldsymbol{u}^*, z^*, \boldsymbol{\xi}^*, \boldsymbol{\omega}^*, \alpha^*, \theta^*)$ solve (24). By the duality theorem of linear programming, we have

$$\gamma^*(\boldsymbol{u}^*) \leq \overline{\boldsymbol{\eta}}^T \boldsymbol{\xi}^* + \underline{\boldsymbol{\eta}}^T \boldsymbol{\omega}^*.$$

Thus

$$\begin{aligned}
&\max_{\boldsymbol{\pi} \in \mathscr{P}_\pi^B} \alpha^* + \frac{1}{1-\beta} \boldsymbol{\pi}^T \boldsymbol{u}^* \\
&= \alpha^* + \frac{1}{1-\beta} \left(\boldsymbol{\pi}^0\right)^T \boldsymbol{u}^* + \frac{\gamma^*(\boldsymbol{u}^*)}{1-\beta} \\
&\leq \alpha^* + \frac{1}{1-\beta} \left(\boldsymbol{\pi}^0\right)^T \boldsymbol{u}^* + \frac{1}{1-\beta} \left(\overline{\boldsymbol{\eta}}^T \boldsymbol{\xi}^* + \underline{\boldsymbol{\eta}}^T \boldsymbol{\omega}^*\right) \\
&\leq \theta^*,
\end{aligned}$$

which, together with other constraints in (24), implies that $(\boldsymbol{x}^*, \boldsymbol{u}^*, \alpha^*, \theta^*)$ is feasible to (20) with $\mathscr{P}_\pi = \mathscr{P}_\pi^B$.

Now assume $(\boldsymbol{x}^*, \boldsymbol{u}^*, \alpha^*, \theta^*)$ is not an optimal solution to (20) with $\mathscr{P}_\pi = \mathscr{P}_\pi^B$, i.e., there exists an optimal solution $(\bar{\boldsymbol{x}}^*, \bar{\boldsymbol{u}}^*, \bar{\alpha}^*, \bar{\theta}^*)$ to (20) such that

$$\bar{\theta}^* < \theta^*.$$

Let $(\bar{z}^*, \bar{\boldsymbol{\xi}}^*, \bar{\boldsymbol{\omega}}^*)$ be an optimal solution to (23) with $\boldsymbol{u} = \bar{\boldsymbol{u}}^*$. By the strong duality theorem of linear programming, we have

$$\begin{aligned}
&\bar{\alpha}^* + \frac{1}{1-\beta} \left(\boldsymbol{\pi}^0\right)^T \bar{\boldsymbol{u}}^* + \frac{1}{1-\beta} \left(\overline{\boldsymbol{\eta}}^T \bar{\boldsymbol{\xi}}^* + \underline{\boldsymbol{\eta}}^T \bar{\boldsymbol{\omega}}^*\right) \\
&= \bar{\alpha}^* + \frac{1}{1-\beta} \left(\boldsymbol{\pi}^0\right)^T \bar{\boldsymbol{u}}^* + \frac{\gamma^*(\bar{\boldsymbol{u}}^*)}{1-\beta} \\
&= \max_{\boldsymbol{\pi} \in \mathscr{P}_\pi^B} \bar{\alpha}^* + \frac{1}{1-\beta} \boldsymbol{\pi}^T \bar{\boldsymbol{u}}^* \\
&\leq \bar{\theta}^*,
\end{aligned}$$

which, together with other constraints in (20) and (23), implies that $(\bar{\boldsymbol{x}}^*, \bar{\boldsymbol{u}}^*, \bar{z}^*, \bar{\boldsymbol{\xi}}^*, \bar{\boldsymbol{\omega}}^*, \bar{\alpha}^*, \bar{\theta}^*)$ is feasible to (24). This contradicts the assumption that $(\boldsymbol{x}^*, \boldsymbol{u}^*, z^*, \boldsymbol{\xi}^*, \boldsymbol{\omega}^*, \alpha^*, \theta^*)$ is an optimal solution to (24) since $\bar{\theta}^* < \theta^*$. Thus $(\boldsymbol{x}^*, \boldsymbol{u}^*, \alpha^*, \theta^*)$ is an optimal solution to (20) with $\mathscr{P}_\pi = \mathscr{P}_\pi^B$.

Conversely, let $(\tilde{\boldsymbol{x}}^*, \tilde{\boldsymbol{u}}^*, \tilde{\alpha}^*, \tilde{\theta}^*)$ solve (20) with $\mathscr{P}_\pi = \mathscr{P}_\pi^B$, and let $(\tilde{z}^*, \tilde{\boldsymbol{\xi}}^*, \tilde{\boldsymbol{\omega}}^*)$ denote an optimal solution to (23) with $\boldsymbol{u} = \tilde{\boldsymbol{u}}^*$. Then $(\tilde{\boldsymbol{x}}^*, \tilde{\boldsymbol{u}}^*, \tilde{z}^*, \tilde{\boldsymbol{\xi}}^*, \tilde{\boldsymbol{\omega}}^*, \tilde{\alpha}^*, \tilde{\theta}^*)$ must solve (24). In fact, if this is not the case, then there exists an optimal solution $(\bar{x}^*, \bar{\boldsymbol{u}}^*, \bar{z}^*, \bar{\boldsymbol{\xi}}^*, \bar{\boldsymbol{\omega}}^*, \bar{\alpha}^*, \bar{\theta}^*)$ of (24) such that $\bar{\theta}^* < \tilde{\theta}^*$. From the discussion of the first part of the proof, $(\bar{\boldsymbol{x}}^*, \bar{\boldsymbol{u}}^*, \bar{\alpha}^*, \bar{\theta}^*)$ is an optimal solution of (20), which contradicts the assumption that $(\tilde{\boldsymbol{x}}^*, \tilde{\boldsymbol{u}}^*, \tilde{\alpha}^*, \tilde{\theta}^*)$ solves (20) since $\bar{\theta}^* < \tilde{\theta}^*$. This completes the proof. □

If $f(\boldsymbol{x}, \boldsymbol{y})$ is a convex function with respect to $\boldsymbol{x}$ and $\mathcal{X}$ is a convex set, then (24) is a convex program. Especially, if $f(\boldsymbol{x}, \boldsymbol{y})$ is linear in $\boldsymbol{x}$ and $\mathcal{X}$ is a convex polyhedron, then the problem is a linear program.

Remark 2 *In the special case where* $\underline{\boldsymbol{\eta}} = \overline{\boldsymbol{\eta}} = 0$*, (24) reduces to the original* CVaR *minimization problem.*

2.2.2 Ellipsoidal Uncertainty

Suppose $\boldsymbol{\pi}$ belongs to an ellipsoid, i.e.,

$$\boldsymbol{\pi} \in \mathscr{P}_{\pi}^{E} \triangleq \left\{ \boldsymbol{\pi} : \boldsymbol{\pi} = \boldsymbol{\pi}^0 + A\boldsymbol{\eta}, \boldsymbol{e}^T\boldsymbol{\eta} = 0, \boldsymbol{\pi}^0 + A\boldsymbol{\eta} \geq 0, \|\boldsymbol{\eta}\| \leq 1 \right\}, \tag{25}$$

where $\|\boldsymbol{\eta}\| = \sqrt{\boldsymbol{\eta}^T\boldsymbol{\eta}}$, $\boldsymbol{\pi}^0$ is a nominal distribution that is the center of the ellipsoid, $A \in \mathcal{R}^{S\times S}$ is the scaling matrix of the ellipsoid. The conditions $\boldsymbol{e}^T\boldsymbol{\eta} = 0$ and $\boldsymbol{\pi}^0 + A\boldsymbol{\eta} \geq 0$ ensure $\boldsymbol{\pi}$ to be a probability distribution.

Consider the following convex program

$$\begin{aligned} \max_{\boldsymbol{\eta}\in\mathcal{R}^S} \ & \boldsymbol{u}^T A\boldsymbol{\eta} \\ \text{s.t.} \quad & \boldsymbol{e}^T\boldsymbol{\eta} = 0, \\ & \boldsymbol{\pi}^0 + A\boldsymbol{\eta} \geq 0, \\ & \|\boldsymbol{\eta}\| \leq 1. \end{aligned} \tag{26}$$

The dual of (26) is the second-order cone program

$$\begin{aligned} \min_{(\zeta,\boldsymbol{\omega},\boldsymbol{\xi},z)\in\mathcal{R}\times\mathcal{R}^S\times\mathcal{R}^S\times\mathcal{R}} \ & \zeta + \left(\boldsymbol{\pi}^0\right)^T \boldsymbol{\omega} \\ \text{s.t.} \quad & -\boldsymbol{\xi} - A^T\boldsymbol{\omega} + \boldsymbol{e}z = A^T\boldsymbol{u}, \\ & \|\boldsymbol{\xi}\| \leq \zeta, \\ & \boldsymbol{\omega} \geq 0. \end{aligned} \tag{27}$$

One can refer to Lobo et al. (1998) and Alizadeh and Goldfarb (2003) for the details on second-order cone programming.

Consider the following minimization problem over $(\boldsymbol{x}, \boldsymbol{u}, \zeta, \boldsymbol{\omega}, \boldsymbol{\xi}, z, \alpha, \theta) \in \mathcal{R}^n \times \mathcal{R}^S \times \mathcal{R} \times \mathcal{R}^S \times \mathcal{R}^S \times \mathcal{R} \times \mathcal{R} \times \mathcal{R}$:

$$\begin{aligned} \min \ & \theta \\ \text{s.t.} \quad & \boldsymbol{x} \in \mathcal{X}, \\ & \alpha + \frac{1}{1-\beta}\left(\boldsymbol{\pi}^0\right)^T \boldsymbol{u} + \frac{1}{1-\beta}\left[\zeta + \left(\boldsymbol{\pi}^0\right)^T \boldsymbol{\omega}\right] \leq \theta, \\ & -\boldsymbol{\xi} - A^T\boldsymbol{\omega} + \boldsymbol{e}z = A^T\boldsymbol{u}, \\ & \|\boldsymbol{\xi}\| \leq \zeta, \ \boldsymbol{\omega} \geq 0, \\ & u_k \geq f(\boldsymbol{x}, \boldsymbol{y}_{[k]}) - \alpha, \\ & u_k \geq 0, \qquad\qquad k = 1, \cdots, S. \end{aligned} \tag{28}$$

This problem is similar to (24). Under some mild condition, such as the existence of interior feasible points for both (26) and (27), the zero duality gap is guaranteed by the strong conic duality theorem. In this case, we can prove the following theorem by using a similar argument to Proposition 2.

Proposition 3 *If* $(\boldsymbol{x}^*, \boldsymbol{u}^*, \zeta^*, \boldsymbol{\omega}^*, \boldsymbol{\xi}^*, z^*, \alpha^*, \theta^*)$ *solves (28), then* $(\boldsymbol{x}^*, \boldsymbol{u}^*, \alpha^*, \theta^*)$ *solves (20) with* $\mathscr{P}_\pi = \mathscr{P}_\pi^E$*; Conversely, if* $(\tilde{\boldsymbol{x}}^*, \tilde{\boldsymbol{u}}^*, \tilde{\alpha}^*, \tilde{\theta}^*)$ *solves (20) with* $\mathscr{P}_\pi = \mathscr{P}_\pi^E$*, then* $(\tilde{\boldsymbol{x}}^*, \tilde{\boldsymbol{u}}^*, \tilde{\zeta}^*, \tilde{\boldsymbol{\omega}}^*, \tilde{\boldsymbol{\xi}}^*, \tilde{z}^*, \tilde{\alpha}^*, \tilde{\theta}^*)$ *solves (28), where* $(\tilde{\zeta}^*, \tilde{\boldsymbol{\omega}}^*, \tilde{\boldsymbol{\xi}}^*, \tilde{z}^*)$ *is an optimal solution to (27) with* $\boldsymbol{u} = \tilde{\boldsymbol{u}}^*$.

If $f(\boldsymbol{x}, \boldsymbol{y})$ is a convex function with respect to $\boldsymbol{x}$ and $\mathcal{X}$ is a convex set, then (28) is a convex program. Furthermore, if $f(\boldsymbol{x}, \boldsymbol{y})$ is a linear function with respect to $\boldsymbol{x}$ and $\mathcal{X}$ is a convex polyhedron, then the problem is a second-order cone program that can be solved efficiently by interior-point methods developed in recent years.

Remark 3 *In the special case where* $A = 0$*, (28) reduces to the original* CVaR *minimization problem.*

3 Robust Portfolio Management Using Worst-Case CVaR

In this section we consider the situation that random returns of financial assets are just specified by a set of distributions, and formulate a portfolio management problem by utilizing worst-case CVaR as the measure of risk.

Suppose there exist n risk assets that can be chosen by the investor in the financial market. Let random vector $\boldsymbol{y} = (y_1, \cdots, y_n)^T \in \mathcal{R}^n$ denote the uncertain returns of the n risk assets, and $\boldsymbol{x} = (x_1, \cdots, x_n)^T \in \mathcal{R}^n$ denote the amount of the investments in the n risk assets decided by the investor. Thus the loss function is defined as

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

By definition, the portfolio return is the negative of the loss, i.e., $\boldsymbol{x}^T \boldsymbol{y}$.

Portfolio optimization tries to find an optimal trade-off between the risk and the return according to the investor's preference, while the robust portfolio selection is performed through the worst-case analysis of risk and return. Thus the robust portfolio selection problem using WCVaR as a risk measure can be represented as

$$\min_{\boldsymbol{x} \in \mathcal{X}} \text{WCVaR}(\boldsymbol{x}),$$

where $\mathcal{X}$ denotes the constraint on the portfolio position, which usually includes the requirement of the worst case minimum expected return. According to the discussion in the previous section, in order to complete the formulation of the robust portfolio selection model, we only need to specify the constraint set $\mathcal{X}$.

Suppose the investor has an initial wealth w_0. Thus the portfolio satisfies

$$\boldsymbol{e}^T \boldsymbol{x} = w_0. \tag{29}$$

To ensure diversification and satisfy the regulations, we impose the bound constraints on the portfolio

$$\underline{\boldsymbol{x}} \leq \boldsymbol{x} \leq \overline{\boldsymbol{x}}, \tag{30}$$

where $\underline{\boldsymbol{x}}$ and $\overline{\boldsymbol{x}}$ are the given lower and upper bounds on the portfolios.

Let μ be the worst-case minimum expected return required by the investor. Mathematically, this can be represented as

$$\min_{p(\cdot) \in \mathscr{P}} \mathrm{E}_p \left(\boldsymbol{x}^T \boldsymbol{y} \right) \geq \mu, \tag{31}$$

where E_p denotes the expectation operator with respect to the distribution $p(\cdot)$ of $\boldsymbol{y}$. Generally, $\mathcal{X}$ is specified by (29), (30) and (31), i.e.,

$$\mathcal{X} \triangleq \left\{ \boldsymbol{x} : \boldsymbol{e}^T \boldsymbol{x} = w_0, \ \underline{\boldsymbol{x}} \leq \boldsymbol{x} \leq \overline{\boldsymbol{x}}, \ \min_{p(\cdot) \in \mathscr{P}} \mathrm{E}_p \left(\boldsymbol{x}^T \boldsymbol{y} \right) \geq \mu \right\}. \tag{32}$$

3.1 Problem Formulations

In this subsection, we discuss robust portfolio selection problems corresponding to the three types of uncertainties described in the previous section. The problems are cast as linear programs and second-order programs.

3.1.1 Mixture Distribution Uncertainty

In the case of mixture distribution uncertainty given by (4), (31) can be written as

$$\sum_{i=1}^{l} \lambda_i \mathrm{E}_{p^i} \left(\boldsymbol{x}^T \boldsymbol{y} \right) \geq \mu, \quad \forall \boldsymbol{\lambda} \in \Lambda, \tag{33}$$

where Λ is defined by (5). By the definition of Λ, it is clear that any $\boldsymbol{x}$ satisfying (33) also satisfies

$$\mathrm{E}_{p^i} \left(\boldsymbol{x}^T \boldsymbol{y} \right) \geq \mu, \quad i = 1, \cdots, l. \tag{34}$$

On the other hand, if (34) holds, then for any $\boldsymbol{\lambda} \in \Lambda$, we have

$$\sum_{i=1}^{l} \lambda_i \mathrm{E}_{p^i} \left(\boldsymbol{x}^T \boldsymbol{y} \right) \geq \sum_{i=1}^{l} \lambda_i \mu = \mu.$$

So (33) is equivalent to (34). Let $\bar{\boldsymbol{y}}^i$ denote the expected value of $\boldsymbol{y}$ with respect to the distribution scenario $p^i(\cdot)$. Then (31) can be simply represented as

$$\boldsymbol{x}^T\bar{\boldsymbol{y}}^i \geq \mu, \quad i = 1, \cdots, l.$$

By (18), the robust portfolio selection problem, under the mixture distribution situation, is formulated as the following linear program with variables $(\boldsymbol{x}, \boldsymbol{u}, \alpha, \theta) \in \mathcal{R}^n \times \mathcal{R}^m \times \mathcal{R} \times \mathcal{R}$:

$$\begin{aligned} \min \ & \theta \\ \text{s.t.} \ & \boldsymbol{e}^T\boldsymbol{x} = w_0, \\ & \boldsymbol{x}^T\bar{\boldsymbol{y}}^i \geq \mu, \\ & \underline{\boldsymbol{x}} \leq \boldsymbol{x} \leq \overline{\boldsymbol{x}}, \\ & \alpha + \frac{1}{1-\beta}\left(\boldsymbol{\pi}^i\right)^T \boldsymbol{u}^i \leq \theta, \\ & u_k^i + \boldsymbol{x}^T\boldsymbol{y}_{[k]}^i + \alpha \geq 0, \\ & u_k^i \geq 0, \qquad k = 1, \cdots, S^i, \quad i = 1, \cdots, l. \end{aligned} \tag{35}$$

3.1.2 Componentwise Bounded Uncertainty in Discrete Distributions

Denote

$$Y = \begin{bmatrix} \boldsymbol{y}_{[1]}^T \\ \vdots \\ \boldsymbol{y}_{[S]}^T \end{bmatrix}. \tag{36}$$

In the case of the componentwise bounded uncertainty in discrete distributions, by (21) and (32), $\mathcal{X}$ is given by

$$\mathcal{X}_B \triangleq \left\{ \boldsymbol{x} : \ \boldsymbol{e}^T\boldsymbol{x} = w_0, \ \underline{\boldsymbol{x}} \leq \boldsymbol{x} \leq \overline{\boldsymbol{x}}, \ (Y\boldsymbol{x})^T\boldsymbol{\pi}^0 + \min_{\{\boldsymbol{\eta}:\ \boldsymbol{e}^T\boldsymbol{\eta}=0,\ \underline{\boldsymbol{\eta}}\leq\boldsymbol{\eta}\leq\overline{\boldsymbol{\eta}}\}} (Y\boldsymbol{x})^T\boldsymbol{\eta} \geq \mu \right\}.$$

The dual problem of the linear program

$$\begin{aligned} \min_{\boldsymbol{\eta}\in\mathcal{R}^S} \ & (Y\boldsymbol{x})^T\boldsymbol{\eta} \\ \text{s.t.} \ & \boldsymbol{e}^T\boldsymbol{\eta} = 0, \\ & \underline{\boldsymbol{\eta}} \leq \boldsymbol{\eta} \leq \overline{\boldsymbol{\eta}} \end{aligned}$$

is written as

$$\begin{aligned} \max_{(\delta,\boldsymbol{\tau},\boldsymbol{\nu})\in\mathcal{R}\times\mathcal{R}^S\times\mathcal{R}^S} \ & \overline{\boldsymbol{\eta}}^T\boldsymbol{\tau} + \underline{\boldsymbol{\eta}}^T\boldsymbol{\nu} \\ \text{s.t.} \ & \boldsymbol{e}\delta + \boldsymbol{\tau} + \boldsymbol{\nu} = Y\boldsymbol{x}, \\ & \boldsymbol{\tau} \leq 0, \ \boldsymbol{\nu} \geq 0. \end{aligned}$$

Define

$$\Phi^B \triangleq \left\{ (\boldsymbol{x}, \delta, \boldsymbol{\tau}, \boldsymbol{\nu}) : \begin{array}{l} \boldsymbol{e}^T\boldsymbol{x} = w_0,\ \underline{\boldsymbol{x}} \le \boldsymbol{x} \le \overline{\boldsymbol{x}},\ \boldsymbol{e}\delta + \boldsymbol{\tau} + \boldsymbol{\nu} = Y\boldsymbol{x}, \\ \boldsymbol{\tau} \le 0,\ \boldsymbol{\nu} \ge 0,\ (Y\boldsymbol{x})^T\boldsymbol{\pi}^0 + \overline{\boldsymbol{\eta}}^T\boldsymbol{\tau} + \underline{\boldsymbol{\eta}}^T\boldsymbol{\nu} \ge \mu \end{array} \right\}$$

and

$$\Phi^B_{\mathcal{X}} \triangleq \left\{ \boldsymbol{x} : \exists(\delta, \boldsymbol{\tau}, \boldsymbol{\nu}) \text{ such that } (\boldsymbol{x}, \delta, \boldsymbol{\tau}, \boldsymbol{\nu}) \in \Phi^B \right\}.$$

By the duality theory of linear programming, it is easy to see that

$$\mathcal{X}_B = \Phi^B_{\mathcal{X}}. \tag{37}$$

By (24) and (37), the robust portfolio selection problem can be written as the following linear program with variables $(\boldsymbol{x}, \boldsymbol{u}, z, \boldsymbol{\xi}, \boldsymbol{\omega}, \alpha, \theta, \delta, \boldsymbol{\tau}, \boldsymbol{\nu}) \in \mathcal{R}^n \times \mathcal{R}^S \times \mathcal{R} \times \mathcal{R}^S \times \mathcal{R}^S \times \mathcal{R} \times \mathcal{R} \times \mathcal{R} \times \mathcal{R}^S \times \mathcal{R}^S$:

$$\begin{array}{ll} \min & \theta \\ \text{s.t.} & \boldsymbol{e}^T\boldsymbol{x} = w_0, \\ & \underline{\boldsymbol{x}} \le \boldsymbol{x} \le \overline{\boldsymbol{x}}, \\ & \left(\boldsymbol{\pi}^0\right)^T Y\boldsymbol{x} + \overline{\boldsymbol{\eta}}^T\boldsymbol{\tau} + \underline{\boldsymbol{\eta}}^T\boldsymbol{\nu} \ge \mu, \\ & \boldsymbol{e}\delta + \boldsymbol{\tau} + \boldsymbol{\nu} = Y\boldsymbol{x}, \\ & \boldsymbol{\tau} \le 0,\ \boldsymbol{\nu} \ge 0, \\ & \alpha + \frac{1}{1-\beta}\left(\boldsymbol{\pi}^0\right)^T \boldsymbol{u} + \frac{1}{1-\beta}\left(\overline{\boldsymbol{\eta}}^T\boldsymbol{\xi} + \underline{\boldsymbol{\eta}}^T\boldsymbol{\omega}\right) \le \theta, \\ & \boldsymbol{e}z + \boldsymbol{\xi} + \boldsymbol{\omega} = \boldsymbol{u}, \\ & \boldsymbol{\xi} \ge 0,\ \boldsymbol{\omega} \le 0, \\ & u_k + \boldsymbol{x}^T\boldsymbol{y}_{[k]} + \alpha \ge 0, \\ & u_k \ge 0, \qquad\qquad k = 1, \cdots, S. \end{array} \tag{38}$$

3.1.3 Ellipsoidal Uncertainty in Discrete Distributions

In the case of the ellipsoidal uncertainty in discrete distributions, by (25) and (32), $\mathcal{X}$ is given by

$$\mathcal{X}_E \triangleq \left\{ \boldsymbol{x} : \begin{array}{l} \boldsymbol{e}^T\boldsymbol{x} = w_0,\ \underline{\boldsymbol{x}} \le \boldsymbol{x} \le \overline{\boldsymbol{x}}, \\ (Y\boldsymbol{x})^T\boldsymbol{\pi}^0 + \min\limits_{\{\boldsymbol{\eta}:\ \boldsymbol{e}^T\boldsymbol{\eta}=0,\ \boldsymbol{\pi}^0+A\boldsymbol{\eta}\ge 0,\ \|\boldsymbol{\eta}\|\le 1\}} (Y\boldsymbol{x})^T A\boldsymbol{\eta} \ge \mu \end{array} \right\},$$

where Y is defined by (36). The dual program of the second-order cone program

$$\begin{array}{ll} \min\limits_{\boldsymbol{\eta}\in\mathcal{R}^S} & (Y\boldsymbol{x})^T A\boldsymbol{\eta} \\ \text{s.t.} & \boldsymbol{e}^T\boldsymbol{\eta} = 0, \\ & \boldsymbol{\pi}^0 + A\boldsymbol{\eta} \ge 0, \\ & \|\boldsymbol{\eta}\| \le 1. \end{array}$$

is given by

$$\begin{aligned}
&\max_{(\sigma,\boldsymbol{\tau},\boldsymbol{\nu},\delta)\in\mathcal{R}\times\mathcal{R}^S\times\mathcal{R}^S\times\mathcal{R}} -\sigma-\left(\boldsymbol{\pi}^0\right)^T\boldsymbol{\tau}\\
&\text{s.t.}\quad \boldsymbol{\nu}+A^T\boldsymbol{\tau}+\boldsymbol{e}\delta=A^TY\boldsymbol{x},\\
&\qquad\ \|\boldsymbol{\nu}\|\le\sigma,\\
&\qquad\ \boldsymbol{\tau}\ge 0.
\end{aligned}$$

Define

$$\Phi^E\triangleq\left\{\ (\boldsymbol{x},\sigma,\boldsymbol{\tau},\boldsymbol{\nu},\delta):\quad \begin{array}{l}\boldsymbol{e}^T\boldsymbol{x}=w_0,\ \underline{\boldsymbol{x}}\le\boldsymbol{x}\le\overline{\boldsymbol{x}},\ \boldsymbol{\nu}+A^T\boldsymbol{\tau}+\boldsymbol{e}\delta=A^TY\boldsymbol{x},\\ \|\boldsymbol{\nu}\|\le\sigma,\ \boldsymbol{\tau}\ge 0,\ (Y\boldsymbol{x})^T\boldsymbol{\pi}^0-\sigma-\left(\boldsymbol{\pi}^0\right)^T\boldsymbol{\tau}\ge\mu\end{array}\right\}$$

and

$$\Phi^E_{\mathcal{X}}\triangleq\left\{\boldsymbol{x}:\exists(\sigma,\boldsymbol{\tau},\boldsymbol{\nu},\delta)\text{ such that }(\boldsymbol{x},\sigma,\boldsymbol{\tau},\boldsymbol{\nu},\delta)\in\Phi^E\right\}.$$

By the conic duality theory, under some mild condition that guarantees zero duality gap, we have

$$\mathcal{X}_B=\Phi^E_{\mathcal{X}}. \tag{39}$$

By (28) and (39), the robust portfolio selection problem can be written as the following second-order cone program with variables $(\boldsymbol{x},\boldsymbol{u},\zeta,\boldsymbol{\omega},\boldsymbol{\xi},z,\alpha,\theta,\sigma,\boldsymbol{\tau},\boldsymbol{\nu},\delta)\in\mathcal{R}^n\times\mathcal{R}^S\times\mathcal{R}\times\mathcal{R}^S\times\mathcal{R}^S\times\mathcal{R}\times\mathcal{R}\times\mathcal{R}\times\mathcal{R}\times\mathcal{R}^S\times\mathcal{R}^S\times\mathcal{R}$:

$$\begin{aligned}
&\min\ \theta\\
&\text{s.t.}\quad \boldsymbol{e}^T\boldsymbol{x}=w_0,\\
&\qquad \underline{\boldsymbol{x}}\le\boldsymbol{x}\le\overline{\boldsymbol{x}},\\
&\qquad \left(\boldsymbol{\pi}^0\right)^TY\boldsymbol{x}-\sigma-\left(\boldsymbol{\pi}^0\right)^T\boldsymbol{\tau}\ge\mu,\\
&\qquad \boldsymbol{\nu}+A^T\boldsymbol{\tau}+\boldsymbol{e}\delta=A^TY\boldsymbol{x},\\
&\qquad \|\boldsymbol{\nu}\|\le\sigma,\ \boldsymbol{\tau}\ge 0,\\
&\qquad \alpha+\frac{1}{1-\beta}\left(\boldsymbol{\pi}^0\right)^T\boldsymbol{u}+\frac{1}{1-\beta}\left[\zeta+\left(\boldsymbol{\pi}^0\right)^T\boldsymbol{\omega}\right]\le\theta,\\
&\qquad -\boldsymbol{\xi}-A^T\boldsymbol{\omega}+\boldsymbol{e}z=A^T\boldsymbol{u},\\
&\qquad \|\boldsymbol{\xi}\|\le\zeta,\ \boldsymbol{\omega}\ge 0,\\
&\qquad u_k+\boldsymbol{x}^T\boldsymbol{y}_{[k]}+\alpha\ge 0,\\
&\qquad u_k\ge 0, \qquad\qquad k=1,\cdots,S.
\end{aligned} \tag{40}$$

3.2 Numerical Examples

In this subsection, we consider two numerical examples to illustrate the robust portfolio optimization problems. Market data simulation analysis and Monte Carlo simulation analysis are presented.

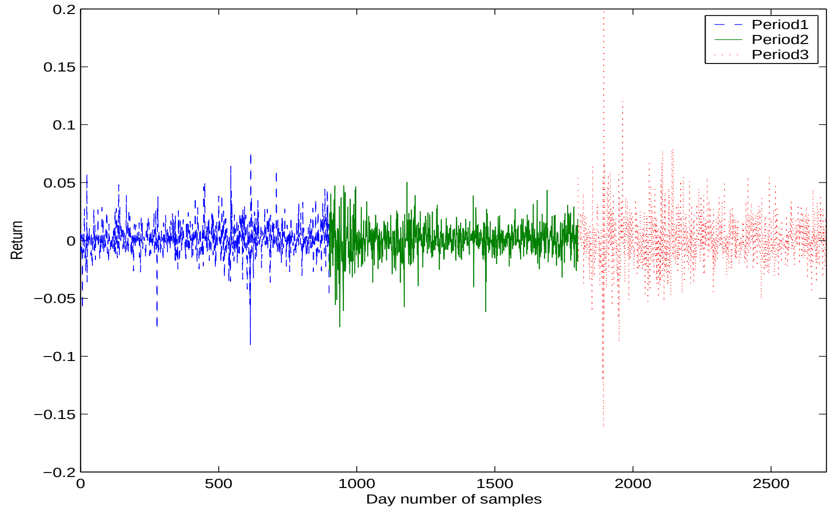

Figure 1: Return of Hang Seng Finance Index of SEHK (27/07/1990 ∼ 11/30/2000).

3.2.1 Market Data Simulation Analysis

The four sectoral sub-indices of Hang Seng Index of Hong Kong Stock Exchange (SEHK): (1) Hang Seng Finance Index (HSNF), (2) Hang Seng Utilities Index (HSNU), (3) Hang Seng Property Index (HSNP), and (4) Hang Seng Commercial/Industrial Index (HSNC), are chosen as the financial assets to construct the portfolios. We consider the day returns of these assets in the example. 2700 samples of returns of these four assets are collected from the time period from July 27, 1990 to November 30, 2000. Figure 1 is constructed by the samples of day return of HSNF. The day returns of other three assets behave similarly to that of HSNF. It can be roughly observed from Figure 1 that the behaviour of returns is not consistent among different time periods. According to this observation, we divide the time period into the following three sub-intervals (900 samples for each time period):

- Period1: 07/27/1990 ∼ 01/06/1994;
- Period2: 01/07/1994 ∼ 06/19/1997;
- Period3: 06/20/1997 ∼ 11/30/2000.

Within each time period, the returns behave similarly, whereas they exhibit remarkable difference between any two time periods.

The expected values and variances of returns of the four assets corresponding to different time periods are listed in Table 1. We find that the expected return of Period1 and the volatility of Period3 are much larger than those of the other two time periods. In this example, the estimation of the statistical parameters is not stable. Thus it is questionable to assume that all the samples are generated by an identical nominal probability distribution. Consequently, the original CVaR, as the measure of risk, is not reliable if all those samples are

Table 1: Expected value and variance of returns of four indices in different time periods.

Period	Mean (10^{-3})				Variance (10^{-3})			
	HSNF	HSNU	HSNP	HSNC	HSNF	HSNU	HSNP	HSNC
Period1	1.8455	1.2522	1.5859	1.1383	0.2106	0.1947	0.2417	0.2062
Period2	0.7264	0.1648	0.2902	0.2625	0.1926	0.1823	0.2740	0.2110
Period3	0.5104	0.6743	-0.1271	0.1872	0.5346	0.4352	0.8138	0.8010

used directly in calculation since the underlying assumption that the probability distribution is precisely known to be a nominal one is violated. In this situation, it is reasonable to assume a mixture distribution of the random returns, and it makes sense for us to perform a worst-case CVaR minimization.

In the example, according to our observation, we assume that the samples are generated by the mixture distribution of three probability distribution scenarios. The samples within each time period are assumed to be generated by the corresponding probability distribution scenario.

SeduMi1.05 (Sturm 2001), a package developed by J. Sturm for optimization over symmetric cones, is employed in our computation. The numerical experiments are implemented on PC (1.5G RAM, CPU 3.06GHz). All the problems are successfully solved within 10 seconds. Especially, the linear programs obtained from the the market data simulation analysis are always solved within 4 seconds.

In this example, we set $\beta = 0.95$, $w_0 = 1$, $\underline{\boldsymbol{x}} = (0, 0, 0, 0)^T$ and $\overline{\boldsymbol{x}} = (1, 1, 1, 1)^T$. Numerical experiments for the nominal and the robust portfolio optimization problems are performed via the linear programming model (35). The former employs the original CVaR as the risk measure, while the latter uses the worst-case CVaR. In the computation of the nominal portfolio optimization problem, we set $l = 1$ and $S^1 = 2700$, i.e., all the samples are used in the model by assuming that they are generated by one nominal probability distribution. In the computation of the robust portfolio optimization problem, we set $l = 3$ and $S^1 = S^2 = S^3 = 900$, where we assume the samples within each time period are generated by the corresponding distribution scenario.

To compare the performances of the nominal portfolio optimization problem and the robust portfolio optimization problem, for various values of the required worst-case expected return μ and for each time period, Table 2 shows the expected values and the CVaRs at confidence level 0.95 of the corresponding portfolios. It is obvious that the larger the required minimal expected/worst-case expected return is set, the larger the associated risk would be. From the expected values, we find that the robust optimal portfolio policy always guarantees

Table 2: Comparison of performances of nominal optimal and robust optimal portfolios.

μ (10^{-3})	Robust (I) Nominal (II)	Mean (10^{-3}) Period1	Period2	Period3	$\text{CVaR}_{0.95}$ Period1	Period2	Period3
0	I	1.3546	0.2618	0.6460	0.0299	0.0299	0.0425
	II	1.4455	0.3478	0.6209	0.0293	0.0295	0.0427
0.5	I	1.6063	0.5000	0.5765	0.0292	0.0299	0.0441
	II	1.4455	0.3478	0.6209	0.0293	0.0295	0.0427
0.55	I	1.6591	0.5500	0.5619	0.0294	0.0304	0.0448
	II	1.4455	0.3478	0.6209	0.0293	0.0295	0.0427
0.95	I	—	—	—	—	—	—
	II	1.7064	0.5948	0.5488	0.0297	0.0308	0.0455

the required worst-case expected value of μ. However the nominal optimal portfolio policy usually results in a small worst-case expected value although it may have a large expected value (see lines for $\mu = 0.0005,\ 0.00055,\ 0.00095$). For the same value of μ, the risk of the robust optimal portfolio policy appears to be larger than the risk of the nominal optimal portfolio policy. It should be mentioned that we only list the CVaRs calculated according to the three sets of samples, and they do not necessarily reveal the real worst-case CVaRs. However, the larger risk is usually rewarded by a higher return. Figure 2 illustrates the evolution of the values of the robust optimal portfolio and the nominal optimal portfolio generated by setting $\mu = 0.0005$. It shows that the robust optimal portfolio almost always outperforms the nominal optimal portfolio. For $\mu = 0.00095$, the robust portfolio optimization problem is infeasible. But we find that, in the sense of worst-case trade-off, the nominal optimal policy generated by setting $\mu = 0.00095$ is dominated by the robust optimal policy generated by setting $\mu = 0.00055$, since we have $0.0005500 > 0.0005488$ for the "worst-case" expected returns and $0.0448 < 0.0455$ for the "worst-case" CVaRs. This together with Figure 2 suggests that the worst-case requirement in the robust portfolio formulation does not affect the average performance of the portfolio substantially.

3.2.2 Monte Carlo Simulation Analysis

In this part, we perform a Monte Carlo simulation analysis for the robust portfolio optimization model under the ellipsoidal uncertainty in distributions. Notice that a nonempty ellipsoid must contain a smaller box, and at the same time, must be contained by a bigger box. Thus, for both the ellipsoidal and componentwise bounded uncertainties, it is pre-

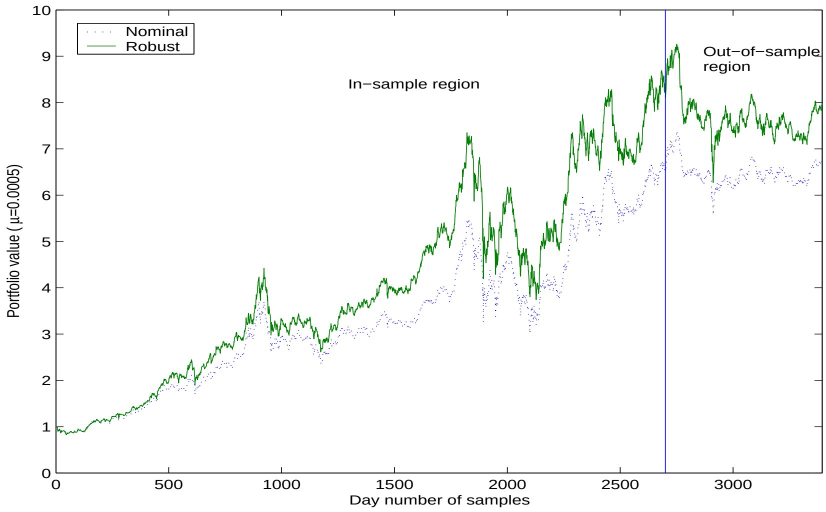

Figure 2: Evolution of values of robust optimal and nominal optimal portfolios ($\mu = 0.0005$).

Table 3: Expected returns.

Asset	Expected value
S&P	0.0101110
Gov Bond	0.0043532
Small Cap	0.0137058

dictable that the simulation results will be similar to each other. As shown in the previous section, the ellipsoidal uncertainty yields a second-order cone program which is more complex than a linear program resulting from the componentwise bounded uncertainty. To reduce the duplicate statements and verify the computational efficiency, we only consider here the case of ellipsoidal uncertainty, i.e, the second-order cone programming model (40).

We take the example given by Rockafellar and Uryasev (2000), where the portfolio is to be constructed by three assets: S&P 500, a portfolio of long-term U.S. government bonds, and a portfolio of small-cap stocks. The expected value and the covariance matrix of returns of these three assets are given in Tables 3 and 4, respectively.

In the example, the discrete sample space of random returns consists of 1000 samples, which are generated via the Monte Carlo simulation approach by assuming a joint normal distribution. We set $\beta = 0.95$, $w_0 = 1$, $\underline{\boldsymbol{x}} = (0, 0, 0)^T$ and $\overline{\boldsymbol{x}} = (1, 1, 1)^T$. For the sake of simplicity, the scaling matrix of the ellipsoid A is assume to be a diagonal matrix ρI_3, where ρ is a nonnegative scalar and I_3 is the 3×3 identity matrix. The larger the value of ρ is, the more uncertain the distribution becomes.

Table 4: Covariance matrix of returns.

	S&P	Gov Bond	Small Cap
S&P	0.00324652	0.00022983	0.00420395
Gov Bond	0.00022983	0.00049937	0.00019247
Small Cap	0.00420395	0.00019247	0.00764097

It should be mentioned that the nominal optimal portfolio is obtained by solving model (40) with $A = 0$, i.e., $\rho = 0$. It can also be obtained by solving model (35). The worst-case CVaR of the nominal optimal portfolio is obtained from solving model (40) by setting $x =$ "nominal optimal portfolio". For both nominal optimal and robust optimal portfolios, we get a set of global minimal worst-case CVaRs associated with different values of ρ, where we set $\mu = -2$ since all the obtained worst-case expected returns are greater than -2. A part of the numerical results is illustrated in Figure 3, which shows that the worst-case CVaR/risk grows as the value of the uncertain parameter ρ increases. More important observation is that the gap between the two curves becomes larger as ρ increases, which demonstrates the advantage of the robust optimization formulation in the situation where the uncertainty grows.

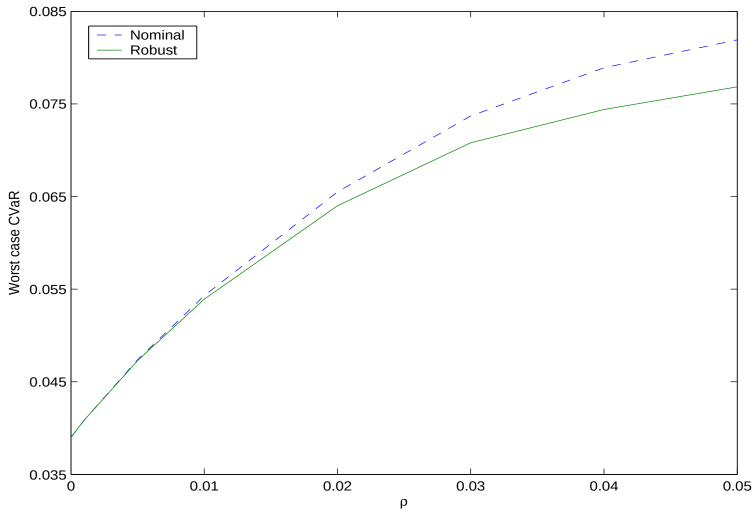

Figure 3: Worst-case CVaR of nominal optimal and robust optimal portfolios.

Table 5 shows a part of the comparison results corresponding to several values of μ and ρ. The phenomenon demonstrated by Figure 3 can also be observed in Table 5. Both nominal optimal and robust optimal portfolios become infeasible when either μ or ρ becomes large. However, by comparison, the robustness of the robust optimal portfolios is evidenced.

Table 5: Worst-case CVaR of nominal optimal and robust optimal portfolios according to different values of μ and ρ.

μ	ρ					
	0.001		0.003		0.005	
	Robust	Nominal	Robust	Nominal	Robust	Nominal
0	0.040870	0.040874	0.044218	0.044257	0.047295	0.047370
0.002	0.040870	0.040874	0.044218	0.044257	0.047445	—
0.004	0.040870	0.040874	0.049347	—	—	—
0.005	0.041453	—	0.096214	—	—	—
0.007	0.069936	—	—	—	—	—

4 Conclusions and Future Directions

This paper focuses on the worst-case CVaR minimization problem for the purpose of dealing with the uncertainty of the probability distributions. Application to robust portfolio optimization is also demonstrated. In comparison with the original CVaR, numerical experiments imply that the portfolio selection model using the worst-case CVaR as the risk measure performs robustly in practice, and provides more flexibility in portfolio decision analysis.

However, we only present here a simple application of worst-case CVaR to portfolio optimization. Many other applications of worst-case CVaR in financial optimization and risk management, such as hedging, index tracking and credit risk management, can also be easily implemented. Moreover, our approach is suitable for modeling the decentralized risk management problems, portfolio selection problems with uncertain exit time, and even those out of financial area.

How to determine the descriptions of the uncertainties, more specifically, how to determine the scenario distributions for the mixture distribution uncertainty, the bounds for the box uncertainty and the scaling matrix for the ellipsoidal uncertainty, are issues left for further investigations. Those are the key factors for the successful practical applications.

We can also formulate the robust portfolio optimization problem in the form of maximizing the worst-case expected return with constraint on the worst-case CVaR. For example, in the case of the mixture distribution uncertaity, noting that

$$\mathrm{WCVaR}_\beta(\boldsymbol{x}) = \min_{\alpha\in\mathcal{R}} \max_{i\in\mathcal{L}} F_\beta^i(\boldsymbol{x},\alpha) \le \theta$$

if and only if there exists α such that

$$\max_{i\in\mathcal{L}} F_\beta^i(\boldsymbol{x},\alpha) \le \theta,$$

we can formulate the corresponding robust portfolio selection problem as the following linear program with variables $(\boldsymbol{x}, \boldsymbol{u}, \alpha, \mu) \in \mathcal{R}^n \times \mathcal{R}^m \times \mathcal{R} \times \mathcal{R}$:

$$\begin{aligned}
\max \quad & \mu \\
\text{s.t.} \quad & \boldsymbol{e}^T \boldsymbol{x} = w_0, \\
& \boldsymbol{x}^T \bar{\boldsymbol{y}}^i \geq \mu, \\
& \underline{\boldsymbol{x}} \leq \boldsymbol{x} \leq \overline{\boldsymbol{x}}, \\
& \alpha + \frac{1}{1-\beta} \left(\boldsymbol{\pi}^i\right)^T \boldsymbol{u}^i \leq \theta, \\
& u_k^i + \boldsymbol{x}^T \boldsymbol{y}_{[k]}^i + \alpha \geq 0, \\
& u_k^i \geq 0, \qquad k = 1, \cdots, S^i, \quad i = 1, \cdots, l,
\end{aligned}$$

where θ is a predetermined bound on the worst-case CVaR. The robust portfolio optimization problem of this form can be similarly formulated as a linear program and a second-order cone program for the other two types of uncertainties.

Acknowledgments

This work is partly supported by the Informatics Research Center for Development of Knowledge Society Infrastructure, Graduate School of Informatics, Kyoto University, Japan. The work of the first author is also supported by the National Science Foundation of China (No. 70401009). The work of the second author is also supported by a Grant-in-Aid for Scientific Research from Japan Society for the Promotion of Science.

References

[1] Acerbi, C., D. Tasche. 2002. On the coherence of expected shortfall. *Journal of Banking and Finance* **26** 1487-1503.

[2] Alizadeh, F., D. Goldfarb. 2003. Second-order cone programming. *Mathematical Programming* Ser. B **95** 3-51.

[3] Andersson, F., H. Mausser, D. Rosen, S. Uryasev. 2001. Credit risk optimization with conditional value at risk criterion. *Mathematical Programming* Ser. B **89** 273-291.

[4] Artzner, P., F. Delbaen, J. M. Eber, D. Heath. 1999. Coherence measures of risk. *Mathematical Finance* **9** 203-228.

[5] Bazaraa, M. S., H. D. Sherali, C. M. Shetty. 1993. *Nonlinear Programming: Theory and Algorithms*, Second Edition. John Wiley & Sons, New York.

[6] Ben-Tal, A., A. Nemirovski. 2002. Robust optimization — methodology and applications. *Mathematical Programming* Ser. B **92** 453-480.

[7] Ben-Tal, A., T. Margalit, A. Nemirovski. 1999. Robust modeling of multi-stage portfolio problems. *High Performance Optimization Techniques*, Chapter 12. Eds. J. B. G. Frenk, K. Roos, T. Terlaky, S. Z. Zhang, Kluwer Academic Publishers.

[8] Black, F., R. Litterman. 1992. Global portfolio optimization. *Financial Analysts Journal* **48** 28-43.

[9] Bogentoft, E., H. E. Romeijn, S. Uryasev. 2001. Asset/Liability management for pension funds using CVaR constraints. *Journal of Risk Finance* **3** 57-71.

[10] Costa, O. L. V. and A. C. Paiva. 2002. Robust portfolio selection using linear-matrix inequality. *Journal of Economic Dynamics & Control* **26** 889-909.

[11] El Ghaoui, L., M. Oks, F. Oustry. 2003. Worst-case Value-at-Risk and robust portfolio optimization: A conic programming approach. *Operations Research* **51** 543-556.

[12] Fan, K. 1953. Minimax theorems. *Proceedings of National Academy of Science* **39** 42-47.

[13] Goldfarb, D., G. Iyengar. 2003. Robust portfolio selection problems. *Mathematics of Operations Research* **28** 1-38.

[14] Hall, J. A., B. W. Brorsen, S. H. Irwin. 1989. The distribution of futures prices: A test of the stable Paretian and mixture of normals hypotheses. *Journal of Financial and Quantitative Analysis* **24** 105-116.

[15] Halldórsson, B. V., R. H. Tütüncü. 2003. An interior-point method for a class of saddle-point problems. *Journal of Optimization Theory and Applications* **116** 559-590.

[16] Konno, H., H. Waki, A. Yuuki. 2002. Portfolio optimization under lower partial risk measures. *Asia-Pacific Financial Markets* **9** 127-140.

[17] Lobo, M. S., S. Boyd. 2000. The worst-case risk of a portfolio. Technical Report, http://faculty.fuqua.duke.edu/∼mlobo/bio/researchfiles/rsk-bnd.pdf.

[18] Lobo, M. S., L. Vandenberghe, S. Boyd, H. Lebret. 1998. Applications of second-order cone programming. *Linear Algebra and Its Applications* **284** 193-228.

[19] Markowitz, H. M. 1952. Portfolio selection. *Journal of Finance* **7** 77-91.

[20] Martellini, L., B. Uroševió. 2001. Portfolio selection with uncertain time of exit. Working paper, http://www-rcf.usc.edu/∼martelli/papers/psute.pdf.

[21] Mulvey, J., G. Erkan. 2003. Decentralized risk management for global P/C insurance companies. Working paper, Department of Operations Research and Financial Engineering, Princeton University.

[22] Ogryczak, W., A. Ruszczyński. 2002. Dual stochastic dominance and related mean-risk models. *SIAM Journal on Optimization* **13** 60-78.

[23] Peel D., G. J. McLachlan. 2000. Robust mixture modelling using the t distribution. *Statistics and Computing* **10** 339-348.

[24] Pflug, G. 2000. Some remarks on the Value-at-Risk and conditional Value-at-Risk. *Probabilistic Constrained Optimization: Methodology and Applications*. Ed. S. Uryasev, Kluwer Academic Publishers, Dordrecht.

[25] Philippe, J. 1996. *Value at Risk: The New Benchmark for Controlling Market Risk*. Irwin Professional Publishing, Chicago.

[26] RiskMetrics^{TM}. 1996. *Technical Document*, 4-th Edition. J. P. Morgan.

[27] Rockafellar, R.T., S. Uryasev. 2000. Optimization of conditional Value-at-Risk. *Journal of Risk* **2** 21-41.

[28] Rockafellar, R. T., S. Uryasev 2002. Conditional Value-at-Risk for general loss distributions. *Journal of Banking and Finance* **26** 1443-1471.

[29] Sturm, J. 2001. Using SeDuMi, a matlab toolbox for optimization over symmetric cones. Department of Ecnometrics, Tilburg University, The Netherlands.

[30] Topaloglou, N., H. Vladimirou, S. A. Zenios. 2002. CVaR models with selective hedging for international asset allocation. *Journal of Banking and Finance* **26** 1535-1561.