Yokohama Publishers
ISSN 1349-8169 ONLINE JOURNAL

Pacific Journal of Optimization
Vol. 15, No. 1, pp. 69 – 90, 2019

DISTRIBUTIONALLY ROBUST REWARD-RISK RATIO PROGRAMMING WITH WASSERSTEIN METRIC*

Yong Zhao, Yongchao Liu† and Xinming Yang

Abstract: It is well-known that reward-risk ratio (RR) is a very important stock market definition. In order to capture the situation that the investor does not have complete information on the distribution of the underlying uncertainty, people extend RR model to distributionally robust reward-risk ratio (DRR) model. In this paper, we study the DRR problem where the ambiguity on the distributions is defined through Wassertein metric. Under some moderate conditions, we show that for a fixed ratio, the DRR problem has the tractable reformulation, which means that we may solve the problem by bisection method. Specifically, we analyze the DRR problems for Sortino-Satchel ratio, Stable Tail Adjusted Return ratio and Omega ratio.

Key words: *distributionally robust optimization, reward-risk ratio, Wasserstein metric*

Mathematics Subject Classification: *90C15, 90C34, 90C47*

1 Introduction

One of the most challenging issues in decision analysis is to find an optimal decision under uncertainty. The solvability of a decision problem and the quality of an optimal decision rely heavily on the information regarding the underlying uncertainty. Suppose the decision maker does not have any information other than the range of the values of the uncertain, and then it might be a reasonable option to choose an optimal decision on the basis of the extreme values of the uncertainty in order to mitigate the risks. This kind of decision making framework is known as robust optimization (RO). RO is based on the pessimistic view by treating all realizations of the uncertainty equally. Over the past two decades, RO problems have been well studied in theory, algorithm and application; see the monograph by Ben-Tal et al. [1] and the survey by Bertsimas et al. [4] for recent development. If the decision maker has complete information on the distribution of the uncertainties, the problem falls into the form of stochastic programming (SP) which usually requires the knowledge of the probability distributions of the uncertainty. SP is similar in style to RO but takes advantage of the fact that probability distributions governing the data are known or can be estimated. This history of SP can be traced back to the middle of the last century and stochastic programming theory offers a variety of models to address the presence of random data in optimization problems such as chance constrained models, two- and multi-stage models,

*This work was supported in part by the NSFC Grant #11571056, #11431004, and the Science and Technology Research Program of Chongqing Municipal Education Commission (Grant No. KJQN201800744).
†Corresponding author

models involving risk measures. We refer interested readers to monographs [21,27] for recent development.

On one hand, one of the criticisms of SP is that in many applications the basic assumption of knowing, or even accurately estimating, the probability distribution of the uncertain data is unrealistic. On the other hand, the worst-case approach of RO could be too conservative. Moreover, the historical data of the uncertainty usually include some endogenous probability information such as mean or covariance. Therefore, an alternative approach, which is less pessimistic w.r.t RO and less optimistic w.r.t SP, is distributionally robust optimization (DRO). DRO problems consider the case that the decision maker is able to construct an ambiguity set of the distributions from historical data, computer simulations or subjective judgements, which contains the true distribution with a certain confidence level. Consequently, the decision maker may hedge against the risk of incomplete information by considering the worst-case distribution in the ambiguity set. DRO has found many applications in operations research, finance and management sciences and has been well investigated through a number of research works by Žáčková [36], Dupačová [7], Shapiro and Ahmed [26]. Over the past a few years, it has gained substantial popularity through further contributions by Delage and Ye [6], Hu and Hong [14], Goldfarb and Iyengar [11], Pflug et al. [20,22], Popescu [23], Wiesemann et al. [31–33], to name a few.

Our paper is related to the works applying DRO to evaluating the reward-risk ratios. Kapsos et al. [17] first propose a so-called distributionally robust Omega ratio model where an investor does not have complete information on the distribution of the underlying uncertainty in portfolio optimization and consequently a robust action is taken against the risk arising from ambiguity of the true distribution. They consider a situation where each distribution in the ambiguity set may be explicitly represented either through a mixture of some known distributions, or a perturbation from a nominal discrete distribution. Tong and Wu [30] investigate DRR optimization problem with composite mixture distributions and they transfer the DRR optimization problem to a convex optimization problem through the dual theorem. Liu et al. [18] propose a DRR model where the ambiguity set is constructed through moment conditions. They first utilize Lagrangian dualization to reformulate the DRR optimization problem as a semi-infinite programming problem and further approximate the semi-infinite constraints with Entropic risk measures. Then the implicit Dinkelbach method is used to solve the approximation problem.

Following the works mentioned above, we would like to study the DRR problem but with the ambiguity set defined through a metric in probability space. Suppose that there are historical data which may help the decision maker to construct an estimation or approximation of the true distribution of the uncertainty. Accordingly it is reasonable to define the ambiguity set as the ball centered at the estimation with proper radius to address the issue of incomplete information such as insufficient number of sample. Different from the works on DRR [17,18,30] where the authors define the ambiguity sets through moment conditions or mixture methods, cannot be more informative than just a set no matter how many data we have, distance type ambiguity set may converge to the true distribution with increasing samples and decreasing diameter of the ball. Motivated by recent works [8,10,13,37] where the authors study DRO problems with ambiguity sets based on the Wasserstein metric, we study DRR optimization problems with the ambiguity set defined through Wasserstein metric. Compared to the work of Esfahani and Kuhn [8], we extend the DRO problem with Wasserstein ball uncertainty of the distributions [8] to the DRR problem, and a key difference is that their model requires the constructed portfolio to attain a fixed, pre-specified reward-risk ratio value, while our formulation generates the portfolio with the largest worst-case ratio value (i.e., β is a decision variable) attainable. Indeed, Ji and Lejeune [15] seem to be the first to consider the DRR problem with Wasserstein ball type ambiguity set. They focus on the case that the uncertain parameter vector ξ has finitely many realizations, and they reformulated the DRR problem as tractable conic programs by leveraging ideas from robust optimization. In our paper, we consider the situations that ξ can have a continuum of realizations, and the reference distribution is discrete. Compared to the results in [15], our results are obtained by using the Wassertein metric to construct the ambiguity set for general distribution, but not restricting on discrete distribution. Therefore, the results, to some extent, extend the corresponding results in [15]. It is then predictable that a lot of difficulties associated with the continuous random variables should appear. Based on some new results in [8, 10], we successfully reformulate the robust ratio problem as a tractable convex problem.

The rest of the paper is organized as follows. In Sections 2 and 3, we focus on the distributionally robust fractional programming (DRFP) problem. Particularly, we introduce the general DRFP problem and the Wasserstein metric in Section 2. Then we provide the reformulation of DRFP model where the ambiguity set is defined through Wasserstein metric in Section 3. In Section 4, we apply the theoretical results in Section 3 to DRR problems, specifically for Sortino-Satchel ratio, Stable Tail Adjusted Return ratio and Omega ratio. By virtue of the structures of the Sortino-Satchel ratio model and Stable Tail Adjusted Return ratio, the bisection algorithm is designed to solve in Section 5. We conduct numerical results to test the applicability of the proposed reformulation and algorithmic framework in Section 6. We end the paper with a conclusion in Section 7.

Throughout the paper, we use the following notations. For vectors $a, b \in \mathbb{R}^n$, the inner product of a, b is denoted by $\langle a, b\rangle := a^\top b$. Denote by $[N] := \{1, 2, \cdots, N\}$. Given a norm $\|\cdot\|$ on $\mathbb{R}^n$, the dual norm is defined as $\|z\|_* := \sup_{\|x\|\leq 1}\langle z, x\rangle$. Let $f : \mathbb{R}^n \to \mathbb{R}$ be a lower semicontinuous convex function, the conjugate of f is defined as $f^*(z) := \sup_{x\in\mathbb{R}^n}\langle z, x\rangle - f(x)$. For a set $\Xi \subset \mathbb{R}^k$, the indicator function $\mathbb{1}_\Xi$ is defined through $\mathbb{1}_\Xi = 1$ if $\xi \in \Xi$; $= 1$ otherwise. The indicator function on the set A is defined as $\delta_A(x) = 0$ if $x \in A$; $\delta_A(x) = \infty$ otherwise.

2 Distributionally robust fractional programming and Wasserstein metric

2.1 Distributionally robust fractional programming

Consider the following fractional programming problem:

$$\max_{x\in X} \quad \frac{\mathbb{E}_P[f(x,\xi)]}{\mathbb{E}_P[g(x,\xi)]}, \tag{2.1}$$

where x is a decision vector, X is a nonempty convex compact subset of $\mathbb{R}^n$, $f : \mathbb{R}^n\times\mathbb{R}^k \to \mathbb{R}$ and $g : \mathbb{R}^n \times \mathbb{R}^k \to \mathbb{R}$ are continuous functions, ξ is a random variable on probability space $(\Xi, \mathbb{F}, P)$ with closed convex set $\Xi \subset \mathbb{R}^k$, $\mathbb{E}_P[\cdot]$ denotes the expected value w.r.t. the probability distribution of ξ. The fractional programs (2.1) have many applications in economics, and management science [9]. In this paper, we mainly focus on the application in portfolio management, where the numerator measures the expected return while the denominator measures the risk of the portfolio. The ratio model (2.1) can be regarded as finding the maximum of expected return per unit risk.

In practice, the information regarding the true distribution of random variable ξ may not be completely accessible. However, it might be possible to construct a family of distributions based on empirical data or subjective judgements which contain the true distribution with a high confidence level. Suppose that there are N independent and identically distributed historical samples $\xi_1, \cdots, \xi_N$. We may base on the empirical probability P_N to construct an ambiguity set:

$$\mathcal{P} := \{P \in \mathscr{P} : d(P, P_N) \leq \theta\},$$

where $d(\cdot, \cdot)$ denotes a predefined distance on the probability space. This leads to the following distributionally robust fractional programming:

$$\text{DRFP} \qquad \max_{x \in X} \inf_{P \in \mathcal{P}} \frac{\mathbb{E}_P[f(x,\xi)]}{\mathbb{E}_P[g(x,\xi)]}. \tag{2.2}$$

In this formulation, robustness is in the sense that given the set of probability measures $\mathcal{P}$, an optimal solution is sought against the worst probability measure which is used to compute the expected value of the objective function. Notice that problem (2.2) might not be well-defined since the denominator may turn into zero. Therefore, we always assume that for every $x \in X$ and every $\xi \in \Xi$, $g(x, \xi) > 0$ throughout this paper.

Recall that a popular method for solving distributionally robust optimization problem is to reformulate the inner maximization problem as a semi-infinite programming problem and further as a tractable convex problem through dual method. We follow the mainstream approaches in the literature to handle the DRR problem. Since the objective function (2.2) is nonlinear w.r.t. the operation of mathematical expectation, it might be very difficult to derive a dual formulation of the robust optimization, especially when the ambiguity set is defined through distance metric. Therefore we consider an equivalent maximization problem with robust constraints:

$$\begin{array}{cl} \sup_{(x,\beta) \in X \times \mathbb{R}} & \beta \\ \text{s.t.} & \inf_{P \in \mathcal{P}} \mathbb{E}_P[f(x,\xi) - \beta g(x,\xi)] \geq 0. \end{array} \tag{2.3}$$

As presented in [18, Proposition 2.1], problems (2.2) and (2.3) are equivalent in sense of the same optimal values and optimal solutions as long as both have finite optimal values and optimal solutions. Thus we may crack DRFP (2.2) by solving the problem (2.3).

2.2 Wasserstein metric

The Wasserstein metric is defined as a distance function between two probability distributions on a given support space Ξ, which is defined on the space $\mathscr{P}(\Xi)$ of all probability distributions P supported on Ξ with $\mathbb{E}_P[\|\xi\|] = \int_\Xi \|\xi\| P(d\xi) < +\infty$.

Definition 2.1. Let $\mathcal{L}$ be the space of all Lipschitz continuous functions $f : \Xi \to \mathbb{R}$ with Lipschitz constant no larger than 1. Then, the Wasserstein metric $d_{\mathrm{W}} : \mathscr{P}(\Xi) \times \mathscr{P}(\Xi) \to \mathbb{R}$ is defined as

$$d_{\mathrm{W}}(P, Q) = \sup_{f \in \mathcal{L}} \Big(\int_\Xi f(\xi) P(d\xi) - \int_\Xi f(\xi) Q(d\xi) \Big), \quad \forall P, Q \in \mathscr{P}(\Xi).$$

The Wasserstein distance exhibits the defining properties of a probability distance metric. That is, $d_{\mathrm{W}}(P, Q) = 0$ if and only if $P = Q$, $d_{\mathrm{W}}(P, Q) = d_{\mathrm{W}}(Q, P)$ and $d_{\mathrm{W}}(P, Q) \leq d_{\mathrm{W}}(P, \hat{P}) + d_{\mathrm{W}}(\hat{P}, Q)$ for any probability distribution $\hat{P}$.

By the Kantorovich-Rubinstein theorem [16], the Wasserstein metric is equivalent to the Kantorovich metric. Then for any $P, Q \in \mathscr{P}(\Xi)$, we have

$$d_{\mathrm{W}}(P, Q) = \inf \left\{ \int_{\Xi^2} \|\xi_1 - \xi_2\| \pi(d\xi_1, d\xi_2) \right\}, \tag{2.4}$$

where π is a joint distribution of ξ_1 and ξ_2 with marginal P and Q, respectively, and the 'inf' is taken over all joint distributions π. Definition 2.1 and the problem (2.4) provide two equivalent characterizations of the Wasserstein metric, which constitute a primal-dual pair of infinite-dimensional linear programs. The decision variable π appearing in the dual linear program can be viewed as a transportation plan for moving a mass distribution P to another one Q. Thus, the Wasserstein distance between P and Q represents the minimum transportation cost.

Having specified the Wasserstein metric, we can now define an ambiguity ball around the nominal distribution. Different decision makers may have different reference distributions, but the most popular way is to set the empirical distribution as the center of the ambiguity set. Suppose there are N independent and identically distributed samples for an unknown true distribution. We may construct an empirical distribution, denoted by P_N, and then define

$$\mathcal{P} := \{P \in \mathscr{P}(\Xi) : d_{\mathrm{w}}(P, P_N) \leq \theta\}, \tag{2.5}$$

where $\mathscr{P}(\Xi)$ denotes the set of all probability with support set contained in Ξ and θ is the ambiguity parameter. Obviously, the parameter θ decides the size of the ambiguity set and the forthcoming question is how to choose the size parameter θ.

If there exists an exponent $\alpha > 1$ such that $\mathbb{E}_P[\exp(\|\xi\|^\alpha)] < \infty$, Zhao and Guan [37, Proposition 1] have shown that for a general k-dimension (e.g., $k > 2$) supporting space Ξ,

$$P(d_{\mathrm{w}}(P, P_N) \leq \theta) \geq 1 - C(\exp(-cN\theta^k)\mathbb{1}_{\{\theta \leq 1\}} + \exp(-cN\theta^\alpha)\mathbb{1}_{\{\theta > 1\}}), \tag{2.6}$$

where N is the number of historical data, and C and c are positive constant numbers. (2.6) provides finite sample guarantee property and asymptotic guarantee property. This means that we may adjust the radius θ of the Wasserstein ball to ensure the ambiguity set contains the true distribution P with a given probability threshold. Moreover, (2.6) implies that the ambiguity set converges to the true distribution P as the sample size N goes to infinity. See [8, 15] for similar statistical evidence.

3 Reformulation

The DRO problem with the ambiguity set defined through Wasserstein metric has been well studied in the past decade. Pflug and Wozabal [22] study the distributionally robust coherent risk measures where the ambiguity set is defined by Wasserstein metric under the assumption that the probability space has a finite support. Wozabal [34, 35] conducts a series of DRO problems which studies the Wasserstein distance metric and presents portfolio optimization illustrations. Esfahani and Kuhn [8], as well as Zhao and Guan [37], study data-driven distributionally robust problems with the ambiguity set defined through Wasserstein metric. In particular, they analyze the conditions which ensure the tractability of the DRO problem. Ji and Lejeune [15] study the DRR problem with Wasserstein ball type ambiguity set. Particularly, they consider the ambiguity set is defined as follows:

$$\mathcal{P}^N := \{P \in \mathscr{P}(\Xi^N) : d_{\mathrm{w}}(P, P_N) \leq \theta\},$$

where P_N is empirical probability with support set Ξ^N and $\mathscr{P}(\Xi^N)$ denotes the set of all probability with support set is contained in Ξ^N. Based on the recent results on RO by Ben-Tal et al. [2], Ji and Lejeune [15] develop a tractable reformulation and associated algorithmic framework to efficiently solve the DRR problem.

We inherit the investigation of distributionally robust reward-risk ratio problems. Different from Ji and Lejeune's work [15], we remove the limitation that the probability in the ambiguity set has the same discrete support with the nominal. Specifically, we study the DRR problem with ambiguity set (2.5). Obviously, $\mathcal{P}^N \subset \mathcal{P}$ as $\mathscr{P}(\Xi^N) \subset \mathscr{P}(\Xi)$.

In what follows, we analyze the conditions which ensure that we can solve the DRFP (2.2) efficiently. For convenience of exposition, we rewrite (2.3) as a minimization problem:

$$\begin{array}{ll} \inf & -\beta \\ \text{s.t.} & \sup_{P\in\mathcal{P}} \mathbb{E}_P[-f(x,\xi)+\beta g(x,\xi)] \le 0, \\ & x\in X, \beta\in\mathbb{R}. \end{array} \tag{3.1}$$

Obviously, the problem (3.1) is a nonconvex semi-infinite problem as there is a bilinear like function $\beta g(x,\xi)$ in the constraints. In general, people cannot expect to solve a nonconvex semi-infinite problem easily. Fortunately, we may take advantage of the special structure of problem (3.1) to design an efficient algorithm. Note that the objective function $-\beta$ is monotonic, and thus we will use the bisection method to solve (3.1), that is, fix β first and then increase or decrease β based on the feasibility of β. Indeed, the idea of employing the bisection method is inspired by existing literatures. In particular, Kapsos et al. [17] update β by a fixed step and then check the feasibility; Liu et al. [18] use the bisection method to solve the DRR problem with moment type ambiguity set; Ji and Lejeune [15] on the other hand, design a bisection method for the DRR problem with the Wasserstein ball restricted on a discrete probability space. We will present the details of the bisection method in Section 5, and right here we first provide some results for checking feasibility.

The following results show that, under some moderate conditions, the constraint of problem (3.1) can be reformulated into a finite dimensional constraint space, and the resulting set of inequalities that provides an equivalent reformulation of the feasible set of (3.1).

Theorem 3.1. *Suppose that for any fixed $x\in X$, $f(x,\cdot)$ is proper and convex and $-g(x,\cdot)$ is proper and convex in ξ. Then the constraint of problem* (3.1) *can be reformulated as:*

$$\begin{array}{ll} \lambda\theta + \frac{1}{N}\sum_{i=1}^N s_i \le 0, & \\ (f)^*(x, z_i-u_i-v_i) + (-\beta g)^*(x,u_i) + \sigma_\Xi(v_i) - \langle z_i,\xi_i\rangle \le s_i, & \forall i\in[N] \\ \|z_i\|_* \le \lambda, & \forall i\in[N]. \end{array} \tag{3.2}$$

Proof. The proof is similar to [8, Theorem 4.2]. Here we give the details of the proof for completeness.

By the definition of the Wasserstein metric, we have

$$\begin{aligned} \sup_{P\in\mathcal{P}} \mathbb{E}_P[-f(x,\xi)+\beta g(x,\xi)] &= \sup_{P\in\mathscr{P}(\Xi)} \left\{\mathbb{E}_P[-f(x,\xi)+\beta g(x,\xi)] : d_{\mathrm{W}}(P,P_N)\le\theta\right\} \qquad (3.3)\\ &= \sup \int_\Xi (-f(x,\xi)+\beta g(x,\xi))P(d\xi) \\ &\quad \text{s.t. } \pi\in\mathscr{P}(\Xi\times\Xi), P\in\mathscr{P}(\Xi), \\ &\qquad \int_{\Xi^2} \|\xi - z\|\pi(d\xi,dz)\le\theta, \\ &\qquad \pi \ \text{ is a joint distribution of } \xi \text{ and } z \\ &\qquad \text{with marginals } P \text{ and } P_N, \text{ respectively.} \end{aligned}$$

By the law of total probability, $\pi = \frac{1}{N}\sum_{i=1}^N Q_i$, where Q_i represents the distribution of ξ

conditional on $z = \xi_i$. Therefore, we can rewrite (3.3) as follows:

$$\begin{array}{ll}\sup_{Q_i \in \mathscr{P}(\Xi)} \frac{1}{N}\sum_{i=1}^N \int_\Xi \big(-f(x,\xi) + \beta g(x,\xi)\big) Q_i(d\xi) \\ \text{s.t.} \quad \frac{1}{N}\sum_{i=1}^N \int_\Xi \|\xi - \xi_i\| Q_i(d\xi) \le \theta. \end{array} \tag{3.4}$$

Consider its Lagrangian dual problem that can be rewrite as follows:

$$L(\lambda) = \sup_{Q_i \in \mathscr{P}(\Xi)} \frac{1}{N}\sum_{i=1}^N \int_\Xi \big(-f(x,\xi) + \beta g(x,\xi)\big) Q_i(d\xi) + \lambda(\theta - \frac{1}{N}\sum_{i=1}^N \int_\Xi \|\xi - \xi_i\| Q_i(d\xi)),$$

where $\lambda \ge 0$ is the dual variable of constraint condition of (3.4). Obviously, the dual problem is

$$\inf_{\lambda \ge 0} L(\lambda).$$

Now, we show that the strong duality result holds. When $\theta > 0$, by virtue of a well-known strong duality result for moment problems [25, Proposition 3.4], there is no duality gap. Hence, we obtain

$$\sup_{P \in \mathcal{P}} \mathbb{E}_P[-f(x,\xi) + \beta g(x,\xi)] = \inf_{\lambda \ge 0} L(\lambda) \tag{3.5}$$

$$= \inf_{\lambda \ge 0} \lambda\theta + \frac{1}{N}\sum_{i=1}^N \sup_{\xi \in \Xi} \big(-f(x,\xi) + \beta g(x,\xi) - \lambda\|\xi - \xi_i\|\big), \tag{3.6}$$

where the equality (3.6) follows from the fact that $\mathscr{P}(\Xi)$ contains all the Dirac distributions supported on Ξ.

One can show that the equality (3.5) continues to hold even for $\theta = 0$. In fact, in which case the Wasserstein ambiguity set reduces to the singleton $\{P_N\}$, $\sup_{P \in \mathcal{P}} \mathbb{E}_P[-f(x,\xi) + \beta g(x,\xi)]$ reduces to the sample average $\frac{1}{N}\sum_{i=1}^N \big(-f(x,\xi_i) + \beta g(x,\xi_i)\big)$. For $\theta = 0$, the variable λ in (3.6) can be increased indefinitely. Since for any given x, β, $-f(x,\xi) + \beta g(x,\xi) - \lambda\|\xi - \xi_i\|$ is concave w.r.t. ξ, we can show that (3.6) converges to the sample average $\frac{1}{N}\sum_{i=1}^N \big(-f(x,\xi_i) + \beta g(x,\xi_i)\big)$ as λ tends to infinity. That is, (3.5) holds for $\theta = 0$.

By introducing auxiliary variables $s_i, i \in [N]$, we can reformulate (3.6) as

$$\begin{array}{rl}\inf_{\lambda, s_i} & \lambda\theta + \frac{1}{N}\sum_{i=1}^N s_i \\ \text{s.t.} & \sup_{\xi \in \Xi} \big(-f(x,\xi) + \beta g(x,\xi) - \lambda\|\xi - \xi_i\|\big) \le s_i, \quad \forall i \in [N] \\ & \lambda \ge 0. \end{array} \tag{3.7}$$

By virtue of the definition of the dual norm, (3.7) can be rewritten as follows:

$$\begin{array}{rl}\inf_{\lambda, s_i} & \lambda\theta + \frac{1}{N}\sum_{i=1}^N s_i \\ \text{s.t.} & \sup_{\xi \in \Xi} \big(-f(x,\xi) + \beta g(x,\xi) - \max_{\|z_i\|_* \le \lambda} \langle z_i, \xi - \xi_i\rangle\big) \le s_i, \quad \forall i \in [N] \\ & \lambda \ge 0, \end{array} \tag{3.8}$$

or equivalently,

$$\begin{array}{rl}\inf_{\lambda, s_i} & \lambda\theta + \frac{1}{N}\sum_{i=1}^N s_i \\ \text{s.t.} & \sup_{\xi \in \Xi} \min_{\|z_i\|_* \le \lambda} \big(-f(x,\xi) + \beta g(x,\xi) - \langle z_i, \xi - \xi_i\rangle\big) \le s_i, \quad \forall i \in [N] \\ & \lambda \ge 0. \end{array} \tag{3.9}$$

By the assumptions, f and g are continuous functions, and for any given x, β, $-f(x,\xi)+\beta g(x,\xi)-\langle z_i,\xi-\xi_i\rangle$ is concave w.r.t. ξ and convex w.r.t. z_i. Since for any finite λ, the set $\{z_i\in\mathbb{R}^n:\|z_i\|_*\le\lambda\}$ is compact, due to the classical minimax theorem [3, Proposition 5.5.4], (3.9) can be rewritten as follows:

$$\begin{array}{rll}\inf_{\lambda,s_i} & \lambda\theta+\frac{1}{N}\sum_{i=1}^N s_i & \\ \text{s.t.} & \min_{\|z_i\|_*\le\lambda}\sup_{\xi\in\Xi}\big(-f(x,\xi)+\beta g(x,\xi)-\langle z_i,\xi-\xi_i\rangle\big)\le s_i, & \forall i\in[N] \\ & \lambda\ge 0. & \end{array}\tag{3.10}$$

or equivalently,

$$\begin{array}{rll}\inf_{\lambda,s_i,z_i} & \lambda\theta+\frac{1}{N}\sum_{i=1}^N s_i & \\ \text{s.t.} & \sup_{\xi\in\Xi}\big(-f(x,\xi)+\beta g(x,\xi)-\langle z_i,\xi-\xi_i\rangle\big)\le s_i, & \forall i\in[N] \\ & \|z_i\|_*\le\lambda, & \forall i\in[N]. \end{array}\tag{3.11}$$

It follows from the definition of conjugacy and the substitution of z_i with $-z_i$, (3.11) is equivalent to the following problem

$$\begin{array}{rll}\inf_{\lambda,s_i,z_i} & \lambda\theta+\frac{1}{N}\sum_{i=1}^N s_i & \\ \text{s.t.} & (f-\beta g+\delta_\Xi)^*(x,z_i)-\langle z_i,\xi_i\rangle\le s_i, & \forall i\in[N] \\ & \|z_i\|_*\le\lambda, & \forall i\in[N]. \end{array}\tag{3.12}$$

Since for any given x, β, the function $f(x,\xi)-\beta g(x,\xi)+\delta_\Xi(\xi)$ is proper, convex and lower semicontinuous w.r.t. ξ. Thus,

$$\begin{aligned}(f-\beta g+\delta_\Xi)^*(x,z_i)&=\mathrm{cl}\big[\inf_{u_i,v_i}\big((f)^*(x,z_i-u_i-v_i)+(-\beta g)^*(x,u_i)+(\delta_\Xi)^*(v_i)\big)\big]\\&=\mathrm{cl}\big[\inf_{u_i,v_i}\big((f)^*(x,z_i-u_i-v_i)+(-\beta g)^*(x,u_i)+\sigma_\Xi(v_i)\big)\big].\end{aligned}$$

As $\mathrm{cl}[h(x)]\le 0$ if and only if $h(x)\le 0$ for any function h. Therefore, from (3.12), we can conclude that (3.2) holds. □

Next, we study the cases that conjugate dual functions involved in problem (3.2) have a closed form.

Theorem 3.2. *Assume*

(a) $\Xi:=\{\xi\in\mathbb{R}^n:C\xi\le d\}$ *where C is a matrix and d is a vector of appropriate dimensions;*

(b) $f(x,\xi)=\max_{1\le k\le K}\{\langle a_k(x),\xi\rangle+b_k(x)\}$;

(c) $g(x,\xi)=\max_{1\le j\le J}\{\langle c_j(x),\xi\rangle+d_j(x)\}$.

Then the constraint of problem (3.1) *can be reformulated as:*

$$\begin{array}{ll}\lambda\theta+\frac{1}{N}\sum_{i=1}^N s_i\le 0, & \\ \langle\eta_{ij},-b(x)-A(x)\xi_i\rangle+\beta(d_j(x)+\langle c_j(x),\xi_i\rangle)+\langle\gamma_{ij},d-C\xi_i\rangle\le s_i, & \forall i\in[N],j\in[J]\\ \|A(x)^\top\eta_{ij}-\beta c_j(x)+C^{\mathrm{T}}\gamma_{ij}\|_*\le\lambda, & \forall i\in[N],j\in[J]\\ \langle\eta_{ij},e\rangle=1, & \forall i\in[N],j\in[J]\\ \gamma_{ij}\ge 0,\eta_{ij}\ge 0, & \forall i\in[N],j\in[J]\end{array}\tag{3.13}$$

where $A(x)$ is a matrix with rows $a_k(x)^{\mathrm{T}},1\le k\le K$, $b(x)$ is the column vector with entries $b_k(x),1\le k\le K$, e is the vector of with each component being 1 and N is the number of samples.

Proof. It follows from the expressions of f and g and the proof of Theorem 3.1 that (3.9) can be reformulated as follows:

$$\begin{array}{rl} \inf_{\lambda,s_i} & \lambda\theta + \frac{1}{N}\sum_{i=1}^N s_i \\ \text{s.t.} & \sup_{\xi\in\Xi}\min_{\|z_{ij}\|_*\leq\lambda}\big(-f(x,\xi)+\beta g_j(x,\xi)-\langle z_{ij},\xi-\xi_i\rangle\big)\leq s_i, \quad \forall i\in[N], j\in[J] \\ & \lambda\geq 0, \end{array} \tag{3.14}$$

where $g_j(x,\xi)=\langle c_j(x),\xi\rangle+d_j(x)$. Since $f(x,\xi)=\max_{1\leq k\leq K}\{\langle a_k(x),\xi\rangle+b_k(x)\}$, then for any given $x\in X$, $f(x,\xi)$ is convex w.r.t. ξ. Then, it follows from the proof of Theorem 3.1 that (3.14) can be re-expressed as

$$\begin{array}{rll} \inf_{\lambda,s_i,z_{ij},u_{ij},v_{ij}} & \lambda\theta + \frac{1}{N}\sum_{i=1}^N s_i & \\ \text{s.t.} & (f)^*(x,z_{ij}-u_{ij}-v_{ij})+(-\beta g_j)^*(x,u_{ij}) & \\ & \quad +\sigma_\Xi(v_{ij})-\langle z_{ij},\xi_i\rangle\leq s_i, & \forall i\in[N], j\in[J] \\ & \|z_{ij}\|_*\leq\lambda, & \forall i\in[N], j\in[J]. \end{array} \tag{3.15}$$

By the definition of the conjugacy operator, we have

$$(f)^*(x,z)=\sup_\xi\{\langle z,\xi\rangle-\max_{1\leq k\leq K}\{\langle a_k(x),\xi\rangle+b_k(x)\}\}=\left\{\begin{array}{l} \inf_{\eta\geq 0}\ \langle\eta,-b(x)\rangle \\ \text{s.t. } A(x)^{\mathrm{T}}\eta=z \\ \qquad \langle\eta,e\rangle=1 \end{array}\right.$$

$$(-\beta g_j)^*(x,u)=\sup_\xi\{\langle u,\xi\rangle+\beta(\langle c_j(x),\xi\rangle+d_j(x))\}=\left\{\begin{array}{ll} \beta d_j(x) & \text{if} \quad u=-\beta c_j(x), \\ \infty & \text{otherwise.} \end{array}\right.$$

Since

$$\begin{array}{rl} \sigma_\Xi(v)=\sup_\xi & \langle v,\xi\rangle \\ \text{s.t.} & C\xi\leq d, \end{array}$$

then it follows from the strong duality of linear programming and the non-empty of the uncertainty set, we have

$$\begin{array}{rl} \sigma_\Xi(v)=\inf_{\gamma\geq 0} & \langle\gamma,d\rangle \\ \text{s.t.} & C^{\mathrm{T}}\gamma=v. \end{array} \tag{3.16}$$

Substituting the above expressions into (3.15), the result holds for $\theta>0$.

When $\theta=0$, the optimal value of (3.13) reduces to the expectation of $-f(x,\xi)+\beta g(x,\xi)$ under the empirical distribution. In fact, for $\theta=0$, the variable λ can be set to any positive value. Since all samples must belong to the uncertainty set, i.e., $d-C\xi_i\geq 0, i=1,\cdots,N$, then it is optimal to set $\gamma_{ij}=0$. This implies that

$$\begin{aligned} s_i &= \min_{\eta_{ij}\geq 0}\big\{\langle\eta_{ij},-b-A(x)\xi_i:\langle\eta_{ij},e\rangle=1\rangle\big\}+\beta\max_{1\leq j\leq J}\{\langle c_j(x),\xi_i\rangle+d_j(x)\} \\ &= \min_{1\leq k\leq K}\{-\langle a_k(x),\xi_i\rangle-b_k(x)\}+\beta\max_{1\leq j\leq J}\{\langle c_j(x),\xi_i\rangle+d_j(x)\}=-f(x,\xi_i)+\beta g(x,\xi_i) \end{aligned}$$

at optimality. Therefore, $\frac{1}{N}\sum_{i=1}^N s_i$ represents the sample average of $-f(x,\xi)+\beta g(x,\xi)$. □

If the condition (b) in Theorem 3.2 is replaced by

(b') $f(x,\xi)=\min_{1\leq k\leq K}\{\langle a_k(x),\xi\rangle+b_k(x)\}$,

the constraint of problem (3.1) is equivalent to a system of nonlinear inequalities.

Theorem 3.3. *Assume the conditions (a) and (c) of Theorem 3.2 and (b') hold. Then the constraint of problem* (3.1) *can be reformulated as:*

$$\begin{array}{ll} \lambda\theta + \frac{1}{N}\sum_{i=1}^N s_i \le 0, & \\ \big(\langle -a_k(x), \xi_i\rangle - b_k(x)\big) + \beta\big(\langle c_j(x), \xi_i\rangle + d_j(x)\big) & \\ \qquad + \langle \gamma_{ijk}, d - C\xi_i\rangle \le s_i, & \forall i \in [N], j\in[J], k\in[K] \\ \|\beta c_j(x) - a_k(x) - C^{\mathrm{T}}\gamma_{ijk}\|_* \le \lambda, \gamma_{ijk} \ge 0, & \forall i \in [N], j\in[J], k\in[K]. \end{array} \tag{3.17}$$

Proof. It follows from the expressions of f and g and the proof of Theorem 3.1 that (3.9) can be reformulated as follows:

$$\begin{array}{ll} \inf_{\lambda, s_i} & \lambda\theta + \frac{1}{N}\sum_{i=1}^N s_i \\ \text{s.t.} & \sup_{\xi\in\Xi}\min_{\|z_{ij}\|_*\le\lambda}\big(- f(x,\xi) + \beta g_j(x,\xi) - \langle z_{ij}, \xi - \xi_i\rangle\big) \le s_i, \quad \forall i\in[N], j\in[J] \\ & \lambda \ge 0, \end{array} \tag{3.18}$$

where $g_j(x,\xi) = \langle c_j(x), \xi\rangle + d_j(x)$. Since $f(x,\xi) = \min_{1\le k\le K}\{\langle a_k(x),\xi\rangle + b_k(x)\}$, then $-f(x,\xi) = \max_{1\le k\le K}\{-\langle a_k(x),\xi\rangle - b_k(x)\}$. Thus, (3.18) can be rewritten as follows:

$$\begin{array}{ll} \inf_{\lambda, s_i} & \lambda\theta + \frac{1}{N}\sum_{i=1}^N s_i \\ \text{s.t.} & \sup_{\xi\in\Xi}\min_{\|z_{ijk}\|_*\le\lambda} -f_k(x,\xi) + \beta g_j(x,\xi) \\ & \qquad -\langle z_{ijk}, \xi - \xi_i\rangle \le s_i, \forall i\in[N], j\in[J], k\in[K] \\ & \lambda \ge 0, \end{array} \tag{3.19}$$

where $f_k(x,\xi) = \langle a_k(x), \xi\rangle + b_k(x)$. Then, it follows from the proof of Theorem 3.1, the result holds for $\theta > 0$.

When $\theta = 0$, the optimal value of (3.17) reduces to the expectation of $-f(x,\xi)+\beta g(x,\xi)$ under the empirical distribution. In fact, for $\theta = 0$, the variable λ can be set to any positive value. Since all samples must belong to the uncertainty set, i.e., $d - C\xi_i \ge 0, i = 1,\cdots,N$, then it is optimal to set $\gamma_{ij} = 0$. This implies that

$$s_i = \max_{1\le k\le K}\{-\langle a_k(x),\xi_i\rangle - b_k(x)\} + \beta\max_{1\le j\le J}\{\langle c_j(x),\xi_i\rangle + d_j(x)\} = -f(x,\xi_i) + \beta g(x,\xi_i)$$

at optimality. Therefore, $\frac{1}{N}\sum_{i=1}^N s_i$ represents the sample average of $-f(x,\xi)+\beta g(x,\xi)$. □

When $f(x,\cdot)$ is an affine function:

(b") $f(x,\xi) = \langle a(x), \xi\rangle + b(x)$,

we have a more concise reformulation of the constraint of (3.1).

Corollary 3.4. *Assume the conditions (a) and (c) of Theorem 3.2 and (b") hold. Then the constraint of problem* (3.1) *can be reformulated as:*

$$\begin{array}{ll} \lambda\theta + \frac{1}{N}\sum_{i=1}^N s_i \le 0, & \\ \langle -a(x), \xi_i\rangle - b(x) + \beta(\langle c_j(x), \xi_i\rangle + d_j(x)) & \\ \qquad + \langle \gamma_{ij}, d - C\xi_i\rangle \le s_i, & \forall i\in[N], j\in[J] \\ \|a(x) + C^{\mathrm{T}}\gamma_{ij} - \beta c_j(x)\|_* \le \lambda, & \forall i\in[N], j\in[J] \\ \gamma_{ij} \ge 0, & \forall i\in[N], j\in[J]. \end{array} \tag{3.20}$$

Remark 3.5. We should keep in mind that the reformulations for the constraint of problem (3.1) are nonconvex. If β is fixed to some arbitrary positive value and $a(x), b(x), a_k(x), b_k(x), c_j(x), d_j(x), k \in [K], j \in [J]$ depend linearly on x, the reformulations (3.20) and (3.17) reduce to convex inequalities. Furthermore, if the Wasserstein metric is defined in terms of the 1-norm or the ∞-norm, the reformulations (3.20) and (3.17) reduce to linear inequalities. In contrast, the reformulation (3.13) turn out to be in non-convex when the matrix $A(x)$ or the vector $b(x)$ depends on x.

4 Reward-Risk Ratio

Since the pioneering work by Markowitz on mean-variance portfolio selection [19], the return-risk analysis framework has been widely used in financial portfolio management. Two criteria essentially underly the portfolio selection approach: the expected return and the risk. One portfolio is preferred to the other if it encompasses higher expected return and lower risk. To overcome the difficulties associated with choosing such variables, Sharpe [28] proposes a ratio optimization which is known as Sharpe ratio. Since the publication of the Sharpe ratio [28], some new performance measures such as STARR ratio, Minimax measure, Sortino ratio, Farinelli-Tibiletti ratio and most recently Rachev ratio and Generalized Rachev ratio have been proposed, for an empirical comparison, see Biglova et al. [5,24] and the references therein.

In this section, we employ the results in Section 3 to investigate the distributionally robust reward-risk ratios. In this setting, the numerator measures the expected return while the denominator measures the risk. We assume that the investor has an amount of money of value 1 and can invest his money into n assets. To simplify the discussion, we ignore the transaction fee, and therefore the total value of portfolio is $\langle x, \xi\rangle$, where $\xi = [\xi_1, \cdots, \xi_n]^{\mathrm{T}}$ denotes the random returns vector of n assets, and $x_1, \cdots, x_n$ the fractions of the initial capital invested in n assets.

4.1 Sortino-Satchel Ratio

The Sortino ratio [29] is a modification of the Sharpe ratio by penalizing only those returns falling below a user-specified target or required rate of return, while the Sharpe ratio penalizes both upside and downside volatility equally. We consider that the numerator measures the expected return while the denominator measures the expected shortfall of return below the benchmark. The former is regarded as a reward and the latter first order lower partial moment as a risk, that is

$$f(x,\xi) = \langle x, \xi\rangle, \quad g(x,\xi) = \max\{\nu - \langle x, \xi\rangle, 0\},$$

where ν is a target return value set by the decision maker.

Corollary 4.1. *Assume that the condition (a) of Theorem 3.2 holds. Consider the reward function $f(x,\xi) = \langle x, \xi\rangle$ and the risk function $g(x,\xi) = \max\{\nu - \langle x, \xi\rangle, 0\}$. Then the constraint of problem* (3.1) *can be reformulated as:*

$$\begin{array}{ll}
\lambda\theta + \frac{1}{N}\sum_{i=1}^N s_i \le 0, & \\
\beta\nu - (1+\beta)\langle x, \xi_i\rangle + \langle \gamma_{i1}, d - C\xi_i\rangle \le s_i, & \forall i \in [N] \\
-\langle x, \xi_i\rangle + \langle \gamma_{i2}, d - C\xi_i\rangle \le s_i, & \forall i \in [N] \\
\|(1+\beta)x + C^{\mathrm{T}}\gamma_{i1}\|_* \le \lambda, & \forall i \in [N] \\
\|x + C^{\mathrm{T}}\gamma_{i2}\|_* \le \lambda, & \forall i \in [N] \\
\gamma_{ij} \ge 0, & \forall i \in [N], j \in [2].
\end{array} \tag{4.1}$$

Proof. By the definition of the conjugacy operator, we have

$$(f)^*(x,z) = \sup_{\xi}\{\langle z,\xi\rangle - \langle x,\xi\rangle\} = \begin{cases} 0 & \text{if} \quad z = x, \\ \infty & \text{otherwise}, \end{cases}$$

$$(-\beta g_1)^*(x,u) = \sup_{\xi}\{\langle u,\xi\rangle + \beta(\langle -x,\xi\rangle + \nu)\} = \begin{cases} \beta\nu & \text{if} \quad u = \beta x, \\ \infty & \text{otherwise}, \end{cases}$$

$$(-\beta g_2)^*(x,u) = \sup_{\xi}\langle u,\xi\rangle = \begin{cases} 0 & \text{if} \quad u = 0, \\ \infty & \text{otherwise}, \end{cases}$$

where $g_1(x,\xi) = \langle -x,\xi\rangle + \nu, g_2(x,\xi) = 0$. From Theorem 3.2, we have

$$\sigma_\Xi(v) = \begin{array}{ll} \inf_{\gamma\geq 0} & \langle\gamma, d\rangle \\ \text{s.t.} & C^{\mathrm{T}}\gamma = v. \end{array}$$

Substituting the above expressions into (3.20), we have the result. □

4.2 Stable Tail Adjusted Return Ratio (STARR)

STARR is a variation of the Sortino-Satchel ratio where the expected one-sided deviation is replaced by a coherent risk measure, which can measure Skewness and fat-tails. Here, we employ the well used coherent risk measure, Conditional Value-at-Risk (CVaR). CVaR, sometimes called expected shortfall has received a great deal of attention as a measure of risk. In a financial context, it has a number of advantages over the commonly used Value-at-Risk (VaR) and has been proposed as the primary tool for banking capital regulation in the draft Basel III standard.

Suppose that $\langle x,\xi\rangle$ has a support contained in $[0,\tau_0]$, and then

$$\mathrm{CVaR}_\alpha(\langle x,\xi\rangle) = \inf_{\tau\in[0,\tau_0]} \tau + \frac{1}{\alpha}\mathbb{E}_P[\max\{\langle x,\xi\rangle - \tau, 0\}],$$

where $\alpha \in (0,1)$ is the confidence level. In fact, if $\tau \leq 0$, we have

$$\tau + \frac{1}{\alpha}\mathbb{E}_P[\max\{\langle x,\xi\rangle - \tau, 0\}] = (1-\frac{1}{\alpha})\tau + \frac{1}{\alpha}\mathbb{E}_P[\langle x,\xi\rangle],$$

which is increasing as $\tau \to -\infty$ since $1 - \frac{1}{\alpha} < 0$. If $\tau \geq \tau_0$, we have $\tau + \frac{1}{\alpha}\mathbb{E}_P[\max\{\langle x,\xi\rangle - \tau, 0\}] = \tau$, and since τ is increasing as $\tau \to \infty$. Because $\tau + \frac{1}{\alpha}\mathbb{E}_P[\max\{\langle x,\xi\rangle - \tau, 0\}]$ is convex in τ, then the minimizer of the function lies in $[0,\tau_0]$.

Corollary 4.2. *Suppose that $\langle x,\xi\rangle$ has a support contained in $[0,\tau_0]$ and the condition (a) of Theorem 3.2 holds. Consider $f(x,\xi) = \langle x,\xi\rangle$, $\mathbb{E}_P[g(x,\xi)] = \inf_\tau \tau + \frac{1}{\alpha}\mathbb{E}_P[\max\{-\langle x,\xi\rangle - \tau, 0\}]$. Then the constraint of problem* (3.1) *can be reformulated as:*

$$\begin{array}{ll} \lambda\theta + \frac{1}{N}\sum_{i=1}^N s_i \leq 0, & \\ \beta(1-\frac{1}{\alpha})\tau + (-\frac{\beta}{\alpha} - 1)\langle x,\xi_i\rangle + \langle\gamma_{i1}, d - C\xi_i\rangle \leq s_i, & \forall i \in [N] \\ \beta\tau - \langle x,\xi_i\rangle + \langle\gamma_{i2}, d - C\xi_i\rangle \leq s_i, & \forall i \in [N] \\ \|x + C^{\mathrm{T}}\gamma_{i1} + \frac{\beta}{\alpha}x\|_* \leq \lambda, & \forall i \in [N] \\ \|x + C^{\mathrm{T}}\gamma_{i2}\|_* \leq \lambda, & \forall i \in [N] \\ \gamma_{ij} \geq 0, \tau \in [-\tau_0, 0], & \forall i \in [N], j \in [2]. \end{array} \tag{4.2}$$

Proof. Similar to the proof of Corollary 3.4, we have

$$(f)^*(x,z) = \begin{cases} 0 & \text{if} \quad z = x, \\ \infty & \text{otherwise} \end{cases}$$

and

$$\sigma_\Xi(v) = \inf_{\gamma \geq 0} \quad \langle \gamma, d \rangle \quad \text{s.t.} \quad C^{\mathrm{T}}\gamma = v.$$

By the expression of $\mathbb{E}_P[g]$, we have

$$\mathbb{E}_P[g(x,\xi)] = \inf_\tau \mathbb{E}_P[\max\{\frac{1}{\alpha}\langle -x, \xi\rangle + (1 - \frac{1}{\alpha})\tau, \tau\}].$$

Let $a_1(x) = -\frac{1}{\alpha}x, b_1 = 1 - \frac{1}{\alpha}, a_2(x) = 0$ and $b_2 = 1$. Then

$$\mathbb{E}_P[g(x,\xi)] = \inf_{\tau \in [-\tau_0, 0]} \mathbb{E}_P[\max_{1 \leq j \leq 2} \langle a_j(x), \xi\rangle + b_j\tau].$$

Subsequently, by [25, Proposition 3.4], we have

$$\begin{aligned} &\inf_{\tau \in [-\tau_0, 0]} \sup_{P \in \mathcal{P}} \mathbb{E}_P[-f(x,\xi) + \beta(\max_{1 \leq j \leq 2} \langle a_j(x), \xi\rangle + b_j\tau)] \\ &= \sup_{P \in \mathcal{P}} \inf_{\tau \in [-\tau_0, 0]} \mathbb{E}_P[-f(x,\xi) + \beta(\max_{1 \leq j \leq 2} \langle a_j(x), \xi\rangle + b_j\tau)]. \end{aligned}$$

Therefore, from Corollary 3.4, we have the result. □

4.3 Omega Ratio

The Omega ratio is another alternative for the widely used Sharpe ratio, which is defined as the probability weighted ratio of gains versus losses for some threshold return target. The Omega ratio is upside/downside ratio performance measures, in which reward is the upside deviation and risk is taken as the downside deviation. Kapsos et al. [17] first study the robust Omega ratio where each distribution in the ambiguity set can be explicitly represented either through a mixture of some known distributions, or a perturbation from a nominal discrete distribution. We focus on the case that the ambiguity set is defined as Wasserstein ball. Here, we set

$$f(x,\xi) = \max\{\langle x, \xi\rangle - \nu, 0\}, \quad g(x,\xi) = \max\{\nu - \langle x, \xi\rangle, 0\},$$

where ν is a threshold that partitions the returns to desirable (gain) and undesirable (loss). The choice of the value 'ν' is left to the investor and can be set for example to be equal to the risk-free rate or zero.

Corollary 4.3. *Assume that the condition (a) of Theorem 3.2 holds. Consider the reward function $f(x,\xi) = \max\{\langle x, \xi\rangle - \nu, 0\}$, the risk function $g(x,\xi) = \max\{\nu - \langle x, \xi\rangle, 0\}$. Then the constraint of problem* (3.1) *can be reformulated as:*

$$\begin{array}{ll} \lambda\theta + \frac{1}{N}\sum_{i=1}^N s_i \leq 0, & \\ \eta_{i1}^1(\nu - \langle x, \xi_i\rangle) + \beta(\nu + \langle -x, \xi_i\rangle) + \langle \gamma_{i1}, d - C\xi_i\rangle \leq s_i, & \forall i \in [N] \\ \eta_{i2}^1(\nu - \langle x, \xi_i\rangle) + \langle \gamma_{i2}, d - C\xi_i\rangle \leq s_i, & \forall i \in [N] \\ \|\eta_{i1}^1 x + \beta x + C^{\mathrm{T}}\gamma_{i1}\|_* \leq \lambda, & \forall i \in [N] \\ \|\eta_{i2}^1 x + C^{\mathrm{T}}\gamma_{i2}\|_* \leq \lambda, & \forall i \in [N] \\ \langle \eta_{ij}, e\rangle = 1, & \forall i \in [N], j \in [2] \\ \gamma_{ij} \geq 0, \eta_{ij} \geq 0, & \forall i \in [N], j \in [2] \end{array} \tag{4.3}$$

where η_{ij}^1 represents the first component of η_{ij}, $j \in [2]$.

Proof. By some simple calculation, we have

$$(f)^*(x,z) = \sup_{\xi}\{\langle z,\xi\rangle - \max\{\langle x,\xi\rangle - \nu, 0\}\} = \left\{ \begin{array}{rl} \inf_{\eta\geq 0} & \langle \eta, -b(x)\rangle \\ \text{s.t.} & A(x)^{\mathrm{T}}\eta = z \\ & \langle \eta, e\rangle = 1 \end{array} \right.$$

$$(-\beta g_1)^*(x,u) = \sup_{\xi}\{\langle u,\xi\rangle + \beta(\langle -x,\xi\rangle + \nu)\} = \left\{ \begin{array}{ll} \beta\nu & \text{if} \quad u = \beta x, \\ \infty & \quad \text{otherwise,} \end{array} \right.$$

$$(-\beta g_2)^*(x,u) = \sup_{\xi}\{\langle u,\xi\rangle\} = \left\{ \begin{array}{ll} 0 & \text{if} \quad u = 0, \\ \infty & \text{otherwise,} \end{array} \right.$$

where $A(x) = \begin{bmatrix} x^{\mathrm{T}} \\ \mathbf{0}^{\mathrm{T}} \end{bmatrix}$, $b(x) = \begin{bmatrix} -\nu \\ 0 \end{bmatrix}$ and $\mathbf{0}$ denotes a vector with each component being 0. Substituting the above expressions into (3.13), we obtain the result. □

Remark 4.4. (i) In the recent work [15], Ji and Lejeune derive robust counterparts for reward-risk ratio expectation constraint by using the framework proposed by Ben-Tal et al. [2]. The method is designed for ambiguous probability distributions with fixed atoms, and encompasses two separate reformulation phases involving the derivation of: 1) the support function of the ambiguity set and 2) the concave conjugate of the reward-risk function. Moreover, the discrete uncertain parameter vector ξ_i is included in the equivalent reformulation of the feasible set of the ambiguous reward-risk ratio constraints. In our paper, we obtain robust counterparts for reward-risk ratio expectation constraint by leveraging tools from robust optimization. Different from Ji and Lejeune's work, the uncertain parameter vector ξ_i does not include in the equivalent reformulation of the feasible set of the ambiguous reward-risk ratio constraints except for independent and identically distributed historical samples.

(ii) Ji and Lejeune [15] study the DRR problem with Wasserstein ball type ambiguity set and show that for many kinds of ratio such as Sharpe, Sortino-Satchel and Omega ratios, the step of checking feasibility can be completed through solving a convex problem. However, problem (4.3) shows that for distributionally robust Omega ratio we cannot check the feasibility through solving a convex optimization problem. The reason is that we release the condition in [15] by considering the Wasserstein ball on the general probability space $\mathscr{P}(\Xi)$ rather than on the discrete probability space $\mathscr{P}(\Xi^N)$.

5 Bisection Algorithmic Methods

For a minimization problem with a monotonous objective value $-\beta$ in $[a,b]$, the bisection algorithm divides the incumbent interval $[a,b]$ in two equal parts. Then, compute the value of the objective function at the midpoint $\beta = (a+b)/2$ of the interval $[a,b]$. If the problem is feasible at β, there is potentially a better optimal value smaller than $-\beta$. The search is continued on $[\beta,b]$, otherwise search focuses on the interval $[a,\beta]$. The process continues until the interval is sufficiently small.

In what follows, we mainly analyze the bisection algorithm for DRR problem with Sortino-Satchel ratio and DRR problem with STARR can be analyzed in a similar way.

Assume that the investor has observed N historical samples from the true distribution P and the support set of distribution P is contained in $\Xi = \{\xi \in \mathbb{R}^n : C\xi \leq d\}$, where C is a matrix and d is a vector of appropriate dimensions. Consider the case that the reward function $f(x,\xi) = \langle x, \xi \rangle$ and the risk function $g(x,\xi) = \max\{\nu - \langle x, \xi \rangle, 0\}$. By Corollary 4.1, we have that the problem (3.1) can be reformulated as follows:

$$\begin{array}{lll} \inf & -\beta & \\ \text{s.t.} & \lambda\theta + \frac{1}{N}\sum_{i=1}^{N} s_i \leq 0 & \\ & \beta\nu - (1+\beta)\langle x, \xi_i \rangle + \langle \gamma_{i1}, d - C\xi_i \rangle \leq s_i, & \forall i \in [N] \\ & -\langle x, \xi_i \rangle + \langle \gamma_{i2}, d - C\xi_i \rangle \leq s_i, & \forall i \in [N] \\ & \|x + C^{\mathrm{T}}\gamma_{i1} + \beta x\|_* \leq \lambda, & \forall i \in [N] \\ & \|x + C^{\mathrm{T}}\gamma_{i2}\|_* \leq \lambda, & \forall i \in [N] \\ & \gamma_{ij} \geq 0, x \in X, & \forall i \in [N], j \in [2]. \end{array} \tag{5.1}$$

For simplicity, we rewrite (5.1) into a more concise form:

$$\begin{array}{lll} (\mathrm{D}) & \inf & -\beta \\ & \text{s.t.} & (x, \lambda, s, \gamma_1, \gamma_2, \beta) \in H, \end{array} \tag{5.2}$$

where H denotes the feasible set of (5.1). Then for each iteration, the feasibility check is done by solving the following problem:

$$\begin{array}{lll} (\mathrm{D}_t) & \inf & -\beta \\ & \text{s.t.} & (x, \lambda, s, \gamma_1, \gamma_2, \beta) \in H, \\ & & \beta = \beta^t, \end{array} \tag{5.3}$$

where β^t is the middle point of the interval $[\beta_L^t, \beta_U^t]$ which is determined by step $t-1$.

Given the tolerance level ε, the bisection algorithm to solve problem (5.1) can be presented as follows

---

Algorithm 1: Pseudo-Code of Bisection Algorithm

Step 1: Determine $[\beta_L^0, \beta_U^0]$ such that $\beta^* \in [\beta_L^0, \beta_U^0]$, and set $t = 0, v^0 = -\beta_L^0$;
Step 2: Set $\beta^t = (\beta_L^t + \beta_U^t)/2$, and solve the problem (D_t);
Step 3: If (D_t) is feasibility, then set $v^t = -\beta^t$; update interval $[\beta_L^{t+1}, \beta_U^{t+1}] : \beta_L^{t+1} = \beta^t$ and $\beta_U^{t+1} = \beta_U^t$, let $t = t+1$. Otherwise, set $v^t = v^{t-1}$; update interval $[\beta_L^{t+1}, \beta_U^{t+1}] : \beta_L^{t+1} = \beta_L^t$ and $\beta_U^{t+1} = \beta^t$, let $t = t+1$;
Step 4: If $\beta_U^t - \beta_L^t \leq \varepsilon$, stop. Return v^t as the optimal value. Otherwise go to **Step 2**.

In Algorithm 1, the suitable values β_L^0 and β_U^0 are determined to ensure that the optimal solution β^* belongs to each of the successive intervals $[\beta_L^t, \beta_U^t]$. The cost on time of Algorithm 1 depends highly on the diameter of the interval $[\beta_L^0, \beta_U^0]$. In practice, people may set the interval based on their experience and/or their expectation. For example, set lower bound as unacceptable ratio and the upper bound as highest ratio we expect. We will present some numerical methods to determine lower and upper bounds of the ratio in the next subsection. Note also $\{\beta_L^t\}$ is an increasing sequence and $\{\beta_U^t\}$ is a decreasing sequence, and hence the sequence $\{-v^t\}$ is monotonically increasing. Therefore, the Algorithm 1 finds the optimal value in a finite number of iterations with a given tolerance ε.

5.1 Determine the initial interval $[\beta_L, \beta_U]$

It is well known that the efficiency of bisection method is highly dependent on the initial interval $[\beta_L, \beta_U]$. In this subsection, we present a method to set the interval $[\beta_L, \beta_U]$. Let us focus on the lower bound β_L first.

Lower bound β_L: For any fixed $x \in X$, denote

$$\beta_L(x) := \inf_{P \in \mathcal{P}} \frac{\mathbb{E}_P[f(x,\xi)]}{\mathbb{E}_P[g(x,\xi)]}. \tag{5.4}$$

Obviously, the optimal ratio β^* such that

$$\beta^* = \max_{x\in X} \inf_{P\in\mathcal{P}} \frac{\mathbb{E}_P[f(x,\xi)]}{\mathbb{E}_P[g(x,\xi)]} \geq \beta_L(x), \quad \forall x \in X.$$

By the equivalent reformulation (2.3), we may estimate a lower bounded $\beta_L(x)$ of β^* by solving

$$\begin{array}{lll} (\mathrm{P}_1) & \max & \beta \\ & \text{s.t.} & \mathbb{E}_P[-f(x,\xi)+\beta g(x,\xi)] \leq 0, \forall P \in \mathcal{P}, \\ & & \beta \in \mathbb{R}_+. \end{array} \tag{5.5}$$

By using Lagrange dual method, problem (5.5) can be reformulated as a convex problem:

$$\begin{array}{llll} (\mathrm{P}_1') & \max & \beta & \\ & \text{s.t.} & \lambda\theta + \frac{1}{N}\sum_{i=1}^N s_i \leq 0 & \\ & & \beta\nu - (1+\beta)\langle x,\xi_i\rangle + \langle \gamma_{i1}, d - C\xi_i\rangle \leq s_i, & \forall i \in [N] \\ & & -\langle x,\xi_i\rangle + \langle \gamma_{i2}, d - C\xi_i\rangle \leq s_i, & \forall i \in [N] \\ & & \|x + C^{\mathrm{T}}\gamma_{i1} + \beta x\|_* \leq \lambda, & \forall i \in [N] \\ & & \|x + C^{\mathrm{T}}\gamma_{i2}\|_* \leq \lambda, & \forall i \in [N] \\ & & \gamma_{ij} \geq 0, & \forall i \in [N], j \in [2]. \end{array} \tag{5.6}$$

Again, the investor may setting the lower bound $\beta_L(x)$ by choosing different x, for example, set x as the portfolio choice of last investment process or the equally weighted strategy.

Upper Bound β_U:

Denote

$$\beta_U^N := \max_{x\in X} \frac{\mathbb{E}_{P_N}[f(x,\xi)]}{\mathbb{E}_{P_N}[g(x,\xi)]}, \tag{5.7}$$

where P_N is the empirical distribution. Obviously, the optimal ratio β^* is given by

$$\beta^* = \max_{x\in X} \inf_{P\in\mathcal{P}} \frac{\mathbb{E}_P[f(x,\xi)]}{\mathbb{E}_P[g(x,\xi)]} \leq \beta_U^N.$$

As P_N is a discrete distribution, problem (5.7) is equivalent to

$$\begin{array}{llll} (\mathrm{P}_2) & \max_{x,u} & \frac{\frac{1}{N}\sum_{i=1}^N \langle x,\xi_i\rangle}{\frac{1}{N}\sum_{i=1}^N u_i} & \\ & \text{s.t.} & u_i \geq \nu - \langle x,\xi_i\rangle, & \forall i \in [N] \\ & & \sum_{i=1}^m x_i = 1, & \\ & & u_i \geq 0, & \forall i \in [N], \end{array} \tag{5.8}$$

where the auxiliary variable $u = (u_1, \cdots, u_N)^{\mathrm{T}}$ is introduced to deal with the max-function. As problem (P_2) is a linear fractional program, it can be further reformulated as an equivalent linear programming problem:

$$\begin{array}{llll} (\mathrm{P}_2') & \max_{z,v,w} & \frac{1}{N}\sum_{i=1}^N \langle z,\xi_i\rangle & \\ & \text{s.t.} & \frac{1}{N}\sum_{i=1}^N v_i = 1 & \\ & & v_i \geq w\nu - \langle z,\xi_i\rangle, & \forall i \in [N] \\ & & \sum_{i=1}^m z_i = w, & \\ & & v_i \geq 0, & \forall i \in [N]. \end{array} \tag{5.9}$$

Obviously, the optimal solution to the problem (P_2) is $x^* = z^*/\sum_{i=1}^N z_i^*$, where z^* is the optimal solution of (P_2'). Therefore,

$$\beta_U = \frac{\mathbb{E}_{P_N}[f(x^*,\xi)]}{\mathbb{E}_{P_N}[g(x^*,\xi)]}.$$

In [15], Ji and Lejeune present an effective method to derive a lower bound and a upper bound for β. They determine a upper bound via solving two subproblems: maximizing the robust reward measure without restriction on the risk and minimizing the robust risk measure without restriction on the reward. For the lower bound, they provide an estimation at each iteration t by virtue of the bisection algorithm with iterative interval compaction method. See Section 5 of [15] for details.

6 Numerical Results

In this section, we consider the DRFP (2.2) and implement Algorithm 1 by evaluating its performance under various scenarios. The experiments are based on the application of the model in portfolio optimization where certain assets are invested in stock markets. We collect the following ten stocks: BAE Systems PLC, BG Group PLC, BHP Billiton PLC, BP PLC, BT Group PLC, Babcock International Group PLC, Barclays PLC, British American Tobacco PLC, British Land Co PLC, British Sky Broadcasting Group (http://finance.google.com) (from 5th Feb 2013 to 15th Nov 2013) with a total of 200 datasets. We consider a simple case when there is a fixed fund normalized to one for investment at the beginning of day 101. We assign the portfolio by using 100 most recent historical data up to the date. Assuming that the stocks are sold at the end of the day and the total accumulated fund is invested in the following day. This is not necessarily practical but we do so by updating the portfolio on daily basis in order to evaluate the performance of the portfolio against the change of data.

For numerical experiments, we consider the case that the reward function $f(x, \xi) = \langle x, \xi \rangle$, the risk function $g(x, \xi) = \max\{1 - \langle x, \xi \rangle, 0\}$ and the support set Ξ is

$$\Xi := \{\xi \in \mathbb{R}^n : C\xi \leq d\} = \{\xi \in \mathbb{R}^{10} : [I, -I]^{\mathrm{T}}\xi \leq (d_{\max}, -d_{\min})^{\mathrm{T}}\},$$

where I denotes the identity matrix and $d_{\max}$, $d_{\min}$ denote the max and min of the 200 history data in component. By Corollary 4.1, the DRFP (2.2) can be reformulated as problem (5.1). We test the cases that the parameter θ of the ambiguity set is $1, 0.5, 0.01, 0.005$ and the $\|\cdot\|_*$ in problem (5.1) is 1-norm, 2-norm and ∞-norm respectively. We implement Algorithm 1 on MATLAB 2015a installed in a PC with Windows 10 operating system. We use CVX (version 1.22) developed by Grant and Boyd [12] to check the feasibility in Steps 3 Algorithm 1. The initial interval of the ratio β is determined by problems (5.6) and (5.9) in Section 5.1 with upper bound 463.8075 and lower bound 25.6081. For convenience in analyzing the interations, we enlarge the interval to $[0, 500]$. Moreover, the tolerance of the bisection method is 10^{-2}.

We record the numerical results in Tables 1-4 and Figures 1-7. Tables 1-4 summarize the daily returns generated by DRO ratio model with the three different norms. In the tables, the number of days when the overall portfolio return exceeds (or equals to) 1 is denoted by "Up", and the lowest, highest and average returns rate denoted by "L", "H" and "A" respectively. We can see that the DRFP (2.2) achieves comparable average daily return for any given parameter θ and norms, but the average investment strategy $(1/M)$ displays more stable performance with a narrower range between the best and worst return rate. With the big ambiguity parameter ($\theta = 1, 0.5$), Table 1 and 2 show that the return rates based on 1-norm and 2-norm are closer to each other and the three norms are stable to the change of the parameter from 0.5 to 1. On the other hand, with small ambiguity parameter ($\theta = 0.01, 0.005$), Table 3 and 4 show that return rate based on 1-norm and 2-norm are sensitive to the change of parameter θ and the strategy based ∞-norm is still stable to the change of the parameter form 0.005 to 0.01. Figures 1-4 depict the evolution of wealth over 100 trading days when managing a portfolio of ten assets on a daily basis with different ambiguity parameters. The Figures 1-4 indicate that all wealth curves have the tendency to going down at the beginning and at the end. Figures 1-2 show that all the wealth curves generated by the three norms dominate the curve of $1/M$ strategy with big ambiguity parameter ($\theta = 1, 0.5$), which means that the robust strategy return by DRFP (2.2) outperforms the average investment strategy. With the decreasing of θ, Figures 3-4 show the difference of four wealth curves is decreasing, which is consistent with the fact that $\mathcal{P} := \{P \in \mathscr{P}(\Xi) : d_{\mathrm{W}}(P, P_N) \leq \theta\}$ tends to P_N as θ tends to zero. Figures 4-7 depict the stability of the norms with respect to the ambiguity parameter θ. Just as shown in Tables 1-4, 1-norm and 2-norm are sensitive to the change of the parameter when θ is small. On the other hand, ∞-norm is more stable to the small change of the parameter θ.

Table 1: $\theta = 1$

	Up	H	L	A
1-norm	52	1.0210	0.9858	1.002
2-norm	52	1.0209	0.9862	1.002
∞-norm	55	1.0207	0.9861	1.003
$1/M$	55	1.0219	0.9872	1.001

Table 2: $\theta = 0.5$

	Up	H	L	A
1-norm	52	1.0210	0.9858	1.003
2-norm	52	1.0210	0.9860	1.002
∞-norm	55	1.0209	0.9861	1.003
$1/M$	55	1.0219	0.9872	1.001

Table 3: $\theta = 0.01$

	Down	H	L	A
1-norm	54	1.0227	0.9860	1.002
2-norm	58	1.0234	0.9814	1.003
∞-norm	58	1.0284	0.9814	1.003
$1/M$	55	1.0219	0.9872	1.001

Table 4: $\theta = 0.005$

	Down	H	L	A
1-norm	56	1.0226	0.9849	1.002
2-norm	53	1.0224	0.9857	1.002
∞-norm	58	1.0285	0.9813	1.003
$1/M$	55	1.0219	0.9872	1.001

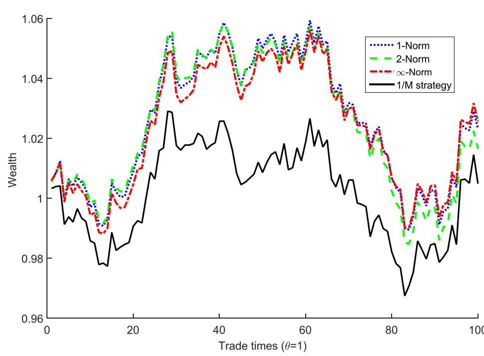

Figure 1: $\theta = 1$

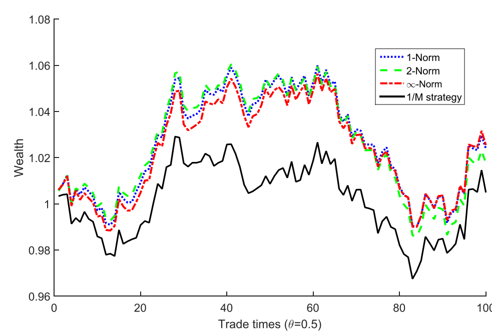

Figure 2: $\theta = 0.5$

7 Conclusions

We consider the distributionally robust reward-risk ratio (DRR) problems, where the ambiguity set is characterized by Wasserstein ball centered at an empirical distribution. The paper makes some contributions to the current research on a few aspects. First, it extends the DRR problem with moment or mixture type ambiguity sets [17, 18, 30] to distance type ambiguity set, which ensures that the ambiguity set may converge to rather than include the true distribution; second, it extends the SDP reformulation of DRO problem with Wasserstein ball uncertainty of the distributions [8] to the DRR problem; third, it extends the current research on DRR with Wasserstein type ambiguity set which restricted on a discrete

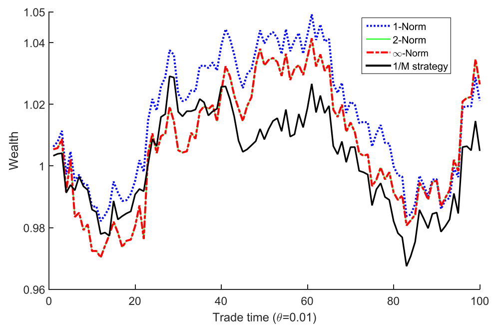

Figure 3: $\theta = 0.01$

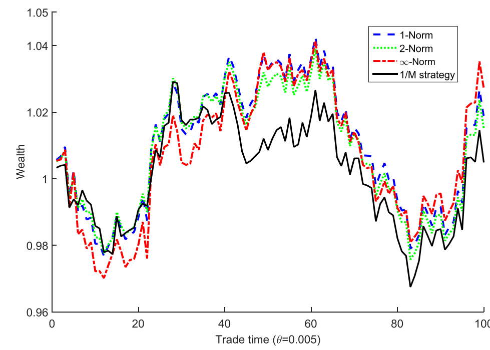

Figure 4: $\theta = 0.005$

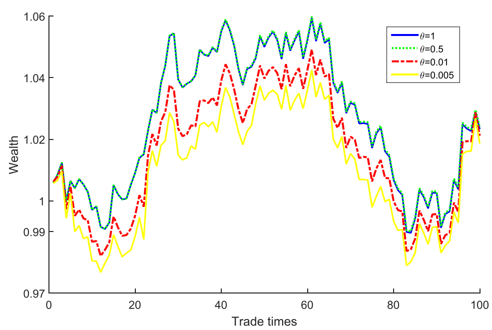

Figure 5: 1-norm

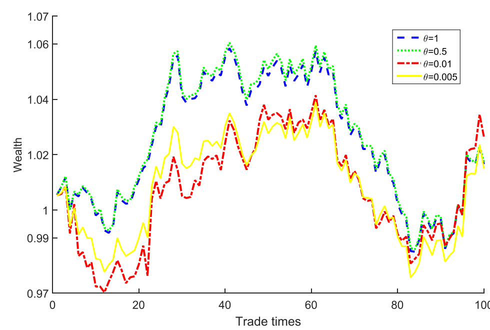

Figure 6: 2-norm

probability space to the general probability space. It might be possible to take this work further in the following direction: the SDP reformulation relies heavily on the piecewise linear structure of the involved functions, and it is very interesting to explore weaker conditions.

Acknowledgements

The authors are grateful to the two anonymous referees for their valuable comments and constructive suggestions, which helped to improve the original manuscript. They also thank Dr. Jin Zhang, Anchao Song for numerous discussions on numerical results.

References

[1] A. Ben-Tal, L. El Ghaoui and A. Nemirovski, *Robust Optimization*, Princeton University Press, NJ, 2009.

[2] A. Ben-Tal, D. den Hertog and J.P. Vial, Deriving robust counterparts of nonlinear uncertain inequalities, *Math. Program.* 149 (2015) 265–299.

[3] D.P. Bertsekas, *Convex Optimization Theory*, Athena Scientific, 2009.

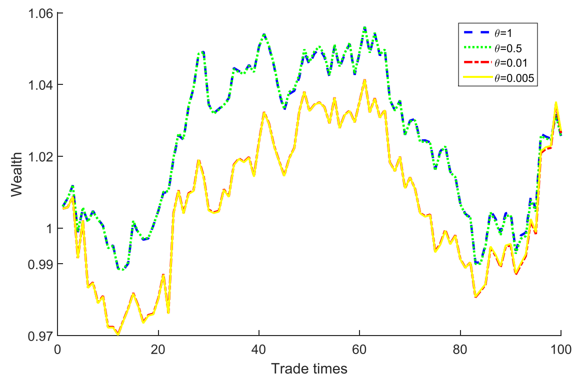

Figure 7: ∞-norm

[4] D. Bertsimas, D.B. Brown and C. Caramanis, Theory and applications of robust optimization, *SIAM Rev.* 53 (2011) 464–501.

[5] A. Biglova, S. Ortobelli, S. Rachev and S. Stoyanov, Different approaches to risk estimation in portfolio theory, *J. Portfol. Manag.* 31 (2004) 103–112.

[6] E. Delage and Y. Ye, Distributionally robust optimization under moment uncertainty with application to data-driven problems, *Oper. Res.* 58 (2010) 592–612.

[7] J. Dupačová, Uncertainties in minimax stochastic programs, *Optimization.* 60 (2011) 1235–1250.

[8] P.M. Esfahani and D. Kuhn, Data-driven distributionally robust optimization using the Wasserstein metric: performance guarantees and tractable reformulations, *Math. Program.* DOI 10.1007/s10107-017-1172-1, 2017.

[9] J.B.G. Frenk and S. Schaible, *Fractional programming*, Handbook of Generalized Convexity and Generalized Monotonicity, . Springer, Berlin, 2004, pp. 333–386.

[10] R. Gao and A. J. Kleywegt, Distributionally robust stochastic optimization with Wasserstein distance. Manuscript, 2016.

[11] D. Goldfarb and G. Iyengar, Robust portfolio selection problems, *Math. Oper. Res.* 28 (2003) 1–38.

[12] M. Grant and S. Boyd, CVX, for convex optimization, http:// www.stanford.edu/boyd/.

[13] G. Hanasusanto and D. Kuhn, Conic programming reformulations of two-stage distributionally robust linear programs over Wasserstein balls. Manuscipt, 2016.

[14] Z. Hu and J. Hong, Kullback-Leibler divergence constrained distributionally robust optimization. Manuscript, 2012.

[15] R. Ji and M.A. Lejeune, Data-driven optimization of reward-risk ratio measures. Manuscript, 2015.

[16] L.V. Kantorovich and G.S. Rubinshtein, On a space of totally additive functions, *Vestn Lening. Univ.* 13 (1958) 52–59.

[17] M. Kapsos, N. Christofides and B. Rustem, Worst-case robust Omega ratio, *Eur. J. Oper. Res.* 234 (2014) 499–507.

[18] Y. Liu, R. Meskarian and H. Xu, Distributionally robust reward-risk ratio optimization with moment constraints, *SIAM J. Optim.* 27 (2017) 957–985.

[19] H. Markowitz, Portfolio selection, *J. Financ.* 1 (1952) 77–91.

[20] G.Ch. Pflug and A. Pichler, Approximations for probability distributions and stochastic optimization problems. *Stochastic Optimization Methods in Finance and Energy*, 163, Springer, New York, 2011, pp. 343–387.

[21] G.Ch. Pflug and A. Pichler, *Multistage Stochastic Optimization*, Springer Series in Operations Research and Financial Engineering, Springer, 2014.

[22] G.Ch. Pflug and D. Wozabal, Ambiguity in portfolio selection, *Quant. Financ.* 7 (2007) 435–442.

[23] I. Popescu, Robust mean-covariance solutions for stochastic optimization, *Oper. Res.* 55 (2007) 98–112.

[24] S. Rachev, T. Jašik, A. Biglova and F. Fabozzi, *Risk and return in momentum strategies: Profitability from portfolios based on risk-adjusted stock ranking criteria*, Technical report, Department of probability and applied statistics, University of California, Santa Barbara, 2005.

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

[26] A. Shapiro and S. Ahmed, On a class of minimax stochastic programs, *SIAM J. Optim.* 14 (2004) 1237–1249.

[27] A. Shapiro, D. Dentcheva and A. Ruszczynski, *Lectures on Stochastic Programming: Modeling and Theory*, SIAM, Philadelphia, 2009.

[28] W.F. Sharpe, Mutual funds performance, *J. Busin.* 39 (1966) 119–138.

[29] F.A. Sortino and S. Satchell, *Managing Downside Risk in Financial Markets*, Butterworth-Heinemann, 2003.

[30] X. Tong and F. Wu, Robust reward-risk ratio optimization with application in allocation of generation asset, *Optimization.* 63 (2014) 1761–1779.

[31] W. Wiesemann, D. Kuhn and B. Rustem, Robust resource allocations in temporal networks, *Math. Program.* 135 (2012) 437–471.

[32] W. Wiesemann, D. Kuhn and B. Rustem, Robust markov decision processes, *Math. Oper. Res.* 38 (2013) 153–183.

[33] W. Wiesemann, D. Kuhn and M. Sim, Distributionally robust convex optimization, *Oper. Res.* 62 (2014) 1358–1376.

[34] D. Wozabal, A framework for optimization under ambiguity, *Ann. Oper. Res.* 193 (2012) 21–47.

[35] D. Wozabal, Robustifying convex risk measures for linear portfolios: A nonparametric approach, *Oper. Res.* 62 (2014) 1302–1315.

[36] J. Žáčková, On minimax solutions of stochastic linear programming problems, *Časopis pro Pěstování Matematiky.* 91 (1966) 423–430.

[37] C. Zhao and Y. Guan, Data-driven risk-averse stochastic optimization with Wasserstein metric, *Oper. Res. Lett.* Doi.org/10.1016/j.orl.2018.01.011, 2018.

Manuscript received 25 June 2017
revised 10 March 2018
accepted for publication 19 March 2018

Yong Zhao
College of Mathematics and Statistics, Chongqing JiaoTong University
Chongqing, 400074, China

E-mail address: zhaoyongty@126.com

Yongchao Liu
School of Mathematical Sciences
Dalian University of Technology, Dalian, 116024, China
E-mail address: lyc@dlut.edu.cn

Xinmin Yang
Department of Mathematics
Chongqing Normal University Chongqing, 400047, China
E-mail address: xmyang@cqnu.edu.cn