A Semi-Infinite Programming Approach for Distributionally Robust Reward-Risk Ratio Optimization with Matrix Moments Constraints[1]

Yongchao Liu
School of Mathematics, Computer Sciences and Engineering
City University London
EC1V 0HB, London, UK
Email: Yongchao.Liu@city.ac.uk
Department of Mathematics, Dalian Maritime University, Dalian 116026, China

Rudabeh Meskarian
Engineering Systems and Design
Singapore University of Technology and Design (SUTD)
Singapore 138682
Email: Rudabeh@sutd.edu.sg

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

January 26, 2015

Abstract

Reward-risk ratio optimization is an important mathematical approach in finance [54]. In this paper, we revisit the model by considering a situation where an investor does not have complete information on the distribution of the underlying uncertainty and consequently a robust action is taken against the risk arising from ambiguity of the true distribution. We propose a distributionally robust reward-risk ratio optimization model where the ambiguity set is constructed through simple inequality moment constraints and develop efficient numerical methods for solving the problem: first, we transform the robust optimization problem into a nonlinear semi-infinite programming problem through Lagrange dualization and then use the well known entropic risk measure to construct an approximation of the semi-infinite constraints, we solve the latter by an implicit Dinkelbach method (IDM). Finally, we apply the proposed robust model and numerical scheme to a portfolio optimization problem and report some preliminary numerical test results.

Key words. Reward-risk ratio, distributionally robust optimization, entropic risk measure, implicit Dinkelbach method

1 Introduction

Since the pioneering work by Markowitz on mean-variance portfolio selection [32], the return-risk analysis framework has been widely used in financial portfolio management. Two criteria essentially

[1]The research is supported by EPSRC grant EP/M003191/1.

underly the portfolio selection approach: the expected return and the risk. One portfolio is preferred to another if it encompasses higher expected return and lower risk. The reward-risk optimization has been discussed based on the various risk measures in the literature, see Stoyanov et al [54] for an excellent treatment and overview of the topic. The proposed models associated with risk and reward are dependent of personal preference and have the following common frameworks [56]: from all feasible portfolios with a given upper bound on risk, find an optimal solution of the maximum return; from all feasible portfolios with a given lower bound on return, find an optimal solution of the minimum risk; from all feasible portfolios with a given risk-aversion parameter, find an optimal solution of the maximum utility function of return and risk. The frameworks rely on the upper (lower) bound of risk (return) or the risk aversion parameters. To overcome the difficulties associated with choosing such variables, performance ratio optimization which is based on the mean-variance analysis was proposed by Sharpe [51]. In the recent progress of risk management, the performance ratio optimization models have attracted a great deal of attentions both in academics and practitioners.

Since the publication of the Sharpe ratio [51], 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. [8], Rachev et al. [42] and the references therein. The new ratios take into account of empirically observed phenomena that, the distributions of asset's returns are fat-tailed and skewed, by incorporating proper reward and risk measures.

In practice, no matter which performance ratio is used, whether the ratio can be precisely evaluated or not depends mainly on the reliability and the accuracy of prediction of the distribution of asset returns [62]. It is well established that, when historical data are used to fit into an economical model, the estimated parameters tend to be unstable. Black and Litterman [9] found that mean-variance portfolio decision is very sensitive with regard to the mean, which indicates that a small error in the estimator of this variable may significantly influence the optimal portfolio strategy. This phenomenon has been further studied by Best and Grauer [7], Broadie [11] and Chopra and Ziemba [13]. Consequently, in the last decades, there has been an increase in study of robustness and worst-case analysis of the portfolio selection problem. Ben-Tal et al. [4] proposed a robust multistage asset allocation model. Goldfarb and Iyengar [23], Halldórsson and Tütüncü [24], Lu [31] investigated the robust mean-variance portfolio selection problem. El Ghaoui et al. [21] and Natarajan et al. [33] studied the robust mean-VaR portfolio selection models. Natarajan et al. [34], Zhu and Fukushima [61], Zhu et al. [62] and Chen et al. [12] explored robust portfolio selection problems using CVaR and LPM measures. Delage and Ye [17] and Popescu [40] studied the distributioally robust portfolio where the ambiguity set defined through moment constraints.

The work by Goldfarb and Iyengar [23] apparently has received wide attention among others particularly in the research community of robust optimization. Instead of assuming precise information on the mean and the covariance matrix of asset returns in a robust mean-variance framework, they introduced some types of uncertainties, such as polytopic uncertainty, box uncertainty and ellipsoidal uncertainty, in the parameters in determining the mean and the covariance matrix, and they then transformed the problem into semidefinite programs or second-order cone programs, which can be efficiently solved by interior-point algorithms developed in recent years. As discussed in [30, 31], the "separable" uncertainty sets typically share two common properties: (1) their actual confidence level, namely, the probability of uncertain parameters falling within the uncertainty set, is unknown, and it can be much higher than the desired one; and (2) they are fully or partially box-type. The associated consequences are that the resulting robust portfolios can be too conservative, and moreover, they are usually highly non-diversified, as observed in the computational experiments conducted in [30, 31, 57].

In this paper, we focus on robust reward-risk ratio optimization. The robust optimization models do not require specification of the exact distribution of the exogenous uncertainties of the model. On the other hand, in stochastic programming, uncertainties are typically modeled as random variables with known distributions and have been used to obtain analytic solutions to important classes of problems; see for instance [45]. This is the general distinction between the approaches of robust optimization and stochastic programming toward modeling problems with uncertainties. In the framework of robust optimization, however, uncertainties are usually modeled as random variables with true distributions that are unknown to the modeler but are constrained to lie within a known support. Each approach has its advantages: if the exact distribution of uncertainties is precisely known, optimal solutions to the robust problem would be overly and unnecessarily conservative. Conversely, if the assumed distribution of uncertainties is in fact different from the actual distribution, the optimal solution using a stochastic programming approach may perform poorly.

A body which aims to bridge the gap between the conservatism of robust optimization and the specificity of stochastic programming is the minimax stochastic programming approach, where optimal decisions are sought for the worst case probability distributions within a family of possible distributions, defined by certain properties such as their support and moments. This approach was pioneered by Zackova [60] and studied in many other works (e.g., Dupacova [18], Shapiro and Kleywegt [27]). This approach has seen numerous applications, dating back to the study by Scarf [46] of an optimal single-product newsvendor problem under an unknown distribution with known mean and variance. El Ghaoui et al. [20] developed worst case Value-at-Risk bounds for a robust portfolio selection problem when only the bounds on the means and covariance matrix of the assets are known. Popescu [40] considered a distributionally robust portfolios optimization model where the mean and covariance are assumed to be known exactly. Delage and Ye [17] studied a distributionally robust optimization model where the true mean and covariance of the underlying uncertainties are subject to uncertainty, and applied the model to portfolio selection.

This paper is aimed at extending this body of work in a similar direction for a reward-risk ratio optimization problem. We seek a solution which is distributionally robust, i.e., feasible for the worst case probability distribution within a specified set of distributions. The main contributions of this paper can be summarized as follows.

- We consider a robust scheme for a reward-risk ratio optimization model, reformulate it as a mathematical program with robust inequality constraints (Proposition 2.1) and further as a semi-infinite program in the case when the ambiguity set is constructed through moments. We then apply the well known entropic approximation scheme to the semi-infinite constraints and give a quantitative stability analysis of the approximation scheme under the Slater conditions (Theorem 3.1). Contrary to the mainstream methods in the literature of robust optimization, the approach does not require any specific structure (e.g. linear, convex or polynomial) of the underlying random functions w.r.t. the random variable. An implicit Dinkelbach method which captures the specific structure of the approximated robust program is developed and convergence of the resulting algorithm is established (Theorem 5.1).

- We investigate a specific case where the ambiguity set is determined by the mean and covariance. Instead of assuming complete information of the two quantities as in [40], we take a similar approach to [17] by allowing some degree of uncertainty of them. However, instead of considering deviation of the true mean and covariance within a specified ellipsoid and positive semidefinite cone as [17], we adopt a simple approach which restricts each component of the mean and covariance to an interval with finite lower and upper bound. In doing so, we convert the distributionally robust approach to a classical moment problem and then solve the latter using the general numerical schemes proposed in this paper (Algorithm 5.1). Statistical analysis has been presented for assessing the likelihood of the true probability distribution to lie in the ambiguity set (Theorem 4.1) and the convergence of the resulting optimal value and optimal solutions (Theorem 4.2). The approach is applied to a portfolio selection problem and preliminary numerical test results shows some promising performance (Section 6).

Throughout this paper, we will use the following notation. For vectors $a, b \in \mathbb{R}^n$, $a^T b$ denotes the scalar product, $\|\cdot\|$ denotes the Euclidean norm of a vector, $\|\cdot\|_\infty$ denotes the maximum norm of continuous functions defined over compact set T, $\|\cdot\|_F$ denotes the Frobenius norm of matrix. For a set A, 'cl$A$' denotes the closure of A, $d(x, A) := \inf_{x' \in A} \|x - x'\|$ denotes the distance from a point x to the set A. For two compact sets $\mathcal{C}$ and $\mathcal{A}$,

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

denotes the deviation of $\mathcal{C}$ from $\mathcal{A}$ and $\mathbb{H}(\mathcal{C}, \mathcal{A}) := \max(\mathbb{D}(\mathcal{C}, \mathcal{A}), \mathbb{D}(\mathcal{A}, \mathcal{C}))$ denotes the Hausdroff distance between $\mathcal{C}$ and $\mathcal{D}$. Finally, for a sequence of subsets $\{A_k\}$ in a metric space, we follow the standard notation [43] by using $\limsup_{k\to\infty} A_k$ to denote its outer limit, that is,

$$\limsup_{k\to\infty} A_k = \{x : \liminf_{k\to\infty} d(x, A_k) = 0\}.$$

2 Robust reward-risk models

We consider general performance ratio optimization with one-sided risk measure, where only downward variations are penalized. Specifically, we consider the following optimization problem:

$$\max_{x\in X} \quad \frac{\mathbb{E}_P[f(x, \xi(\omega)) - Y(\xi(\omega))]}{\mathbb{E}_P[(Y(\xi(\omega)) - f(x, \xi(\omega)))_+]}, \tag{2.1}$$

where x is a decision vector, X is a nonempty convex subset of $\mathbb{R}^n$, $f : \mathbb{R}^n \times \mathbb{R}^k \to \mathbb{R}$ is a continuous return function and it is concave w.r.t. x for every fixed ξ, $\xi : \Omega \to \Xi \subset \mathbb{R}^k$ is a random variable defined on probability space $(\Omega, \mathcal{F}, P)$ with support Ξ, $Y(\xi(\omega))$ is a benchmark return, $\mathbb{E}_P[\cdot]$ denotes the expected value w.r.t. the probability distribution of $\xi(\omega)$, and $(a)_+ = \max(a, 0)$. In this setup, the numerator measures the expected excess return over the benchmark while the denominator measures the expected shortfall of return below the benchmark. The former is regarded as a reward and the latter as a risk. To ease notation, we will use ξ to denote either the random vector $\xi(\omega)$ or an element of $\mathbb{R}^k$ depending on the context.

Problem (2.1) might not be well defined since the denominator may turn into zero. In the literature of portfolio optimization, one often assumes that the risk of the active return is positive for all feasible portfolios [54]. Another difficulty is the fractional form of the objective. If $f(\cdot, \xi)$ is concave for every ξ, then the objective function is quasiconcave. In that case, one may introduce a new variable $\tau \in \mathbb{R}$ and reformulate (2.1) as follows:

$$\begin{array}{ll} \max\limits_{(x,\tau)\in X\times\mathbb{R}} & \tau \\ \text{s.t.} & \mathbb{E}_P[f(x, \xi) - Y(\xi) - \tau(Y(\xi) - f(x, \xi))_+] \geq 0, \\ & x \in X. \end{array} \tag{2.2}$$

This kind of reformulation is well known, see for instance Pinar et al. [39].

The third issue which concerns the reward-risk ratio model is the information on the underlying uncertainty and indeed this is the main focus of this paper. In practice, complete information on the true distribution of random variable ξ may not be available. However, it might be possible to construct a set of distributions based on empirical data or subjective judgements which contain or approximate the true distribution. In the literature of robust optimization, the set of distributions is also called *ambiguity set* indicating ambiguity of true distribution. Let $\mathcal{P}$ denote the ambiguity set which contains the true probability distribution. We consider a robust scheme for problem (2.1) as follows:

$$\max_{x\in X}\min_{P\in\mathcal{P}} \frac{\mathbb{E}_P[f(x,\xi)-Y(\xi)]}{\mathbb{E}_P[(Y(\xi)-f(x,\xi))_+]}. \tag{2.3}$$

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. Note that the robust problem (2.3) depends on the choice of $\mathcal{P}$ and an optimal solution of problem (2.3) provides a lower bound of optimal value of the true problem (2.1). In order for the robust formulation (2.3) to be well defined, we assume that there exists a positive number ϵ such that

$$\min_{x\in X}\min_{P\in\mathcal{P}}\mathbb{E}_P[(Y(\xi)-f(x,\xi))_+]\geq\epsilon. \tag{2.4}$$

Since the objective function (2.3) is nonlinear w.r.t. the operation of mathematical expectation, it might be very difficult to derive a dual formulation of the robust optimization problem particularly when the ambiguity set is defined through moments. To circumvent the issue, we consider an equivalent maximization problem with robust constraints:

$$\begin{array}{ll}\displaystyle\max_{(x,\tau)\in X\times\mathbb{R}} & \tau \\ \text{s.t.} & \displaystyle\min_{P\in\mathcal{P}}\mathbb{E}_P[f(x,\xi)-Y(\xi)-\tau(Y(\xi)-f(x,\xi))_+]\geq 0.\end{array} \tag{2.5}$$

Compared to (2.3), (2.5) is relatively easier to tackle as both the objective and constraint functions are linear w.r.t. $\mathbb{E}_P[\cdot]$. The following proposition confirms the equivalence of the two problems.

Throughout the paper, we write ∞ for $+\infty$ and by convention $\frac{0}{0}=\infty$.

Proposition 2.1 *Problems (2.5) and (2.3) are equivalent when both have a finite optimal value and optimal solutions.*

Proof. Let (x^*,τ^*) and $(\hat{x},\hat{\tau})$ be the pairs of the optimal solution and the optimal value of problems (2.5) and (2.3) respectively. We proceed the proof in two steps.

Step 1. $\tau^*\leq\hat{\tau}$. Since (x^*,τ^*) must be a feasible solution to (2.5),

$$\mathbb{E}_P[(f(x^*,\xi)-Y(\xi))-\tau^*(Y(\xi)-f(x^*,\xi))_+]\geq 0 \tag{2.6}$$

for all $P\in\mathcal{P}$. In what follows, we show that (2.6) implies

$$\min_{P\in\mathcal{P}}\frac{\mathbb{E}_P[f(x^*,\xi)-Y(\xi)]}{\mathbb{E}_P[(Y(\xi)-f(x^*,\xi))_+]}\geq\tau^*. \tag{2.7}$$

Let

$$\mathcal{P}_0:=\{P\in\mathcal{P}:\mathbb{E}_P[(Y(\xi)-f(x^*,\xi))_+]=0\}$$

and

$$\mathcal{P}^*:=\arg\min_{P\in\mathcal{P}}\frac{\mathbb{E}_P[f(x^*,\xi)-Y(\xi)]}{\mathbb{E}_P[(Y(\xi)-f(x^*,\xi))_+]}.$$

We consider two cases: 1. $\mathcal{P}_0 = \emptyset$; 2. $\mathcal{P}_0 \neq \emptyset$, that is, there exists $P_0 \in \mathcal{P}$ such that $\mathbb{E}_{P_0}[(Y(\xi) - f(x^*,\xi))_+] = 0$. The implication of (2.6) to (2.7) follows from a simple rearrangement in Case 1 in that $\mathbb{E}_P[(Y(\xi) - f(x^*,\xi))_+] > 0$ for all $P \in \mathcal{P}$, therefore it suffices to show it in Case 2.

Let $P_0 \in \mathcal{P}_0$. Then $\mathbb{E}_{P_0}[(Y(\xi) - f(x^*,\xi))_+] = 0$. Since $(Y(\xi) - f(x^*,\xi))_+ \geq 0$, the equality implies $(Y(\xi) - f(x^*,\xi))_+ = 0$ almost surely with respect to P_0 (a.s. P_0 for short) or equivalently $Y(\xi) \leq f(x^*,\xi)$ a.s. P_0. Subsequently, we have $\mathbb{E}_{P_0}[f(x^*,\xi) - Y(\xi)] \geq 0$ and hence

$$\frac{\mathbb{E}_{P_0}[f(x^*,\xi) - Y(\xi)]}{\mathbb{E}_{P_0}[(Y(\xi) - f(x^*,\xi))_+]} = \infty.$$

If $P_0 \in \mathcal{P}^*$, then

$$\min_{P\in\mathcal{P}} \frac{\mathbb{E}_P[f(x^*,\xi) - Y(\xi)]}{\mathbb{E}_P[(Y(\xi) - f(x^*,\xi))_+]} = \infty$$

and (2.7) holds trivially. Let us therefore assume $\mathcal{P}_0 \cap \mathcal{P}^* = \emptyset$. In that case, (2.7) is equivalent to

$$\min_{P\in\mathcal{P}\backslash\mathcal{P}_0} \frac{\mathbb{E}_P[f(x^*,\xi) - Y(\xi)]}{\mathbb{E}_P[(Y(\xi) - f(x^*,\xi))_+]} \geq \tau^*. \tag{2.8}$$

Clearly (2.8) is implied by (2.6) through a simple rearrangement because $\mathcal{P}\backslash\mathcal{P}_0 \subset \mathcal{P}$. This completes the proof for case 2.

Observe that $x^* \in X$ is a feasible solution to (2.3) whereas $\hat{x}$ is an optimal solution of (2.3). Therefore

$$\hat{\tau} \geq \min_{P\in\mathcal{P}} \frac{\mathbb{E}_P[f(x^*,\xi) - Y(\xi)]}{\mathbb{E}_P[(Y(\xi) - f(x^*,\xi))_+]} \geq \tau^*.$$

Step 2. $\hat{\tau} \leq \tau^*$. Since $\hat{x}$ is an optimal solution to (2.3),

$$\frac{\mathbb{E}_P[f(\hat{x},\xi) - Y(\xi)]}{\mathbb{E}_P[(Y(\xi) - f(\hat{x},\xi))_+]} \geq \hat{\tau}$$

for all $P \in \mathcal{P}$. Multiplying both sides of the inequality above by $\mathbb{E}_P[(Y(\xi) - f(\hat{x},\xi))_+]$ and taking minimum w.r.t. P over $\mathcal{P}$, we immediately obtain

$$\min_{P\in\mathcal{P}} \mathbb{E}_P[(f(\hat{x},\xi) - Y(\xi)) - \hat{\tau}(Y(\xi) - f(\hat{x},\xi))_+] \geq 0,$$

which means $(\hat{x}, \hat{\tau})$ is a feasible solution of problem (2.5) and hence $\hat{\tau} \leq \tau^*$. ■

Proposition 2.1 paves the way for us to develop a complete numerical treatment of robust reward-risk ratio optimization problem (2.3) via (2.5). Note that the proposition does not require condition (2.4). To ensure the optimal value and the optimal solutions of (2.5) to be bounded, we make the following assumption.

Assumption 2.1 *Assume that*

(a) X is a compact and convex set;

(b) the random variable is defined in a finite dimensional space with compact support;

(c) $Y(\cdot)$ is continuous;

(d) there exists a positive number ϵ such that (2.4) holds.

Parts (a)-(c) of the assumption are standard, see similar assumptions by Dupačová [18, 19] for distributionally robust minimax problems. Part (d) provides a sufficient condition for the well-definedness of the robust formulation (2.5).

Proposition 2.2 *Under Assumption 2.1, problem (2.5) has a finite optimal value.*

Proof. Assumption 2.1 (d) ensures that for all $P \in \mathcal{P}$

$$\begin{aligned}\left|\frac{\mathbb{E}_P[f(x,\xi)-Y(\xi)]}{\mathbb{E}_P[(Y(\xi)-f(x,\xi))_+]}\right| &\leq \frac{1}{\epsilon}\left|\mathbb{E}_P[f(x,\xi)-Y(\xi)]\right| \\ &\leq \sup_{x\in X}\sup_{P\in\mathcal{P}}\frac{1}{\epsilon}\left|\mathbb{E}_P[f(x,\xi)-Y(\xi)]\right| \\ &\leq \frac{1}{\epsilon}\sup_{x\in X}\sup_{\xi\in\Xi}\left|f(x,\xi)-Y(\xi)\right|.\end{aligned}$$

Under Assumption 2.1 (b)-(c), Ξ is compact and functions f and Y are continuous. Therefore the last term of $\sup_{x\in X}\sup_{\xi\in\Xi}|f(x,\xi)-Y(\xi)|$ is bounded and so is the optimal value of (2.3). The conclusion follows from the equivalence between (2.3) and (2.5) as we have shown in Proposition 2.1. ∎

For the convenience of exposition, let us rewrite (2.5) as a minimization problem

$$\begin{array}{ll}\min\limits_{(x,\tau)\in X\times\mathbb{R}} & -\tau \\ \text{s.t.} & \max\limits_{P\in\mathcal{P}} -\mathbb{E}_P[(f(x,\xi)-Y(\xi))-\tau(Y(\xi)-f(x,\xi))_+]\leq 0.\end{array} \tag{2.9}$$

Moreover, for the simplicity of notation, let

$$G(x,\xi,\tau) := -f(x,\xi)+Y(\xi)+\tau(Y(\xi)-f(x,\xi))_+.$$

Then we can write (2.9) in a concise form

$$\begin{array}{ll}\min\limits_{(x,\tau)\in X\times\mathbb{R}} & -\tau \\ \text{s.t.} & \max\limits_{P\in\mathcal{P}} \mathbb{E}_P[G(x,\xi,\tau)]\leq 0.\end{array} \tag{2.10}$$

Observe that problem (2.10) is nonconvex as G is not convex w.r.t. (x,τ). However, for fixed τ, G is convex. This and the specific form of the objective function allow us to give a simple characterization of optimality as stated in the proposition below.

Proposition 2.3 *Under Assumption 2.1, the following assentations hold.*

(i) For each $P \in \mathcal{P}$ and $x \in X$, $\mathbb{E}_P[G(x,\xi,\tau)]$ is strictly increasing in τ;

(ii) there exists a finite τ^* such that

$$\max_{P\in\mathcal{P}} \mathbb{E}_P[G(x,\xi,\tau^*)] \geq 0, \forall x\in X; \tag{2.11}$$

(iii) if there exists $x^* \in X$ such that the equality in (2.11) holds, then $-\tau^*$ is the optimal value and (x^*,τ^*) is the optimal solution of problem (2.10).

Proof. Part (i). Under Assumption 2.1 (d), it is easy to see that $\mathbb{E}_P[G(x,\xi,\cdot)]$ is strictly increasing in τ.

Part (ii). Since X and Ξ are compact sets by Assumption 2.1 and $-f(x,\xi)+Y(\xi)$ is continuous in (x,ξ),

$$\max_{P\in\mathcal{P}} \mathbb{E}_P[-f(x,\xi)+Y(\xi)] \leq \max_{(x,\xi)\in X\times\Xi} -f(x,\xi)+Y(\xi) < \infty.$$

On the other hand, under Assumption 2.1 (d), it follows by (2.4)

$$\max_{P\in\mathcal{P}} \mathbb{E}_P[\tau(Y(\xi)-f(x,\xi))_+] \geq \tau\epsilon.$$

Therefore, there must exists τ^* sufficiently large such that

$$\max_{P\in\mathcal{P}} \mathbb{E}_P[G(x,\xi,\tau^*)] \geq 0, \forall x\in X.$$

Part (iii). The equality of (2.11) implies feasibility of (x^*,τ^*) to problem (2.10). By Part (i) and (2.11), for any $x\in X$ and any $\delta>0$

$$\max_{P\in\mathcal{P}} \mathbb{E}_P[G(x,\xi,\tau^*+\delta)] > 0,$$

which implies that the optimal value of problem (2.10) cannot be smaller than $-\tau^*$. The conclusion follows. ■

The monotonicity of $\mathbb{E}_P[G(x,\xi,\cdot)]$ plays a key role in designing the implicit Dinkelbach method (Algorithm 5.1) and the Parts (ii)-(iii) ensure the boundedness of the optimal solution to problem (2.10) and the sufficiently condition for optimality.

3 Dual formulation and entropic approximation

Problem (2.10) is not numerically solvable as it stands unless the ambiguity is specified and the max operation in the constraint is removed. In this section, we consider a dual formulation for the problem in the case when the ambiguity set $\mathcal{P}$ is constructed through moments. Depending on the context of practical applications and availability of information on the distribution of the unknown parameters, there are various ways to define the distributional set $\mathcal{P}$, for instances Zhu and Fukushima [61] considered a mixture distribution approach for building the ambiguity set where component distributions are drawn from various sources such as historical data, market information and investor's subjective views. More recently, Zhu et al. [63] proposed a new scheme which effectively determines a set of mixture distributions by maximum likelihood approach with a given set of sampled data.

Our focus here will be on the case when the ambiguity set is constructed through moments. The underlying consideration is that given some historical data, it is often easier to estimate the moments of the random parameters than to derive their probability distributions in this setting. Moment problems have been studied by Stieltjes [53] in the nineteenth century. Schmüdgen [47], Putinar [41], and Curto and Fialkow [15] derived necessary and sufficient conditions for sequences of moments with different settings. The problem of moments is also related to optimization over polynomials (the dual theory of moment). For instance, Lasserre [28] and Parrilo [37] among others proposed relaxation hierarchies for optimization over polynomials using moment results. Bertsimas and Popescu [6] studied further the optimal inequalities given moment information. Moment problems in finance such as option pricing problems have been investigated in the literature; see [5] and references therein.

In this section, we consider the robust optimization problem (2.10) where the ambiguity set $\mathcal{P}$ is constructed through moment conditions:

$$\mathcal{P} := \left\{ P \in \mathscr{P} : \begin{array}{l} \mathbb{E}_P[\psi_i(\xi)] = 0, \ i = 1, \ldots, p \\ \mathbb{E}_P[\psi_i(\xi)] \leq 0, \ i = p+1, \ldots, q \end{array} \right\}, \tag{3.12}$$

where $\psi_i : \Xi \to \mathbb{R}$, $i = 1, \ldots, p$, are random functions, and $\mathscr{P}$ denotes the set of probability measures on probability space $(\Omega, \mathcal{F})$ of random variable ξ. Note that in this setup, we restrict $\psi(\xi)$ to be scalar functions so that we may focus on the main analysis and methodology that we want to convey in this paper. It would be challenging and practically interesting to consider the case when some ψ_i is a matrix such as correlation matrix.

For the simplicity of discussion, we make the following assumptions throughout this section.

Assumption 3.1 *Let $\psi_i(\xi)$, $i = 1, \cdots, q$, be defined as in (3.12) and Ξ be the support set of ξ. Let $\psi := (\psi_1, \cdots, \psi_q)$. The following inclusion holds:*

$$0_q \in \text{int}\{\mathbb{E}_P[\psi(\xi)] : P \in \mathscr{P}\} - \mathcal{K},$$

where 'int' denotes the interior of a set and $\mathcal{K} := 0_q \times \mathbb{R}^{q-p}_+$.

The assumption is similar to Slater constraint qualification. It means that ψ_i cannot be selected arbitrarily. It is a standard condition for deriving Lagrange dual of moment problems, see Shapiro [48]. Under Assumption 3.1, we can reformulate problem (2.9), through [48, Proposition 3.4], as the following semi-infinite programming problem

$$\begin{array}{ll} \min\limits_{x \in X, \tau \in \mathbb{R}, \lambda \in \mathbb{R}^p \times \mathbb{R}^{q-p}_+} & -\tau \\ \text{s.t.} & \sum\limits_{i=1}^{q} \lambda_i \psi_i(\xi) \geq G(x, \xi, \tau), \ \forall \xi \in \Xi. \end{array} \tag{3.13}$$

If, in addition, Assumption 2.1 holds, then the optimal value of the primal and the dual problems is finite. Again, by [48, Proposition 3.4], the set of optimal solutions is nonempty and bounded.

In what follows, we propose an approximation scheme for the semi-infinite constraints. This is much needed because the index set Ξ is the support set of random vector ξ which could be very large particularly when ξ has several components. In the mainstream robust optimization, one often assumes that Ξ takes a specific structure such as polyhedral and the underlying functions are linear w.r.t. ξ. In that way, one can derive a reformulation of the semi-infinite constraints as a finite number system of equalities and inequalities. Here we do not make any assumption as such to maximize the generality of the model.

For the convenience of exposition, let

$$R(x, \tau, \lambda, \xi) := G(x, \xi, \tau) - \sum_{i=1}^{q} \lambda_i \psi_i(\xi) \tag{3.14}$$

and

$$W := X \times \mathbb{R} \times \mathbb{R}^p \times \mathbb{R}^{q-p}_+ \text{ and } w := (x, \tau, \lambda) \in W. \tag{3.15}$$

The semi-infinite constraint in (3.13) can be written as

$$\sup_{\xi\in\Xi} R(w,\xi) \le 0. \tag{3.16}$$

Of course, writing the semi-infinite constraints as above does not bring us any numerical convenience. Our plan here is to consider an approximation of $\sup_{\xi\in\Xi} R(w,\xi)$ through entropic risk measure so that the constraint can be handled numerically more conveniently. This kind of approximation scheme was recently considered by Anderson et al. [1] using CVaR and Liu and Xu with entropic approximation [29]. The rationale behind the scheme is that if we regard $\sup_{\xi\in\Xi} R(w,\xi)$ as a extremely robust risk measure, then the CVaR of $R(w,\xi)$ would be a relaxation depending on the confidence level to be set. This kind of relaxation might not significantly affect the optimal decision through stability analysis but does allow one to take a few more samples near the extremum to calculate the CVaR (as opposed to a single sample at the extremely robust constraint) and in that way "smooth up" or "stabilize" the numerical calculation. Here we take a similar initiative but adopt entropic risk measure approximation in that the latter is a smooth function and fits to a broader class of random functions.

For a random $Z \in L^\infty$ the entropic risk measure is defined as

$$e_\gamma(Z) := \rho(Z) = \frac{1}{\gamma}\ln \mathbb{E}_P[e^{-\gamma Z}],$$

see [22] for a thorough treatment of the concept. The following lemma states how entropic risk measure approximation works for a general random function.

Lemma 3.1 *(Entropic approximation of a random function, [29, Proposition 2.1]) Let* $h : \mathbb{R}^n \times \mathbb{R}^k \to \mathbb{R}$ *be a continuous function and* X *be a subset of* $\mathbb{R}^n$*. Let* $\xi : \Omega \to \mathbb{R}^k$ *be a random variable defined on the probability space* $(\Omega, \mathcal{F}, P)$ *with support set* Ξ*. Let* $H(x)$*,* $F(\cdot)$ *and* Ω_x *denote respectively the essential supremum, the cumulative distribution function and the support set of* $-h(x,\xi)$*. Let* $\mathrm{Diam}(\Omega_x)$ *denote the diameter of the support set* Ω_x *which is the distance between* $H(x)$ *and essential infimum of* $-h(x,\xi)$*. Assume: (a)* $X \subset \mathbb{R}^n$ *is a compact set, (b) for each fixed* $x \in X$*,*

$$\inf_{\xi\in\Xi} h(x,\xi) > -\infty.$$

Then for each fixed $x \in X$*,*

$$\lim_{\gamma\to\infty} e_\gamma(h(x,\xi)) = H(x).$$

Assume in addition that (c)

$$\inf_{x\in X}\inf_{\xi\in\Xi} h(x,\xi) > -\infty,$$

and (d) for any fixed small positive number ϵ*, there exists* $\delta(\epsilon) \in (0,1)$ *such that*

$$1 - F(H(x) - \epsilon) \ge \delta(\epsilon), \quad \forall x \in X_\epsilon,$$

where $X_\epsilon := \{x \in X : \mathrm{Diam}(\Omega_x) > 2\epsilon\}$*. Then*

$$|e_\gamma(h(x,\xi)) - H(x)| < 2\epsilon + \left|\frac{1}{\gamma}\ln\delta(\epsilon)\right|.$$

Following Lemma 3.1, we may consider an approximation of the inequality system (3.16) by

$$e_\gamma(-R(w,\xi)) := \frac{1}{\gamma}\log \mathbb{E}[e^{\gamma R(w,\xi)}] \le 0. \tag{3.17}$$

Remark 3.1 It is important to distinguish the expectation $\mathbb{E}[\cdot]$ here from the expectation $\mathbb{E}_P[\cdot]$ in the preceding section. From Lemma 3.1, we can see the conclusion holds for any probability distribution of $\tilde{\xi}$ with support set Ξ. However, in problem (2.1), we don't assume any knowledge of the true distribution P except the support set Ξ. The expectation $\mathbb{E}[\cdot]$ in (3.17) should be understood as taken w.r.t. any distribution of any random variable $\tilde{\xi}$ with support set Ξ. In other words here the ξ does not have to be identical to the ξ in (2.1). We use the same letter to ease the notation. In a simple case, we may set ξ to be a random variable with uniform distribution over Ξ. Of course, the selection of ξ and its distribution will affect the rate of convergence of the entropic risk measure to its essential supremum.

With (3.17), we can construct an approximation of problem (3.13) by

$$\begin{array}{cl} \min\limits_{x\in X,\tau\in\mathbb{R},\lambda\in\mathbb{R}^p\times\mathbb{R}^{q-p}_+} & -\tau \\ \text{s.t.} & e_\gamma(-R(x,\lambda,\tau,\xi)) \leq 0. \end{array} \tag{3.18}$$

This is a one stage stochastic minimization problem with a single stochastic constraint. One can easily apply the well known sample average approximation (SAA) methods for solving the problem. The convergence results are well documented, see for instance Shairo [50], Pflug [38], Römisch [44] for the convergence of optimal values when SAA is applied to general stochastic optimization problem and Anderson et al. [1] for a similar scheme to (3.18) with CVaR approximation.

Recall that we have explained immediately after formulation (3.13) that the set of the optimal solutions to problem (3.13) is nonempty and bounded under Assumption 3.1. Thus, we may restrict the variables λ_i, $i = 1, \cdots, q$ in problem (3.18) to take finite values. Specifically, we assume that there exists a positive constant C_0 such that

$$|\lambda_i| \leq C_0, \text{for } i = 1, \cdots, q.$$

Note also that problem (3.18) is nonconvex in that R is a nonconvex function of (x, λ, τ). However, for fixed τ the function is convex w.r.t. other variables. We will come back to this in Section 5 when we discuss numerical schemes for the problem. Here, we concentrate on approximation of (3.18) to (3.13) in terms of the optimal value and the optimal solutions as γ increases. The proposition below addresses the approximation of the set of feasible solutions as γ increases and its impact on the optimal values.

Proposition 3.1 *Let $\mathcal{F}$ and $\mathcal{F}(\gamma)$ denote the feasible sets of (3.13) and (3.18) respectively, let ϑ^* and $\vartheta(\gamma)$ denote the corresponding optimal values [2]. Assume that the mathematical expectation in (3.17) is taken w.r.t. a probability distribution from the ambiguity set $\mathcal{P}$. Then the following assertions hold.*

(i) $\mathcal{F} \subset \mathcal{F}(\gamma)$ *for all* $\gamma > 0$;

(ii) $\mathcal{F}(\gamma)$ *is monotonically decreasing, that is, for* $\gamma_1 < \gamma_2$, $\mathcal{F}(\gamma_2) \subset \mathcal{F}(\gamma_1)$;

(iii) $\vartheta(\gamma)$ *is monotonically increasing and* $\vartheta(\gamma) \leq \vartheta^*$;

(iv) if, in addition, Assumptions 2.1 and 3.1 hold, then both ϑ^* and $\vartheta(\gamma)$ are finite, and the set of optimal solutions to (3.18) is nonempty and bounded.

[2] By writing $\mathcal{F}(\gamma)$ and $\vartheta(\gamma)$, we mean that we are looking into the feasible set and the optimal value as a function (multi-valued for the former) of γ and investigate specifically how these quantities change w.r.t. variation of γ.

Proof. Part (i). Compare constraints of problems (3.13) and (3.18), the only difference is the second constraint. Since $e_\gamma(-R(w,\xi)) \le \sup_{\xi\in\Xi} R(w,\xi)$, then $\mathcal{F} \subset \mathcal{F}(\gamma)$. Part (ii). The monotonicity follows from that of the constraint function $e_\gamma(-R(w,\xi))$ increases w.r.t. γ. Part (iii) follows from Part (i).

Part (iv). Let us show boundedness of the optimal values first. Under Assumption 2.1, we have shown in Proposition 2.2 that ϑ^* is finite. In what follows, we demonstrate the boundedness of $\vartheta(\gamma)$. By Jensen's inequality

$$\mathbb{E}\left[e^{\gamma R(x,\lambda,\tau,\xi)}\right] \ge e^{\gamma \mathbb{E}[R(x,\lambda,\tau,\xi)]}.$$

Through the definition of R in (3.14), we have

$$\mathbb{E}[R(x,\lambda,\tau,\xi)] = \mathbb{E}[G(x,\xi,\tau)] - \sum_{i=1}^{q} \lambda_i \mathbb{E}[\psi_i(\xi)].$$

Moreover, since the mathematical expectation in (3.17) is taken w.r.t. a probability distribution from the ambiguity set $\mathcal{P}$, it follows by (2.4)

$$\mathbb{E}[G(x,\xi,\tau)] = -\mathbb{E}[f(x,\xi)] + \mathbb{E}[Y(\xi)] + \tau\mathbb{E}[(Y(\xi) - f(x,\xi))_+] \ge -\mathbb{E}[f(x,\xi)] + \mathbb{E}[Y(\xi)] + \tau\epsilon.$$

Furthermore, since $|\lambda_i| \le C_0$, for $i = 1, \cdots, q$, then the inequality above means $\mathbb{E}[G(x,\xi,\tau)] \to +\infty$ as $\tau \to +\infty$, and hence

$$e_\gamma(-R(w,\xi)) = \frac{1}{\gamma}\mathbb{E}\left[e^{\gamma R(x,\lambda,\tau,\xi)}\right] \to +\infty.$$

The discussions above show that a large τ value would violate the constraint of (3.13) regardless of the value of γ, which means τ must be bounded at its optimum.

The nonemptiness and boundedness of the set of optimal solutions to (3.18) follow from the boundedness of τ and the fact that the other variables of the problem are restricted to take a value from a compact set. ■

The proposition states some important properties of $\mathcal{F}(\gamma)$ and $\vartheta(\gamma)$ but it is short of characterizing their approximation to $\mathcal{F}$ and ϑ^*. In what follows, we give a quantitative description of the approximation of the feasible set and the optimal value. To this end, we need some conditions on the constraints.

Assumption 3.2 *There exist positive constants C and δ such that*

$$d(w,\mathcal{F}) \le C\left(\sup_{\xi\in\Xi} R(w,\xi)\right)_+ \tag{3.19}$$

for any $d(w,\mathcal{F}) \le \delta$.

Assumption 3.2 is the error bounded condition which plays an important role in the area of stability analysis, interested readers can find more details about error bound on the survey papers [3, 36]. A sufficient condition for Assumption 3.2 is that for every $w \in \mathcal{F}$, the constrained qualification due to Borwein [10] holds, that is, for any $\alpha \in [0,\infty)$, $0 \in \alpha\partial(\sup_{\xi\in\Xi} R(w,\xi)) + \mathcal{N}_{\mathcal{F}}(w_0)$ implies $\alpha = 0$, see [10, Theorem 3.2].

We are now ready to discuss the stability of problem (3.18) against variation of the parameter γ.

Theorem 3.1 *Assume: (a) $\mathcal{F}$ is a compact set, (b) ψ_i, $i = 1, \cdots, q$, is continuous, (c) Ω is a compact subset of a finite dimensional space, (d) Assumption 2.1 and the conditions of Lemma 3.1 hold for function $R(w, \xi)$. Then*

(i) for any $\epsilon > 0$, there exists a positive number γ_0 such that when $\gamma \in (\gamma_0, +\infty)$,

$$\mathbb{H}(\mathcal{F}(\gamma), \mathcal{F}) \leq \epsilon;$$

(ii) if, in addition, Assumption 3.2 holds then there exists a sufficiently large γ^* such that

$$\mathbb{H}(\mathcal{F}(\gamma), \mathcal{F}) \leq C \sup_{w \in W} \Delta_\gamma(w), \forall \gamma \in [\gamma^*, +\infty), \tag{3.20}$$

where C is the error bound parameter and

$$\Delta_\gamma(w) := \left(\sup_{\xi \in \Xi} R(w, \xi) - e_\gamma(-R(w, \xi)) \right);$$

(iii) the difference between the optimal values of $\vartheta(\gamma)$ and ϑ^* is bounded by $C \sup_{w \in W} \Delta_\gamma(w)$, i.e.,

$$|\vartheta(\gamma) - \vartheta^*| \leq C \sup_{w \in W} \Delta_\gamma(w).$$

Proof. By Proposition 3.1 (i), $\mathbb{D}(\mathcal{F}, \mathcal{F}(\gamma)) = 0$, therefore in Parts (i) and (ii), we only need to show the inequalities hold for $\mathbb{D}(\mathcal{F}(\gamma), \mathcal{F})$.

Part (i). Let ϵ be a fixed small positive number. Define

$$H(\epsilon) := \inf_{\substack{w \in W \\ d(w, \mathcal{F}) \geq \epsilon}} \sup_{\xi \in \Xi} R(w, \xi).$$

Then $H(\epsilon) > 0$. Let $\delta := H(\epsilon)/2$. Under conditions (d), it follows by Lemma 3.1 that there exists a positive number γ_0 such that

$$\sup_{w \in W} \left[\sup_{\xi \in \Xi} R(w, \xi) - e_\gamma(-R(w, \xi)) \right] \leq \delta,$$

for $\gamma \geq \gamma_0$. For any $w \in W$ with $d(w, \mathcal{F}) \geq \epsilon$,

$$e_\gamma(-R(w, \xi)) = \sup_{\xi \in \Xi} R(w, \xi) + e_\gamma(-R(w, \xi)) - \sup_{\xi \in \Xi} R(w, \xi) \geq H(\epsilon) - H(\epsilon)/2 = H(\epsilon)/2 > 0,$$

which implies $w \notin \mathcal{F}(\gamma)$. This means that for every $w \in \mathcal{F}(\gamma)$, we have $d(w, \mathcal{F}) < \epsilon$, that is, $\mathbb{D}(\mathcal{F}(\gamma), \mathcal{F}) \leq \epsilon$.

Part (ii). Under Assumption 3.2, it follows by Part (i) that there exists a sufficiently large γ^* such that

$$d(w, \mathcal{F}) \leq C \left(\sup_{\xi \in \Xi} R(w, \xi) \right)_+$$

for all $w \in \mathcal{F}(\gamma)$ when $\gamma \geq \gamma^*$. Since $w \in \mathcal{F}(\gamma)$ is equivalent to $e_\gamma(-R(w,\xi)) \leq 0$, then for any $w \in \mathcal{F}(\gamma)$,

$$\begin{aligned} d(w,\mathcal{F}) &\leq C\left(\sup_{\xi\in\Xi} R(w,\xi)\right)_+ - Ce_\gamma(-R(w,\xi)) \\ &\leq C\sup_{w\in W}\left(\sup_{\xi\in\Xi} R(w,\xi) - e_\gamma(-R(w,\xi))\right). \end{aligned}$$

This shows

$$\mathbb{D}(\mathcal{F}(\gamma),\mathcal{F}) \leq C\sup_{w\in W}\left(\sup_{\xi\in\Xi} R(w,\xi) - e_\gamma(-R(w,\xi))\right).$$

Part (iii). The conclusion follows from Part (i) by applying classical stability result [26, Theorem 1]. Here we include a proof for completeness. Let w^* and w_γ be an optimal solution of program (3.13) and program (3.18) respectively. Let τ^* and τ_γ be the corresponding first components. Then $\vartheta^* = -\tau^*$ and $\vartheta(\gamma) = -\tau_\gamma$. By Part (ii), there exists $\bar{w} \in \mathcal{F}$ such that

$$\|w_\gamma - \bar{w}\| \leq C\sup_{w\in W}\Delta_\gamma(w).$$

Let $\bar{\tau}$ be the corresponding component of $\bar{w}$. Then $\vartheta^* = -\tau^* \leq -\bar{\tau}$. Consequently we have

$$\vartheta^* \leq -\bar{\tau} \leq -\tau_\gamma + |\bar{\tau} - \tau_\gamma| \leq \vartheta(\gamma) + C\sup_{w\in W}\Delta_\gamma(w).$$

Exchanging the role of w_γ and w^* under the symmetry of Hausdorff distance between $\mathcal{F}(\gamma)$ and $\mathcal{F}$, we have

$$\vartheta(\gamma) \leq \vartheta^* + C\sup_{w\in W}\Delta_\gamma(w).$$

The conclusion follows. ■

Under condition (d), it follows by Lemma 3.1 that $\Delta_\gamma(w)$ goes to 0 uniformly for all $w \in W$ as $\gamma \to \infty$, Theorem 3.1 states that $|\vartheta(\gamma) - \vartheta^*| \to 0$ and it gives rise to a linear bound for $|\vartheta(\gamma) - \vartheta^*|$ in terms of $\Delta_\gamma(w)$.

4 Ambiguity set for data-driven problems

Delage and Ye [17] developed a distributionally robust optimization model for so-called data-driven problems where the distribution of the underlying random vector relies solely on historical data (see page 596 [17]). They argued that for these problems, it might be safer to rely on estimates of the mean and covariance matrix of the random vector. Let us denote the mean and covariance matrix by $\bar{\mu}$ and $\bar{\Sigma}$ respectively here. Then Delage and Ye's ambiguity set can be described as follows:

$$\mathcal{P}(\bar{\mu},\bar{\Sigma},\gamma_1,\gamma_2) := \left\{P \in \mathscr{P} : \begin{array}{l} \mathbb{E}_P[\xi-\bar{\mu}]^T\bar{\Sigma}^{-1}\mathbb{E}_P[\xi-\bar{\mu}] \leq \gamma_1 \\ \mathbb{E}_P[(\xi-\bar{\mu})(\xi-\bar{\mu})^T] \preceq \gamma_2\bar{\Sigma} \end{array}\right\}, \tag{4.21}$$

where γ_i, $i = 1,2$ are parameters. The parameters are introduced in that one may not be entirely confident in the estimates of the mean and covariance in data-driven problems. The formulation allows one to construct an ambiguity set where the true mean and covariance do not have to be matched precisely and this particularly helpful when $\bar{\mu}$ and $\bar{\Sigma}$ are estimated through empirical data.

Delage and Ye [17] explained that the first constraint in (4.21) assumes that the mean of ξ lies in an ellipsoid of size γ_1 centered at $\bar{\mu}$ and the second constraint forces the centered second moment to lie in a positive semidefinite cone defined with matrix inequality. A significant advantage of this particular way to construct the ambiguity set is that through a simple duality argument, the resulting distributionally robust optimization problem can be converted into a tractable convex semi-definite program.

In this section, we consider a variation of (4.21). Instead of using ellipsoid constraints for the mean and semidefinite constraints for the centered covariance, we propose simple box constraints for each component of these quantities. Specifically we define the ambiguity set as follows:

$$\mathcal{P}^* := \left\{ P \in \mathscr{P} : \begin{array}{ll} -\epsilon \leq (\mathbb{E}_P[\xi] - \bar{\mu})_i \leq \epsilon, & i = 1, \cdots, m \\ \|\mathbb{E}_P[(\xi - \bar{\mu})(\xi - \bar{\mu})^T] - \bar{\Sigma}\|_* \leq \sigma & \end{array} \right\}, \tag{4.22}$$

where $\|A\|_* = \max |a_{ij}|$. It is easy to verify that $\|\cdot\|_*$ is a norm for the matrix but without the sub-multiplicative property. Tütüncü and Koenig [57] apparently were among the first to consider box constraints of the mean and co-variance to define a set of means and co-variance matrices in portfolio optimization. They developed a minimax robust Markowitz's mean-variance model where the worst mean and convariance matrix are selected from the set. Our motivations here for this particular way of construction of the ambiguity set are three-fold: (a) the semidefinite constraint gives a holistic specification on the property of the centered covariance of random vector ξ but in practice this may not be entirely justified, and it might also be more convenient to give a lower and upper bound for each component of the centered covariance; (b) in some practical cases, we may have more information on the covariance of some random components than others, e.g., we know precisely the covariance between ξ_1 and ξ_2, our model allows us to set a different σ value for each component of the covariance matrix although we use a unified σ in this paper for simplicity of exposition; (c) the ambiguity set defined as such can be easily fitted into the framework of the moment problem in Section 3 for which we have already developed a new numerical scheme. In fact, it is easy to recast (4.22) as

$$\mathcal{P}^* := \left\{ P \in \mathscr{P} : \begin{array}{l} -\epsilon \leq \mathbb{E}[\psi_i(\xi)] \leq \epsilon, \ i = 1, \ldots, k \\ -\sigma \leq \mathbb{E}_P[\psi_j(\xi)] \leq \sigma, \ j = k+1, \ldots, q \end{array} \right\}, \tag{4.23}$$

where k is the dimension of random ξ, $q = \frac{k^2+3k}{2}$, $\psi_i(\xi) = \xi_i - \bar{\mu}_i$ and $\psi_j(\xi)$, $j = k+1, \cdots, q$, is the elements of the upper triangular of matrix $(\xi - \bar{\mu})(\xi - \bar{\mu})^T - \bar{\Sigma}$.

Analogous to the arguments in [17], in practice we may use samples to construct an estimate of the true mean and covariance. Let $\xi^1, \cdots, \xi^N$ be an independent and identically distributed sample of ξ and

$$\mu^N := \frac{1}{N} \sum_{s=1}^N \xi^s, \qquad \Sigma^N := \frac{1}{N} \sum_{s=1}^N (\xi^s - \mu^N)(\xi^s - \mu^N)^T.$$

Then we may consider the following sample based ambiguity set

$$\mathcal{P}^N := \left\{ P \in \mathscr{P} : \begin{array}{ll} -\epsilon \leq (\mathbb{E}_P[\xi] - \mu^N)_i \leq \epsilon, & i = 1, \cdots, m \\ \|\mathbb{E}_P[(\xi - \mu^N)(\xi - \mu^N)^T] - \Sigma^N\|_* \leq \sigma & \end{array} \right\}, \tag{4.24}$$

or equivalently

$$\mathcal{P}^N := \left\{ P \in \mathscr{P} : \begin{array}{l} -\epsilon \leq \psi_i^N(\xi) \leq \epsilon, \ i = 1, \ldots, k \\ -\sigma \leq \mathbb{E}_P[\psi_j^N(\xi)] \leq \sigma, \ j = k+1, \ldots, q \end{array} \right\},$$

where k, q are defined as above, $\psi_i^N(\xi) = \xi_i - \mu_i^N$ and $\psi_j^N(\xi)$, $j = k+1, \cdots, p$, are the elements of the upper triangular of matrix $(\xi - \mu^N)(\xi - \mu^N)^T - \Sigma^N$.

In order to justify the specific way for constructing the ambiguity set, we need to address a few theoretical questions: (a) does $\mathcal{P}^N$ converge to $\mathcal{P}^*$ as the sample size increases? (b) Is $\mathcal{P}^N$ statistically meaningful in the sense whether there is a significant likelihood such that the true probability distribution of ξ lies in $\mathcal{P}^N$? (c) does the optimal value and the optimal solutions obtained on the basis of $\mathcal{P}^N$ converge to their true counterpart? In the rest of this section, we address these questions in sequel. Note that throughout this section, we concentrate on asymptotic analysis of the ambiguity set and other statistical quantities for fixed entropic parameter γ. Thus, this section may be regarded as stability analysis of the distributionally robust reward risk ratio problem w.r.t. change of sample data.

Definition 4.1 Let $P, Q \in \mathscr{P}$ and $\mathscr{M}$ denote the set of measurable functions defined in the probability space $(\Xi, \mathscr{B})$. The *total variation metric* between P and Q is defined as (see e.g., page 270 in [2])

$$d_{TV}(P,Q) := \sup_{h\in\mathscr{M}} (\mathbb{E}_P[h(\xi)] - \mathbb{E}_Q[h(\xi)]),$$

where

$$\mathscr{M} := \{h : \mathbb{R}^k \to \mathbb{R} | h \text{ is } \mathscr{B} \text{ measurable}, \sup_{\xi\in\Xi} |h(\xi)| \le 1\}.$$

Using the total variation metric, we can define the distance from a point to a set, deviation from one set to another and Hausdorff distance between two sets in the space of $\mathscr{P}$. Specifically, let

$$d_{TV}(Q, \mathcal{P}^*) := \inf_{P\in\mathcal{P}^*} d_{TV}(Q,P),$$

$$\mathbb{D}_{TV}(\mathcal{P}^N, \mathcal{P}^*) := \sup_{Q\in\mathcal{P}^N} d_{TV}(Q, \mathcal{P}^*)$$

and

$$\mathbb{H}_{TV}(\mathcal{P}^N, \mathcal{P}^*) := \max\{\mathbb{D}_{TV}(\mathcal{P}^N, \mathcal{P}^*), \mathbb{D}_{TV}(\mathcal{P}^*, \mathcal{P}^N)\}.$$

Here $\mathbb{H}_{TV}(\mathcal{P}^N, \mathcal{P}^*)$ defines Hausdorff distance between $\mathcal{P}^N$ and $\mathcal{P}^*$ under the total variation metric in space $\mathscr{P}$.

Proposition 4.1 *Suppose that Ξ is a compact set. Then $\mathbb{H}_{TV}(\mathcal{P}^N, \mathcal{P}^*) \to 0$ as $N \to \infty$.*

Proof. When Ξ is a compact set, both $\mathcal{P}^N$ and $\mathcal{P}^*$ are compact sets under the total variation metric. For any given $Q \in \mathcal{P}^N$, it follows by [55, Lemma 2] that there exists a positive constant L such that

$$\begin{aligned} \mathbb{D}_{TV}(Q, \mathcal{P}^*) &\le L\left(\|(\mathbb{E}_Q[\Psi_I(\xi)] - \epsilon e)_+\| + \|(-\mathbb{E}_Q[\Psi_I(\xi)] - \epsilon e)_+\|\right) \\ &\quad + L\left(\|(\mathbb{E}_Q[\Psi_J(\xi)] - \sigma e)_+\| + \|(-\mathbb{E}_Q[\Psi_J(\xi)] + \sigma e)_+\|\right) \\ &\le L\left(\|(\mathbb{E}_Q[\Psi_I(\xi) - \Psi_I^N(\xi)])_+\| + \|(\mathbb{E}_Q[\Psi_J(\xi) - \Psi_J^N(\xi)])_+\|\right) \\ &\quad + L\left(\|(\mathbb{E}_Q[\Psi_I^N(\xi) - \Psi_I(\xi)])_+\| + \|(\mathbb{E}_Q[\Psi_J^N(\xi) - \Psi_J(\xi)])_+\|\right), \end{aligned}$$

where

$$\Psi_I := \begin{bmatrix} \psi_1 \\ \vdots \\ \psi_k \end{bmatrix}, \Psi_J := \begin{bmatrix} \psi_{k+1} \\ \vdots \\ \psi_q \end{bmatrix}, \Psi_I^N := \begin{bmatrix} \psi_1^N \\ \vdots \\ \psi_k^N \end{bmatrix}, \Psi_J^N := \begin{bmatrix} \psi_{k+1}^N \\ \vdots \\ \psi_q^N \end{bmatrix},$$

e is the vector with each component being 1 and for a vector a, $(a)_+ := \max(0, a)$ with the maximum being taken componentwise. Likewise, for any given $Q \in \mathcal{P}^*$,

$$
\begin{aligned}
\mathbb{D}_{TV}(Q, \mathcal{P}^N) \;\; \leq \;\; & L\left(\|(\mathbb{E}_Q[\Psi_I(\xi) - \Psi_I^N(\xi)])_+\| + \|(\mathbb{E}_Q[\Psi_J(\xi) - \Psi_J^N(\xi)])_+\|\right) \\
& +L\left(\|(\mathbb{E}_Q[\Psi_I^N(\xi) - \Psi_I(\xi)])_+\| + \|(\mathbb{E}_Q[\Psi_J^N(\xi) - \Psi_J(\xi)])_+\|\right).
\end{aligned}
$$

Since $\mu^N \to \bar{\mu}$ and $\Sigma^N \to \bar{\Sigma}$, it is not difficult to see that Ψ_I^N and Ψ_J^N converge to Ψ_I and Ψ_J uniformly over Ξ as $N \to \infty$. Thus both $\mathbb{D}_{TV}(Q, \mathcal{P}^*)$ and $\mathbb{D}_{TV}(Q, \mathcal{P}^N)$ converge to zero as N tends to infinity, which implies $\mathbb{H}_{TV}(\mathcal{P}^N, \mathcal{P}^*) \to 0$. The proof is complete. ■

Next, we address question (b). We need some intermediate technical results which can be easily established by Shawe-Taylor and Cristianini's theorems, see [52].

Lemma 4.1 *If ξ is essentially bounded by a positive number ρ, then for any given positive small number δ,*

(i) with probability at least $1-\delta$ over the choice of sample ξ,

$$\|\mathbb{E}[\xi] - \mu^N\| \leq \frac{\rho^2}{\sqrt{N}}\left(2 + \sqrt{2\ln\frac{2}{\delta}}\right),$$

(ii) with probability at least $(1-\delta)^2$ over the choice of sample ξ,

$$\|\mathbb{E}[(\xi - \mu^N)(\xi - \mu^N)^T] - \Sigma^N\|_* \leq \frac{\rho^2 + 2\rho^3}{\sqrt{N}}\left(2 + \sqrt{2\ln\frac{2}{\delta}}\right).$$

Proof. Part (i) follows straightforwardly from [52, Theorem 3]. We only prove Part (ii). Define

$$A^N := \frac{1}{N}\sum_{i=1}^N \xi_i\xi_i^T, \quad A := \mathbb{E}[\xi\xi^T].$$

Then

$$
\begin{aligned}
& \|\mathbb{E}[(\xi - \mu^N)(\xi - \mu^N)^T] - \Sigma^N\|_* \\
& = \|\mathbb{E}[(\xi - \mu^N)(\xi - \mu^N)^T] - \frac{1}{N}\sum_{i=1}^N (\xi_i - \mu^N)(\xi_i - \mu^N)^T\|_* \\
& = \|A - 2\mathbb{E}[\xi](\mu^N)^T + \mu^N(\mu^N)^T - A^N + \mu^N(\mu^N)^T\|_* \\
& \leq \|A - A^N\|_* + 2\|\mathbb{E}[\xi](\mu^N)^T - \mu^N(\mu^N)^T\|_* \\
& \leq \|A - A^N\|_F + 2\|\mathbb{E}[\xi] - \mu^N\|_F \cdot \|(\mu^N)^T\|_F.
\end{aligned}
$$

Consequently, by [52, Corollary 5] and Part (i) that with probability at least $(1-\delta)^2$

$$\|\mathbb{E}[(\xi - \mu^N)(\xi - \mu^N)^T] - \Sigma^N\|_* \leq \frac{\rho^2}{\sqrt{N}}\left(2 + \sqrt{2\ln\frac{2}{\delta}}\right) + \frac{2\rho^3}{\sqrt{N}}\left(2 + \sqrt{2\ln\frac{2}{\delta}}\right).$$

The proof is complete. ■

With Lemmas 4.1, we are ready to demonstrate that with appropriate choice of parameters ϵ and σ, the true distribution lies in $\mathcal{P}^N$ with probability at least $(1-\delta)^2$ for the given small positive number δ.

Theorem 4.1 *Suppose that ξ is essentially bounded by a positive number ρ and the parameters ϵ and σ are chosen as follows*

$$\epsilon_N := \frac{\rho^2}{\sqrt{N}}\left(2+\sqrt{2\ln\frac{2}{\delta}}\right), \qquad \sigma_N := \frac{\rho^2+2\rho^3}{\sqrt{N}}\left(2+\sqrt{2\ln\frac{2}{\delta}}\right).$$

Then with probability at least $(1-\delta)^2$ over the choice of sample ξ, the true distribution of ξ lies in the ambiguity set $\mathcal{P}^N$.

Note that Shawe-Taylor and Cristianini presented a similar result when the matrix norm is replaced by the Frobenius norm. Theorem 4.1 is an analogue of their result under the new matrix norm $\|\cdot\|_*$.

Using the ambiguity set $\mathcal{P}^N$, we can derive the dual problem of (2.9) coupled by the entropic approximation as follows:

$$\begin{array}{cl} \min\limits_{x\in X,\tau\in\mathbb{R},\lambda\in\mathbb{R}^p\times\mathbb{R}^{q-p}_+} & -\tau \\ \text{s.t.} & e_\gamma(-R^N(x,\lambda,\tau,\xi))\leq 0, \end{array} \tag{4.25}$$

where $R^N(x,\tau,\lambda_0,\lambda,\xi) := G(x,\xi,\tau) - \sum_{i=1}^{p}\lambda_i\psi_i^N(\xi)$.

Assume $\mathcal{F},\mathcal{F}(\gamma),\vartheta^*$ and $\vartheta(\gamma)$ are as defined in Proposition 3.1. Let $\mathcal{F}^N(\gamma)$ and $\vartheta^N(\gamma)$ denote the set of feasible solutions and the optimal value of problem (4.25), $S^N(\gamma)$, $S(\gamma)$ and S^* denote the optimal solution of problems (4.25), (3.18) and (3.13) respectively. Let $\mathcal{F}^s(\gamma)$ denote the set of strict feasible solutions of problem (3.18).

The following theorem summarizes the convergence of problem (4.25) to problems (3.18) and (3.13) in terms of the optimal value and the optimal solutions.

Theorem 4.2 *Assume: (a) D is a compact set and $D\cap S(\gamma)\neq\emptyset$, $D\cap S^N(\gamma)\neq\emptyset$; (b) $\mathrm{cl}\mathcal{F}^s(\gamma)\cap S(\gamma)\neq\emptyset$. Then*

(i) $\limsup_{N\to\infty} S^N(\gamma)\cap D\subseteq S(\gamma)\cap D$ *and* $\lim_{N\to\infty}\vartheta^N(\gamma)=\vartheta(\gamma)$;

(ii) if, in addition, Assumption 3.2 holds, there exist $\hat{N}$ and $\hat{\gamma}$ sufficiently large such that

$$\mathbb{D}(\mathcal{F}^N(\gamma),\mathcal{F}) \leq C\sup_{w\in W}\Delta_\gamma^N(w),$$

and

$$\vartheta^* - \vartheta^N(\gamma) \leq C\sup_{w\in W}\Delta_\gamma^N(w),$$

for $\gamma\geq\hat{\gamma}$ and $N\geq\hat{N}$, where C is a positive constant and

$$\Delta_\gamma^N(w) := \left(\sup_{\xi\in\Xi} R(w,\xi) - e_\gamma(-R^N(w,\xi))\right)_+.$$

Proof. Part (i). Since D is a compact set and $\Psi_I^N(\cdot)$ and $\Psi_J^N(\cdot)$ converge to $\Psi_I(\cdot)$ and $\Psi_J(\cdot)$ uniformly on Ξ, it is easy to show that $\limsup_{N\to\infty}\mathcal{F}^N(\gamma)\subseteq\mathcal{F}(\gamma)$. By taking a subsequence if necessary we may assume for the simplicity of notation that $(x^N, \tau^N, \lambda^N) \to (x^*, \tau^*, \lambda^*) \in \mathcal{F}(\gamma)$, which implies

$$\lim_{N\to\infty} v^N = \lim_{N\to\infty} -\tau^N = -\tau^* \geq v^*.$$

Under condition (b), there exists an optimal solution $(\bar{x}, \bar{\tau}, \bar{\lambda}) \in S(\gamma)$ such that $(\bar{x}, \bar{\tau}, \bar{\lambda}) \in \mathrm{cl}\mathcal{F}^s(\gamma)$. It is easy to show that there exists $(x^N, \tau^N, \lambda^N) \in \mathcal{F}^N(\gamma)$ such that $||(x^N, \tau^N, \lambda^N) - (\bar{x}, \bar{\tau}, \bar{\lambda})\| \to 0$ as $N \to \infty$. Therefore

$$v^* = -\bar{\tau} = \lim_{N\to\infty} -\tau^N \geq \lim_{N\to\infty} v^N = \tau^*.$$

This shows $(x^*, \tau^*, \lambda^*) \in S(\gamma)$. The convergence of optimal value follows from the continuity of the object.

Part (ii). By Part (i) and Theorem 3.1, there exist $\hat{N}$ and $\hat{\gamma}$ sufficiently large such that for any $N \geq \hat{N}$ and $\gamma \geq \hat{\gamma}$

$$(\mathcal{F}^N(\gamma), \mathcal{F}) \leq \delta,$$

where δ is the parameter in Assumption 3.2. For any $w \in \mathcal{F}^N(\gamma)$, it follows by (3.19)

$$\begin{aligned} d(w, \mathcal{F}) &\leq C \left(\sup_{\xi\in\Xi} R(w, \xi) \right)_+ - Ce_\gamma(-R^N(w, \xi)) \\ &\leq C \sup_{w\in W} \left(\sup_{\xi\in\Xi} R(w, \xi) - e_\gamma(-R^N(w, \xi)) \right)_+ . \end{aligned}$$

This shows

$$\mathbb{D}(\mathcal{F}^N(\gamma), \mathcal{F}) \leq C \sup_{w\in W} \left(\sup_{\xi\in\Xi} R(w, \xi) - e_\gamma(-R^N(w, \xi)) \right)_+ .$$

Let w^* and w_γ^N be an optimal solution of problem (3.13) and problem (4.25) respectively, and τ^* and τ_γ^N be the corresponding components to τ. Then $\vartheta^* = -\tau^*$ and $\vartheta^N(\gamma) = -\tau_\gamma^N$. By Part (ii), there exists $\bar{w} \in \mathcal{F}$ such that

$$\|w_\gamma^N - \bar{w}\| \leq C \sup_{w\in W} \Delta_\gamma^N(w).$$

Let $\bar{\tau}$ be the corresponding component of $\bar{w}$. Then we have

$$\vartheta^* \leq -\bar{\tau} \leq -\tau_\gamma^N + |\bar{\tau} - \tau_\gamma^N| \leq \vartheta^N(\gamma) + C \sup_{w\in W} \Delta_\gamma^N(w).$$

The proof is complete. ■

Compared to Theorem 3.1, Theorem 4.2 has only managed a single sided quantitative analysis of the feasible set and the optimal value because we are not certain $\mathcal{F}$ lies in $\mathcal{F}^N(\gamma)$. Note also that $\Delta_\gamma^N(w) \to 0$ as $N, \gamma \to \infty$. This can be easily observed through the inequality below

$$\Delta_\gamma^N(w) \leq \left(\sup_{\xi\in\Xi} R(w, \xi) - e_\gamma(-R(w, \xi)) \right)_+ + \left(e_\gamma(-R(w, \xi) - e_\gamma(-R^N(w, \xi)) \right)_+ ,$$

where the first term goes to zero as $\gamma \to \infty$ by Lemma 3.1 and the second term tends to zero as $N \to \infty$ by [45, Proposition 7, Chapter 6].

5 Implicit Dinkelbach method

In this section, we propose an iterative scheme for solving problem (3.18). As we discussed in Section 2, the problem is nonconvex. However, for each fixed τ, the underlying function in the first constraint is convex. This motivates us to exploit the specific structure of the problem and propose an algorithm which follows the framework of the Dinkelbach method. To this end, we write (3.18) in a slightly neater form

$$\begin{array}{ll} \min\limits_{x\in X,\tau\in\mathbb{R},\lambda\in\Lambda} & -\tau \\ \text{s.t.} & e_\gamma(-R(x,\lambda,\tau,\xi)) \leq 0, \end{array} \tag{5.26}$$

where $\Lambda := \mathbb{R}^p \times \mathbb{R}_+^{q-p}$. The following algorithm presents an iterative scheme for solving (5.26).

Algorithm 5.1 *(Implicit Dinkelbach method)*

Step 1. Given x_0, λ_0, set $k = 0$.

Step 2. For given x_k, λ_k, solve τ_k as a solution to the following equation:

$$e_\gamma(-R(x_k,\lambda_k,\tau,\xi)) = 0. \tag{5.27}$$

Step 3. For given τ_k, solve

$$\begin{array}{ll} \min & e_\gamma(-R(x,\lambda,\tau_k,\xi)) \\ \text{s.t.} & x \in X, \\ & \lambda \in \Lambda, \end{array} \tag{5.28}$$

and denote the optimal value and the optimal solution by $\Delta(\tau_k)$ and (x_{k+1}, λ_{k+1}) respectively.

Step 4. If $\Delta(\tau_k) = 0$, stop. Return τ_k as the optimal value and $(x_{k+1}, \lambda_{k+1}, \tau_k)$ as the optimal solution. Otherwise go to Step 2.

We call the algorithm implicit Dinkelbach function method as the numerical scheme resembles the algorithmic procedures of Dinkelbach method at Steps 2 and 3 when we update variable τ_k and (x_k, λ_k). The word *implicit* is used to reflect the fact that τ_k is imbedded in a nonlinear equation. There are two main differences from the standard Dinkelbach method: (a) the objective function in the minimization problem (5.28) is nonlinear w.r.t. parameter τ or w.r.t (x, λ), (b) τ_k is determined in Step 2 by solving a nonlinear equation.

Proposition 5.1 *Let τ_k be generated by Algorithm 5.1 and $\Delta(\tau)$ be defined as the optimal value of program (5.28). Then*

(i) $\Delta(\tau)$ is a continuous function;

(ii) for each k, $-\tau_{k+1} < -\tau_k$;

(iii) $\Delta(\tau_k) \leq 0$ and τ_k is optimal if and only if $\Delta(\tau_k) = 0$.

Proof. Part (i). Since X and Λ are assumed to be bounded, it is easy to verify that $e_\gamma(-R(x,\lambda,\tau,\xi))$ is uniformly continuous w.r.t. τ. By [43, Theorem 1.17], the optimal value function is continuous w.r.t. τ.

Parts (ii) and (iii). By the definition of τ_k and $\Delta(\tau_k)$, we know $\Delta(\tau_k) \leq 0$. For fixed γ, x and λ, $e_\gamma(-R(x, \lambda, \tau, \xi))$ is strictly increasing in τ as by definition $R(x, \lambda, \tau, \xi)$ does, see Proposition 2.3 and (3.14). Therefore (5.27) has a unique solution. Moreover, τ_k is the optimal value of problem (5.26) if and only if $\Delta(\tau_k) \geq 0$. To see this, assume for the sake of a contradiction that $\Delta(\tau_k) < 0$, that is,

$$e_\gamma(-R(x_{k+1}, \lambda_{k+1}, \tau_k, \xi)) < 0.$$

Then we can find a positive number δ such that

$$e_\gamma(-R(x_{k+1}, \lambda_{k+1}, \tau_k + \delta, \xi)) = 0.$$

Thus $-\tau_{k+1} = -\tau_k - \delta < -\tau_k$ which contradicts the assumption that τ_k is optimal. Conversely, if $\Delta(\tau_k) = 0$, then

$$\min_{(x,\lambda) \in X \times \Lambda} e_\gamma(-R(x, \lambda, \tau_k, \xi)) = 0.$$

Since $e_\gamma(-R(x, \lambda, \cdot, \xi))$ is strictly increasing in τ,

$$e_\gamma(-R(x, \lambda, \tau_k + \delta, \xi)) > 0, \quad \forall \delta > 0, \forall (x, \lambda) \in X \times \Lambda,$$

which means that $\tau_k + \delta$ is not feasible to problem (5.26) for any $\delta > 0$. This shows $-\tau_k$ is the optimal value and $(x_{k+1}, \lambda_{k+1}, \tau_k)$ is the optimal solution of problem (5.26). ∎

The following theorem states that the Algorithm 5.1 either terminates in a finite number of iterations or generates a sequence of approximation of optimal values converging to the optimal value of problem (5.26).

Theorem 5.1 *Let $\{-\tau_k\}$ be a sequence generated by Algorithm 5.1. Under Assumption 2.1, the sequence is monotonically decreasing and it converges to the optimal value of problem (5.26).*

Proof. The Monotonicity follows from Proposition 5.1. In what follows, we show the convergence. Let us first consider the case when the algorithm terminates after k iterations, i.e., $\Delta(\tau_k) = 0$. By Proposition5.1, τ_k is the optimal value.

Now suppose that $\{-\tau_k\}$ is an infinite sequence. Under Assumption 2.1, the sequence is lower bounded. Therefore there exists some positive number τ^* such that $-\tau_k \downarrow -\tau^*$, which means τ^* is the upper bound of the sequence $\{\tau_k\}$. It suffices to show that $\Delta(\tau^*) \geq 0$.

Assume for the sake of a contradiction that $\Delta(\tau^*) < 0$. Denote by (x^*, λ^*) the corresponding optimal solution to problem (5.28) for the given τ^*. Then there exists a positive constant α_0 such that

$$e_\gamma(-R(x^*, \lambda^*, \tau^*, \xi)) \leq -\alpha_0.$$

Since $e_\gamma(-R(x^*, \lambda^*, \tau, \xi))$ is monotonically increasing w.r.t. τ, and $\tau_k < \tau^*$, then

$$e_\gamma(-R(x^*, \lambda^*, \tau_k, \xi)) < -\alpha_0, \forall k.$$

Denote by (x_{k+1}, λ_{k+1}) the corresponding optimal solution of problem (5.28) for the given τ_k. Then the optimality of $(x_{k+1}, \lambda_{k+1}, \tau_k)$ means

$$e_\gamma(-R(x_{k+1}, \lambda_{k+1}, \tau_k, \xi)) \leq e_\gamma(-R(x^*, \lambda^*, \tau_k, \xi)) < -\alpha_0.$$

Since $\tau_k \to \tau^*$ and $e_\gamma(\cdot)$ is continuous, there exists a sufficiently large k_0 such that

$$e_\gamma(-R(x_{k_0+1}, \lambda_{k_0+1}, \tau^*, \xi)) \leq -\alpha_0/2.$$

On the other hand, looking at Step 2 of the algorithm, τ_{k_0+1} is chosen to satisfy

$$e_\gamma(-R(x_{k_0+1}, \lambda_{k_0+1}, \tau_{k_0+1}, \xi)) = 0.$$

Combining last two equation and taking into account the monotonicity of $e_\gamma(-R(x_{k_0+1}, \lambda_{k_0+1}, \cdot, \xi))$, we obtain $\tau_{k_0+1} > \tau^*$. This contradicts the fact that τ^* is the upper bound of the sequence $\{\tau_k\}$. The proof is complete. ■

6 Numerical tests

We evaluate the performance of the robust ratio portfolio optimization model (2.1) by using a historical data set of 10 assets (Aberdeen Asset Management plc, Admiral Group PLC, AMEC PLC, Anglo American PLC, Antofagasta PLC, AstraZeneca PLC, Aviva PLC, Babcock International Group PLC, BAE Systems PLC and Barclays PL) over a time horizon of 4 years (from 7th Dec 2009 to 18th Nov 2013) with a total of 1001 records on the historical stock returns (these are obtained from http://finance.google.com with adjustment for stock splitting). We have carried out out-of-sample tests with a rolling window of 500 days, that is, we use first 500 data to calculate the optimal portfolio strategy for day 501 and then move on a rolling basis. To simplify the discussions, we ignore the transaction fee, therefore the total value of portfolio is

$$f(x, \xi) := r_1 x_1 + r_2 x_2 + \cdots + r_{10} x_{10},$$

where r_i is the random return of asset i.

Table 6.1: Daily return

Strategy	L	H	A	Down	Up
Our DRO Model	0.9564	1.0590	1.006	235	265
Popescu's DRO model	0.8954	1.0839	1.004	244	256
SP model	0.9532	1.0664	1.003	235	265
EW model (bench mark)	0.9717	1.0541	1.004	241	259

In implementing the proposed robust ratio portfolio optimization model (2.1) and the numerical scheme, we use the ambiguity set defined in (4.24). The parameters for the confidence interval δ in Theorem 4.1 is fixed at 0.01 and the corresponding parameters $\epsilon_{500} = 1.22$ and $\sigma_{500} = 4.01$. Moreover, we use the equally weighted portfolio as a benchmark strategy $y(\xi)$ and set the small positive number satisfying (2.4) by 0.01. The entropic parameter γ is fixed at 1000. If we set the robust parameters $\epsilon_{500} = 0$ and $\sigma_{500} = 0$, the ambiguity set collapses to Popescu's exact mean-covariance distributionally robust model [40]. In the numerical experiment, we compare our model with Popescu's model for the reward risk ratio problem (2.3). Moreover, we also compare the model with the stochastic programming model where (2.1) is solved by approximating the true probability distribution P with empirical data. We implement Algorithm 5.1 on MATLAB 7.6 installed in a PC with Windows 7 operating system. We use the built-in optimization solver *fmincon* to solve minimization problem (5.27).

Table 6.1 summarizes the daily returns generated by the three portfolio models, where "L", "H" and "A" denote respectively the lowest, highest and average return from day 501 to day 1000. We record the number of days when the overall portfolio return rate falls below 1 and exceeds 1, denoted by "Down" and "Up". We can see that our distributionally robust optimization model achieves comparable average daily return and displays stable performance with a narrow range between the best and worst return curves. Moreover, in comparison with the benchmark strategy,

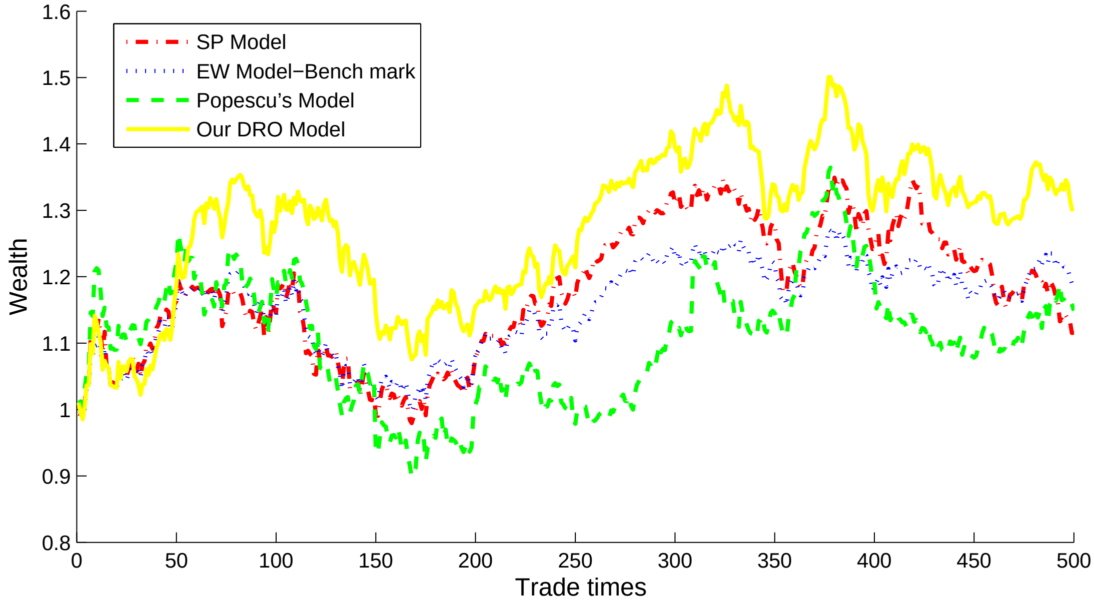

Figure 6.1: wealth evolution with the trading times.

both the stochastic programming model and our distributionally robust model generate more "Up" times. Figure 6.1 depicts the performance of three portfolios. The figure indicates that the wealth curve of our DRO model outperforms SP model and Popescu's model and it lies well above the the bench mark wealth curve. At the end of the time horizon, the total wealth from our PRO model is 1.3026 compared to 1.1158 by SP model and 1.1491 by Popescu's model. The SP wealth curve is very close to the bench mark curve but it falls below at the end of investment horizon. Compared to the SP model, Popescu's model displays higher average daily return and wealth at the end of horizon but it fails at the stability with highest and lowest daily return. The experiments show that our model outperforms Popescu's model and the SP model. When the assets number n is large, our model may fail since the resulting problem size may increases significantly.

Acknowledgements. The authors are deeply indebted to Victor DeMiguel for numerous discussions on an earlier version of the paper.

References

[1] E. Anderson, H. Xu and D. Zhang, CVaR approximation for minimax and robust convex optimization, March, 2013.

[2] K. B. Athreya and S. N. Lahiri, *Measure theory and probability theory*, Springer, NewYork, 2006.

[3] D. Azé, *A survey on error bounds for lower semicontinuous functions*, ESAIM: Proceedings, Vol. 13, pp. 1–17, 2003.

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

[5] D. Bertsimas and I. Popescu, On the relation between option and stock prices: a convex optimization, *Operations Research,* Vol. 50, pp. 358–374, 2002.

[6] D. Bertsimas and I. Popescu, Optimal inequalities in probability theory: a convex optimization approach, *SIAM Journal of Optimization,* Vol. 15, pp. 780–804, 2005.

[7] M. J. Best and R. Grauer, On the sensitivity of mean-variance efficient portfolios to changes in asset means: some analytical and computational results, *Review of Financial Studies,* Vol. 4, pp. 315–342, 1991.

[8] A. Biglova, S. Ortobelli, S. Rachev and S. Stoyanov, Different approaches to risk estimation in portfolio theory, *The Journal of Portfolio Management,* Vol. 31, pp. 103–112, 2004.

[9] F. Black and R. Litterman, Global portfolio optimization, *Financial Analysts Journal,* Vol. 48, pp. 28–43, 1992.

[10] J. M. Borwein, Stability and regular points of inequality systems, *Journal of Optimization Theory and Applications*, Vol. 48, pp. 9–52, 1986.

[11] M. Broadie, Computing efficient frontiers using estimated parameters, *Annals of Operations Research,* Vol. 45, pp. 21–58, 1993.

[12] L. Chen, S. He and S. Zhang, Tight bounds for some risk measures, with applications to robust portfolio selection, *Operations Research,* Vol. 59, pp. 847–865, 2011.

[13] V. K. Chopra and W. T. Ziemba, The effects of errors in means, variances, and covariance on optimal portfolio choice, *Journal of Portfolio Management,* Vol. 19, pp. 6–11, 1993.

[14] F. H. Clarke, *Optimization and nonsmooth analysis*, Wiley, New York, 1983.

[15] R. E. Curto and L. A. Failow, *Solution of the truncated complex moment problem for flat data,* Memoirs of the American Mathematical Society, 1996.

[16] E. Delage, *Distributionally robust optimization in context of data-driven problems,* PhD Thesis, Stanford University. Palo Alto, CA, 2010.

[17] E. Delage and Y. Ye, Distributionally robust optimization under moment uncertainty with application to data driven problems, *Operations Research,* Vol. 58, pp. 595–612, 2010.

[18] J. Dupačová, The minimax approach to stochastic programming and an illustrative application, *Stochastics*, Vol. 20, pp. 73–88, 1987.

[19] J. Dupačová, Uncertanities in minimax stochastic programs, *Optimization*, Vol. 60, pp. 10–11, 2011.

[20] L. El-Ghaoui and H Lebret, Robust solutions to least-square problems to uncertain data matrices, *SIAM Journal on Matrix Analysis and Applications,* Vol. 18, pp. 1035–1064, 1997.

[21] L. El-Ghaoui, L. M. Oks and F. Oustry, Worst-case Value-at-Risk and robust portfolio optimization: A conic programming approach, *Operations Research,* Vol. 51, pp. 889–909, 2003.

[22] H. Föllmer and A. Schied, *Stochastic Finance: An Introduction in Discrete Time*, Walter de Gruyter, Berlin, 2004.

[23] D. Goldfarb and G. Iyengar, Robust portfolio selection problems, *Mathematics of Operations Research,* Vol. 28, pp. 1–38, 2003.

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

[25] R. Kaas, M. Goovaerts, J. Dhaene and M. Denuit, *Modern Actuarial Risk Theory Using R,* Springer, 2008.

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

[27] A. J. Kleywegt, A. Shapiro and T. Homem-de Mello, The sample average approximation method for stochastic discrete optimization, *SIAM Journal of Optimization,* Vol. 12, pp. 479–502, 2002.

[28] J. B. Lasserre, Global optimization with polynomials and the problem of moments, *SIAM Journal on Optimization,* Vol. 11, pp. 796–817, 2001.

[29] Y. Liu and H. Xu, Entropic approximation for mathematical programs with robust equilibrium constraints, *SIAM Journal on Optimization*, Vol. 24, pp. 933–958, 2014.

[30] Z. Lu, A new cone programming approach for robust portfolio selection, *Tech. Rep., Department of Mathematics, Simon Fraser University, Burnaby, BC,* 2006.

[31] Z. Lu, Robust portfolio selection based on joint ellipsoidal uncertainty set, *Optimization Methods and Software,* Vol. 26, pp. 89–104, 2011.

[32] H. M. Markowitz, Portfolio Selection, *The Journal of Finance*, Vol. 1 , pp. 77–91, 1952.

[33] K. Natarajan, D. Pachamanova and M. Sim, Incorporating asymmetric distributional information in robust value-at-risk optimization, *Management Science,* Vol. 54, pp. 573–585, 2008.

[34] K. Natarajan, D. Pachamanova and M. Sim, Constructing risk measures from uncertainty sets, *Operations Research,* Vol. 57, pp. 1129–1141, 2009.

[35] W. Ogryczak and A. Ruszyński, On consistency of stochastic dominance and mean-semiderivation models, *Mathematical Programming,* Vol. 89, pp. 217–232, 2001.

[36] J-S. Pang, *Error bounds in mathematical programming, Mathematical Programming,* Vol. 79, pp. 299–332, 1997.

[37] P. Parrilo, *Structured semidefinite programs and semialgebraic geometry methods in robustness and optimization,* PhD Thesis, California Institute of Technology, 2000.

[38] G. Ch. Pflug, Stochastic optimization and statistical inference, A. Rusczyński and A. Shapiro, eds. *Stochastic Program.*, Handbooks in OR & MS, Vol. 10, North-Holland Publishing Company, Amsterdam, 2003.

[39] M. Pinar and R. H. Tütüncü, Robust profit opportunities in risky financial portfolios, *Operations Research Letters,* Vol. 33, pp. 331–340, 2005.

[40] I. Popescu, Robust mean-covariance solutions for stochastic optimization, *Operations Research,* Vol. 55, pp. 98–112, 2007.

[41] M. Putinar, Positive polynomials on compact semialgebraic sets, *Indian university Mathematics Journal,* Vol. 42, pp. 969–984, 1993.

[42] 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.

[43] R.T. Rockafellar and R.J-B. Wets, *Variational analysis*, Springer, Berlin, 1998 (3rd printing 2009).

[44] W. Römisch, Stability of stochastic programming problems, in *Stochastic Programming*, A. Ruszczyński and A. Shapiro, eds., Elsevier, Amsterdam, pp. 483–554, 2003.

[45] A Ruszczyński and A. Shapiro, *Stochastic Programming,* Elsevire, Amsterdam, 2003.

[46] H. Scarf, *A minmax solution of an inventory problem,* K. S. Arrow, s. Karlin, H. E. Scarf. Studies in the Mathematical Theory of Inventory and Production, Stanford University Press, 1958.

[47] K. Schmüdgen, The k-moment problem for compact semialgebraic sets, *Mathematische Annalen,* Vol. 289, pp. 203–206, 1991.

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

[49] A. Shapiro, Asymptotic analysis of stochastic programs, *Annals of Operations Research*, Vol. 30, pp. 169–186, 1991.

[50] A. Shapiro, Monte Carlo sampling methods, A. Rusczyński and A. Shapiro, eds. *Stochastic Program.*, Handbooks in OR & MS, Vol. 10, North-Holland Publishing Company, Amsterdam, 2003.

[51] W. F. Sharpe, Mutual funds performance, *Journal of Business,* Val. 39, pp. 119–138, 1966.

[52] J. Shawe-Taylor and N. Cristianini, estimating the moments of a random vector with applications, *Proc. GRETSI 2003 Conference*, pp. 47–52.

[53] T. J. Stieltjes, Recherches sur les fractions continues, *Annales de la Facculte des Sciences de Toulouse,* Vol. 8, pp. 1–122, 1894.

[54] S. Stoyanov, S. Rachev and F. Fabozzi, Optimal financial portfolios, *Applied Mathematics Finance,* Vol. 14, pp. 403–438, 2007.

[55] H. Sun and H. Xu, Asymptotic Convergence Analysis for Distributional Robust Optimization and Equilibrium Problems, *Mathematics of Operations Research*, to appear.

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

[57] R. H. Tütüncü and M . Koenig, Robust asset allocation, *Annals of Operations Research,* Vol. 132, pp. 157–187, 2004.

[58] Z. Wang, P. W. Glynn and Y. Ye, Likelihood robust optimization for data-driven Newsvendor problems, manuscript, 2012.

[59] W. Wiesemann, D. Kuhn and B. Rustem, Robust Markov decision process, *Optimization Online*, 2012.

[60] J. Zackova, On minimax solution of stochastic linear programming problems, *Casopis pro Pestovani Matematiky,* Vol. 91, pp. 423–430, 1966.

[61] S. Zhu and M. Fukushima, Worst-case conditional value-at-risk with application to robust portfolio management, *Operations Research,* Vol. 57, pp. 1155–1168, 2009.

[62] S. Zhu, D. Li and S. Y. Wang, Robust portfolio selection under downside risk measures, *Quantitative Finance,* Vol. 9, pp. 869–885, 2009.

[63] S. Zhu, M. Fan and D. Li, Robust portfolio selection via learning with mixture model, February, 2013.