

DISTRIBUTIONALLY ROBUST REWARD-RISK RATIO OPTIMIZATION WITH MOMENT CONSTRAINTS

YONGCHAO LIU*, RUDABEH MESKARIAN[†], AND HUIFU XU[‡]

Abstract. Reward-risk ratio optimization is an important mathematical approach in finance. 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 to mitigate the risk arising from ambiguity of the true distribution. We consider a distributionally robust reward-risk ratio optimization model varied from ex ante Sharpe ratio where the ambiguity set is constructed through prior moment information and the return function is not necessarily linear. We transform the robust optimization problem into a nonlinear semi-infinite programming problem through standard 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. The proposed robust formulation and numerical schemes can be easily applied to stochastic fractional programming problems.

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

AMS subject classifications.

1. Introduction. Since the pioneering work by Markowitz on mean-variance portfolio selection [30], 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 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. [48] 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 [50]: 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 [44, 45]. In the recent progress of risk management, the performance ratio optimization models have attracted substantial amount of attention from academics and practitioners.

Since the publication of the Sharpe ratio [44], 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

* School of Mathematical Sciences, Dalian University of Technology, Dalian, 116024, China (lyc@dlut.edu.cn). The work of this author was carried out in the School of Mathematical Sciences, University of Southampton as a postdoctoral research fellow supported by EPSRC grant EP/M003191/1.

[†] Engineering Systems and Design, Singapore University of Technology and Design, Singapore. (Rudabeh@sutd.edu.sg)

[‡]School of Mathematical Sciences, University of Southampton, SO17 1BJ, Southampton, UK (H.Xu@soton.ac.uk).

empirical comparison, see Biglova et al. [8], Rachev et al. [39] 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 [56]. 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 [19], Halldórsson and Tütüncü [22], Lu [29] investigated the robust mean-variance portfolio selection problem. El Ghaoui et al. [16] studied the robust mean-VaR portfolio selection models. Rujeerapaiboon et al. [41] studied the distributionally robust growth-optimal portfolios. Natarajan et al. [31], Zhu and Fukushima [55], Zhu et al. [56] and Chen et al. [12] explored robust portfolio selection problems using CVaR and LPM measures. Delage and Ye [14] and Popescu [38] studied the distributionally robust portfolio where the ambiguity set is defined through moment constraints.

The work by Goldfarb and Iyengar [19] apparently has received wide attention among others particularly in the research community of robust optimization. Instead of assuming complete information on the mean and the covariance matrix of asset returns, they introduced some types of uncertainties, such as polytopic uncertainty, box uncertainty and ellipsoidal uncertainty, in the parameters in order to determine the mean and the covariance matrix. They then transformed the problem into semi-definite programs (SDP) or second-order cone programs, which can be efficiently solved by interior-point algorithms developed in recent years. As discussed in [28, 29], the “separable” uncertainty sets typically share two common properties: (a) 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 level; and (b) they are fully or partially of box-type. The associated consequences are that the resulting robust portfolios may be too conservative, and moreover, they are usually highly non-diversified, as observed in the computational experiments conducted in [28, 29, 51].

In this paper, we concentrate on robust reward-risk ratio optimization. Kapsos et al. [24] first proposed a so-called robust omega ratio (a probability weighted ratio of gains versus losses for some threshold return target) model where in the absence of the knowledge of the true probability distribution, the worst probability distribution from an ambiguity set of distributions is considered for the omega ratio. They considered a situation where each distribution in the ambiguity set can be explicitly represented either through the mixture of some known distributions, or a perturbation from a nominal discrete distribution. Gorissen [20] studied the robust fractional programming problem where both the objective and the constraints contain some uncertain parameters and developed tractable formulation for solving the problem under some conditions. The established results were used to study the Markowitz portfolio model

when the expected return and the covariance are not fully obtainable.

Our research in this paper is to investigate a distributionally robust reward risk ratio model varied from the standard ex ante Sharpe ratio model with the ambiguity set being constructed through prior moment information. Specifically, 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. Since the reformulated optimization problem is intrinsically nonconvex, it cannot be further reformulated as a tractable convex SDP as in [52], we apply the well-known entropic approximation scheme to tackle the semi-infinite constraints and present a quantitative stability analysis of the approximation scheme under the Slater condition (Theorem 3.3). Contrary to the mainstream approaches in the literature of robust optimization, the approximation scheme does not require any specific structure (e.g. linear, convex or polynomial) of the underlying random functions w.r.t. the random variables, hence our framework (the model and numerical scheme) is applicable to a broader class of reward functions, which is a significant departure from [24].

To solve the approximated optimization problem, we propose an implicit Dinkelbach method which captures the specific structure of the problem: differing from the standard Dinkelbach method, the new method updates the optimal value at each iteration by solving a nonlinear equation. Convergence of the resulting algorithm is established (Theorem 5.2). In comparison with the work by Kapsos et al. [24], this paper has a few distinctions: (a) the model is more general in the sense that the underlying return function can be nonlinear; (b) the ambiguity set is constructed through prior moment information which complements the model in [24]; (c) the numerical scheme (Algorithm 5.1) updates the robust ratio automatically by solving a monotonic nonlinear equation rather than by a fixed increment as in [24]. Moreover, since the underlying functions are nonlinear, we solve a nonlinear convex programming problem rather than a linear programming problem at each iteration as in [24]. While this might be viewed as a disadvantage in terms of numerical efficiency, the new robust approach and numerical scheme can be easily applied to general stochastic fractional programs which have a wide range of applications in operational research and management sciences.

To see how the proposed new framework of modelling and numerical scheme works, we look into 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 [38], we take a similar approach to [14] by considering some degree of uncertainty for these quantities. We take a simple approach which restricts each component of the mean and covariance to an interval with finite lower and upper bound rather than consider the deviation of the true mean and covariance within a specified ellipsoid and positive semi-definite cone as [14], The main advantage of our approach is that it allows us to convert the distributionally robust formulation to a classical moment problem so that we can solve the latter with the general numerical schemes discussed above (Algorithm 5.1). Statistical analysis is presented for assessing the likelihood of the true probability distribution to lie in the ambiguity set (Theorem 4.4) and the convergence of the resulting optimal value and optimal solutions (Theorem 4.5). A case study on portfolio selection is carried out and the preliminary numerical test results show some promising performance of the new robust modelling and numerical framework (Section 6).

Throughout this paper, we will use the following notation. By convention, we

use $a^T b$ to denote the scalar product of two vectors $a, b \in \mathbb{R}^n$, $\|\cdot\|$ the Euclidean norm of a vector, $\|\cdot\|_\infty$ infinity norm of continuous functions defined over a compact set, and $\|\cdot\|_F$ the Frobenius norm of matrix. We write ‘clA’ for the closure of a set A , $d(x, A) := \inf_{x' \in A} \|x - x'\|$ for the distance from a point x to A , $\mathbb{D}(\mathcal{C}, \mathcal{A}) := \sup_{x \in \mathcal{C}} d(x, \mathcal{A})$, the deviation of set \mathcal{C} from \mathcal{A} and $\mathbb{H}(\mathcal{C}, \mathcal{A}) := \max(\mathbb{D}(\mathcal{C}, \mathcal{A}), \mathbb{D}(\mathcal{A}, \mathcal{C}))$ for the Hausdorff distance between \mathcal{A} and \mathcal{C} . Finally, for a sequence of subsets $\{S_k\}$ in a metric space, we follow the standard notation [40] by using $\limsup_{k \rightarrow +\infty} S_k$ to denote its outer limit, that is, $\limsup_{k \rightarrow +\infty} S_k = \{x : \liminf_{k \rightarrow +\infty} d(x, S_k) = 0\}$.

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

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

where x is a decision vector, X is a nonempty convex and compact subset of \mathbb{R}^n , $f : \mathbb{R}^n \times \mathbb{R}^k \rightarrow \mathbb{R}$ is a continuous return function and it is concave w.r.t. x for every fixed ξ , ξ is a random variable on probability space (Ξ, \mathbb{F}, P) with $\Xi \subset \mathbb{R}^k$, $Y(\xi)$ is a benchmark return, $\mathbb{E}_P[\cdot]$ denotes the expected value w.r.t. the probability distribution of ξ , 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. The model is a variation of the well known ex ante Sharpe ratio [45] where the standard deviation in the denominator is replaced by one sided deviation from zero rather than the mean value. Therefore $\mathbb{E}_P[(Y(\xi) - f(x, \xi))_+]$ is not even a standard deviation risk functional (see [34, Definition 2.21])¹. The reason why we choose single-sided dispersion measure is that we view the reward falling below the benchmark as unfavorable (a risk) and we use first order instead of higher order purely for simplicity of the model. The framework in this paper (robust formulation and numerical schemes) works for higher order single sided measure. Moreover, since the reward function f may be nonlinear, model (2.1) might have applications beyond finance such as supply chain management and energy in that the reward function may be regarded as a utility/performance function whereas $Y(\xi)$ is just benchmark such as $f(x_0, \xi)$.

Problem (2.1) might not be well defined since the denominator may turn out to be zero. In the literature of portfolio optimization, one often assumes that the risk of the active return is positive for all feasible portfolios [48]. Another difficulty is the fractional form of the objective. If $f(\cdot, \xi)$ is concave for every ξ , then the objective function is quasi-concave. In that case, one may introduce a new variable $\tau \in \mathbb{R}$ and reformulate (2.1) as follows:

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

This kind of reformulation is well known, see for instance Pinar and Tütüncü [36]. In the case when the expected one sided deviation in (2.1) is replaced by a coherent risk

¹It is not a risk measure because it does not satisfy the translation invariance property.

measure, we may use the dual representation of the risk measure and derive a similar formulation with semi-infinite constraints indexed by a set of functionals [35].

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 the true distribution during the decision making. Let \mathcal{P} denote the non-empty ambiguity set which contains the true probability distribution. We consider a distributionally robust formulation for problem (2.1) as follows:

$$(2.3) \quad \sup_{x \in X} \inf_{P \in \mathcal{P}} \frac{\mathbb{E}_P[f(x, \xi) - Y(\xi)]}{\mathbb{E}_P[(Y(\xi) - f(x, \xi))_+]}$$

In this formulation, the optimal decision is based on the the worst probability distribution from \mathcal{P} , a conservative approach to mitigate the risk arising from the ambiguity of the true distribution. Note that problem (2.3) depends on the choice of \mathcal{P} and an optimal solution of problem (2.3) provides a lower bound for the optimal value of the true problem (2.1) if the true probability distribution is contained in \mathcal{P} .

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 problem particularly when the ambiguity set is defined through moments. To deal with the issue, we consider an equivalent maximization problem with robust constraints:

$$(2.4) \quad \begin{array}{ll} \sup & \tau \\ (x, \tau) \in X \times \mathbb{R} & \\ \text{s.t.} & \inf_{P \in \mathcal{P}} \mathbb{E}_P[f(x, \xi) - Y(\xi) - \tau(Y(\xi) - f(x, \xi))_+] \geq 0. \end{array}$$

Compared to (2.3), (2.4) is relatively easier to tackle as both the objective and constraint functions are linear w.r.t. $\mathbb{E}_P[\cdot]$. The following proposition gives a theoretical guarantee on the equivalence of the two problems.

Throughout the paper, we use the convention $\frac{0}{0} = +\infty$.

PROPOSITION 2.1. *Problems (2.4) and (2.3) are equivalent when both have finite optimal value and optimal solutions, that is, if $\{\tau^*, (x^*, \tau^*)\}$ is a pair ² of optimal value, optimal solution of problem (2.4), then $\{x^*, \tau^*\}$ is a pair of the optimal value and optimal solution of (2.3), and vice versa.*

Proof. Let $\{\tau^*, (x^*, \tau^*)\}$ and $(\hat{\tau}, \hat{x})$ be a pair of optimal value and optimal solution of problems (2.4) and (2.3) respectively. It suffices to show that

$$(2.5) \quad \tau^* = \inf_{P \in \mathcal{P}} \frac{\mathbb{E}_P[f(x^*, \xi) - Y(\xi)]}{\mathbb{E}_P[(Y(\xi) - f(x^*, \xi))_+]} = \inf_{P \in \mathcal{P}} \frac{\mathbb{E}_P[f(\hat{x}, \xi) - Y(\xi)]}{\mathbb{E}_P[(Y(\xi) - f(\hat{x}, \xi))_+]} = \hat{\tau}.$$

Since (x^*, τ^*) must be a feasible solution to (2.4),

$$(2.6) \quad \inf_{P \in \mathcal{P}} \mathbb{E}_P[(f(x^*, \xi) - Y(\xi)) - \tau^*(Y(\xi) - f(x^*, \xi))_+] \geq 0.$$

In what follows, we show that (2.6) implies

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

²It is unique if there is a unique optimal solution.

Assume for the sake of a contradiction that (2.7) fails to hold. Then there exists a small positive number ϵ and a sequence $\{P^N\} \subset \mathcal{P}$ such that

$$(2.8) \quad \frac{\mathbb{E}_{P^N}[f(x^*, \xi) - Y(\xi)]}{\mathbb{E}_{P^N}[(Y(\xi) - f(x^*, \xi))_+]} < \tau^* - \epsilon$$

when N is sufficiently large. For a fixed P^N , we have

$$\mathbb{E}_{P^N}[(f(x^*, \xi) - Y(\xi)) - \tau^*(Y(\xi) - f(x^*, \xi))_+] \leq -\epsilon \mathbb{E}_{P^N}[(Y(\xi) - f(x^*, \xi))_+].$$

If $\mathbb{E}_{P^N}[(Y(\xi) - f(x^*, \xi))_+] > 0$, then the inequality above contradicts (2.6). On the other hand, if $\mathbb{E}_{P^N}[(Y(\xi) - f(x^*, \xi))_+] = 0$, then $\mathbb{E}_{P^N}[f(x^*, \xi) - Y(\xi)] \geq 0$, which entails the left hand side of (2.8) to be positive infinity. This is not possible because of the boundedness of τ^* . Thus we get a contradiction as desired.

Since $x^* \in X$ is a feasible solution to (2.3) whereas \hat{x} is an optimal solution of (2.3), then

$$(2.9) \quad \hat{\tau} \geq \inf_{P \in \mathcal{P}} \frac{\mathbb{E}_P[f(x^*, \xi) - Y(\xi)]}{\mathbb{E}_P[(Y(\xi) - f(x^*, \xi))_+]} \geq \tau^*.$$

On the other hand, as \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

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

which means $(\hat{x}, \hat{\tau})$ is a feasible solution of problem (2.4). Thus

$$(2.10) \quad \tau^* \geq \inf_{P \in \mathcal{P}} \frac{\mathbb{E}_P[f(\hat{x}, \xi) - Y(\xi)]}{\mathbb{E}_P[(Y(\xi) - f(\hat{x}, \xi))_+]} \geq \hat{\tau}.$$

Combining (2.9) and (2.10), we arrive at (2.5). The proof is complete. \square

Proposition 2.1 paves the way for us to develop a complete numerical treatment of robust reward-risk ratio optimization problem (2.3) via (2.4). It is unclear whether we will be able to derive a similar result in the case when the expected one sided deviation in (2.3) is replaced by a coherent risk measure. To ensure the optimal value and the optimal solutions of (2.4) to be bounded, we make the following assumption.

ASSUMPTION 2.1. *Assume that*

- (a) X is a compact convex set;
- (b) $\Xi \subset \mathbb{R}^k$ is a compact set;
- (c) $f(\cdot, \cdot)$ and $Y(\cdot)$ are continuous;
- (d) there exists a positive number ϵ such that

$$(2.11) \quad \min_{x \in X} \inf_{P \in \mathcal{P}} \mathbb{E}_P[(Y(\xi) - f(x, \xi))_+] \geq \epsilon.$$

Parts (a)-(c) of Assumption 2.1 are standard, see similar assumptions by Du-
pačová [15] for distributionally robust minimax problems. Part (d) provides a suffi-
cient condition for the well-definedness of the robust formulation (2.4) (or (2.3)). It

is possible to relax conditions (a) and (b) by allowing X and Ξ to be unbounded. However this will entail a lot of technical details which might detract us from the main focus of the paper. We leave them to interested readers.

PROPOSITION 2.2. *Under Assumption 2.1, problem (2.4) has a finite optimal value.*

Proof. Condition (2.11) ensures that for all $P \in \mathcal{P}$ and all $x \in X$

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

Under Assumption 2.1, Ξ and X are 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 of problems (2.3) and (2.4) as we have shown in Proposition 2.1. \square

For the convenience of exposition, let us rewrite (2.4) as a minimization problem

$$(2.12) \quad \begin{array}{ll} \inf_{(x, \tau) \in X \times \mathbb{R}} & -\tau \\ \text{s.t.} & \sup_{P \in \mathcal{P}} -\mathbb{E}_P[(f(x, \xi) - Y(\xi)) - \tau(Y(\xi) - f(x, \xi))_+] \leq 0. \end{array}$$

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 (2.12) in a concise form

$$(2.13) \quad \begin{array}{ll} \inf_{(x, \tau) \in X \times \mathbb{R}} & -\tau \\ \text{s.t.} & \sup_{P \in \mathcal{P}} \mathbb{E}_P[G(x, \xi, \tau)] \leq 0. \end{array}$$

Observe that problem (2.13) is non-convex as G is not convex in (x, τ) . However, for fixed τ , G is convex in x . 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 assertions 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*

$$(2.14) \quad \sup_{P \in \mathcal{P}} \mathbb{E}_P[G(x, \xi, \tau^*)] \geq 0, \forall x \in X;$$

- (iii) *if for the given τ^* satisfying (2.14), there exists $x^* \in X$ such that the equality in (2.14) holds, that is, $\sup_{P \in \mathcal{P}} \mathbb{E}_P[G(x^*, \xi, \tau^*)] = 0$, then $-\tau^*$ is the optimal value and (x^*, τ^*) is the optimal solution of problem (2.13).*

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). By Assumption 2.1 (a)-(c), X and Ξ are compact sets and $-f(x, \xi) + Y(\xi)$ is continuous in (x, ξ) , therefore

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

On the other hand, under Assumption 2.1 (d), it follows by (2.11)

$$(2.15) \quad \sup_{P \in \mathcal{P}} \mathbb{E}_P[\tau(Y(\xi) - f(x, \xi))_+] \geq \tau\epsilon.$$

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

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

Part (iii). The equality of (2.14) implies feasibility of (x^*, τ^*) to problem (2.13). Moreover, by using (2.15), we can easily show that $\sup_{P \in \mathcal{P}} \mathbb{E}_P[G(x, \xi, \tau)]$ is strictly monotonically increasing. By Part (i) and (2.14), for any $x \in X$ and any $\delta > 0$

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

which implies that the optimal value of problem (2.13) cannot be smaller than $-\tau^*$. The conclusion follows. \square

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 (2.13) and the sufficient condition for optimality.

3. Dual formulation and entropic approximation. Problem (2.13) 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 random parameters, there are various ways to define the distributional set \mathcal{P} , for instances Zhu and Fukushima [55] 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. [57] 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 [47] in the nineteenth century. The problem of moments is related to optimization over polynomials (the dual theory of moment). For instance, Lasserre [26] and Parrilo [33] 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 problem (2.13) where the ambiguity set \mathcal{P} is constructed as follows:

$$(3.1) \quad \mathcal{P} := \left\{ P \in \mathcal{D} : \begin{array}{l} \mathbb{E}_P[\psi_i(\xi)] = 0, \quad i = 1, \dots, p \\ \mathbb{E}_P[\psi_i(\xi)] \leq 0, \quad i = p + 1, \dots, q \end{array} \right\},$$

where $\psi_i : \Xi \rightarrow \mathbb{R}$, $i = 1, \dots, q$, are continuous functions, and \mathcal{D} denotes the set of probability measures on probability space (Ξ, \mathbb{F}) of random variable ξ . Note that

in this setup, we restrict $\psi_i(\xi)$ 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 a correlation matrix.

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

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

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

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

The assumption is similar to the Slater constraint qualification in the literature. It means that ψ_i cannot be selected arbitrarily. It is a standard condition for deriving Lagrange dual of moment problems, see Shapiro [43] and Xu et al. [54, Proposition 2.1] for some detailed discussions on the condition. We should also note that Assumption 3.1 is a sufficient condition rather than a necessary condition, see [54, Example 2.8] for a counter-example where Assumption 3.1 fails but there is no dual gap. Under Assumption 3.1, we can reformulate problem (2.12), through [43, Proposition 3.4], as the following semi-infinite programming problem

$$(3.2) \quad \begin{aligned} & \min_{x \in X, \tau \in \mathbb{R}, \lambda \in \mathbb{R}^p \times \mathbb{R}_+^{q-p}} && -\tau \\ & \text{s.t.} && \sum_{i=1}^q \lambda_i \psi_i(\xi) \geq G(x, \xi, \tau), \quad \forall \xi \in \Xi. \end{aligned}$$

If, in addition, Assumption 2.1 holds, then the optimal value of the primal and the dual problems is finite. Moreover, by [43, Proposition 3.4], the set of optimal solutions is nonempty and bounded.

Following the mainstream work in DRO (see e.g. Wiesemann et al. [52] and references therein), it might be possible to reformulate the semi-infinite constraints of problem (3.2) into some semi-definite constraints when ψ_i and Ξ have certain specific structure. However, since $G(\cdot)$ is non-convex in (x, τ) even in the simplest case where f is a linear function of x , the constraints are intrinsically non-convex and this discourages us to adopt an SDP reformulation approach because the end optimization problem would be non-tractable after all. This motivates us to propose an approximation scheme for handling the semi-infinite constraints. Compared to the work [24], our approach will not require $f(x, \xi)$ to be linear function or ξ to be a discrete random variable, which means it is applicable to a broader class of problems including nonlinear portfolio optimization problems where both stocks and options are involved [58].

For the convenience of exposition, let

$$(3.3) \quad R(x, \tau, \lambda, \xi) := G(x, \xi, \tau) - \sum_{i=1}^q \lambda_i \psi_i(\xi)$$

and

$$(3.4) \quad W := X \times \mathbb{R} \times \mathbb{R}^p \times \mathbb{R}_+^{q-p} \text{ and } w := (x, \tau, \lambda) \in W.$$

The semi-infinite constraints in (3.2) can be written as

$$(3.5) \quad \sup_{\xi \in \Xi} R(w, \xi) \leq 0.$$

Of course, writing the semi-infinite constraints as above does not bring us any computational convenience. Our plan here is to develop an approximation of $\sup_{\xi \in \Xi} R(w, \xi)$ through entropic risk measure so that the constraint can be handled numerically more conveniently. Anderson et al. [1] seem to be the first to consider a CVaR approximation. The rationale behind their approximation scheme is that one may regard $\sup_{\xi \in \Xi} R(w, \xi)$ as an extremely robust risk measure of the random variable $R(w, \xi)$ whereas CVaR of $R(w, \xi)$ is a kind of deviation from the extremum conservatism. In the context of (3.5), the deviation may allow one to capture more samples at the tail when it is plugged into sample average approximation scheme without significantly affecting the set of solutions (set of feasible solution in a relevant optimization problem) and consequently “smooth up” or “stabilize” the numerical solution of the optimization problem. Here we take a similar initiative but adopt entropic risk measure approximation in that the latter is a smooth function and is suited for a broader class of optimization problems.

For a random variable $Z \in L^\infty(\Xi, \mathbb{F}, P)$ ³, the entropic risk measure is defined as

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

where γ is a positive number. It is well known (see e.g. [17, formulation (4)]) that $e_\gamma(Z)$ is monotonically increasing in γ and

$$\lim_{\gamma \rightarrow +\infty} e_\gamma(Z) = \text{ess sup}(-Z),$$

where ess sup denotes essential supremum of the random variable, see [17, 18] for a thorough treatment of the subject. The following lemma states the uniform approximation of entropic risk measure for a general class of random functions. It will pave the way for stability analysis of the optimal solutions of the approximation problem in the forthcoming Theorem 3.3.

LEMMA 3.1. (*Entropic approximation of a random function, [27, Proposition 2.1]*) *Let $h : \mathbb{R}^n \times \mathbb{R}^k \rightarrow \mathbb{R}$ be a continuous function and X be a subset of \mathbb{R}^n . Let ξ be a random variable on the probability space (Ξ, \mathbb{F}, P) with support set $\Xi \subset \mathbb{R}^k$. Let $H(x)$, $F_x(\cdot)$ and Ξ_x denote respectively the essential supremum, the cumulative distribution function and the support set of $-h(x, \xi)$, that is, Ξ_x is the smallest set that satisfies $P(-h(x, \xi) \in \Xi_x) = 1$. Let $\text{Diam}(\Xi_x)$ denote the diameter of 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 $x \in X$, $\lim_{\gamma \rightarrow +\infty} e_\gamma(h(x, \xi)) = H(x)$. 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*

$$F_x(H(x) - \epsilon) \leq 1 - \delta(\epsilon), \quad \forall x \in X_\epsilon,$$

where $X_\epsilon := \{x \in X : \text{Diam}(\Xi_x) > 2\epsilon\}$. Then

$$(3.6) \quad |e_\gamma(h(x, \xi)) - H(x)| < 2\epsilon + \left| \frac{1}{\gamma} \ln \delta(\epsilon) \right|.$$

Using Lemma 3.1, we propose to approximate the inequality system (3.5) by

$$(3.7) \quad e_\gamma(-R(w, \xi)) := \frac{1}{\gamma} \ln \mathbb{E}[e^{\gamma R(w, \xi)}] \leq 0.$$

³ $L^\infty(\Xi, \mathbb{F}, P)$ denotes the set of essentially bounded measurable functions.

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 (3.7) 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 practice, there are two simple ways to choose the probability distribution in (3.7). One is to choose uniform distribution if Ξ is bounded, the other is to use the true distribution of ξ if the samples of the random variables are obtainable. In the latter approach, we may use the samples to compute sample average approximation (SAA) of the expected value in $e_\gamma(\cdot)$. In that case, the value of parameter γ may be estimated by the sample size and the tolerance. For example, if there are 100 samples " Z_1, \dots, Z_{100} " in decreasing order, then

$$e_\gamma^{100}(-Z) - Z_1 = \frac{1}{\gamma} \ln(1 + e^{\gamma(Z_2 - Z_1)} + \dots + e^{\gamma(Z_{100} - Z_1)}) - \frac{1}{\gamma} \ln 100.$$

Since $Z_i \leq Z_1$ for $i = 2, \dots, 100$, the quantity in the logarithm is less than 100. Thus

$$\begin{aligned} |e_\gamma^{100}(-Z) - Z_1| &= Z_1 - e_\gamma^{100}(-Z) \\ &= \frac{1}{\gamma} \ln 100 - \frac{1}{\gamma} \ln(1 + e^{\gamma(Z_2 - Z_1)} + \dots + e^{\gamma(Z_{100} - Z_1)}) \\ &\leq \frac{1}{\gamma} \ln 100. \end{aligned}$$

If the precision is ϵ , then we may set $\gamma \geq \frac{\ln 100}{\epsilon}$. Of course, the choice of the distribution will affect the rate of convergence of the entropic risk measure to its essential supremum, we explain this in the following example.

EXAMPLE 3.2. Consider $h(x, \xi) = -x\xi$, where $x \in [0, 1] \subset \mathbb{R}$ and (a) ξ follows a uniform distribution over interval $[-1, 1]$, (b) ξ follows a distribution with density

$$g(\xi) = \begin{cases} -\xi, & \text{for } \xi \in [-1, 0], \\ \xi, & \text{for } \xi \in (0, 1]. \end{cases}$$

Let $\epsilon < 1$ be a small positive number. Then $H(x) = x$, $\Xi_x = [-x, x]$ and $X_\epsilon = \{x \in [0, 1] : 2\epsilon < 2x\} = (\epsilon, 1]$. For case (a), it is easy to derive that

$$1 - F_x(H(x) - \epsilon) = \frac{\epsilon}{2x} \geq \frac{\epsilon}{2}, \quad \forall x \in X_\epsilon.$$

This shows that condition (3.6) holds with $\delta(\epsilon) = \frac{\epsilon}{2}$ and hence $|e_\gamma(h(x, \xi)) - H(x)| \leq 2\epsilon + \left| \frac{1}{\gamma} \ln \frac{\epsilon}{2} \right|$. For case (b),

$$1 - F_x(H(x) - \epsilon) \geq \frac{2\epsilon - \epsilon^2}{2}, \quad \forall x \in X_\epsilon.$$

and $|e_\gamma(h(x, \xi)) - H(x)| \leq 2\epsilon + \left| \frac{1}{\gamma} \ln \frac{2\epsilon - \epsilon^2}{2} \right|$. Observe that $\left| \frac{1}{\gamma} \ln \frac{\epsilon}{2} \right| > \left| \frac{1}{\gamma} \ln \frac{2\epsilon - \epsilon^2}{2} \right|$. Thus case (b) provides a tighter bound for the uniform entropic approximation.

With (3.7), we can construct an approximation of problem (3.2) by

$$(3.8) \quad \begin{aligned} &\min_{x \in X, \tau \in \mathbb{R}, \lambda \in \mathbb{R}^p \times \mathbb{R}_+^{q-p}} && -\tau \\ &\text{s.t.} && e_\gamma(-R(x, \lambda, \tau, \xi)) \leq 0. \end{aligned}$$

This is a one stage stochastic minimization problem with a single stochastic constraint. One can easily apply the well known SAA methods for solving the problem. The convergence results are well documented, see for instance [42] for the convergence of optimal values when SAA method is applied to general stochastic optimization problem and Anderson et al. [1] for a similar scheme to (3.8) with CVaR approximation.

Recall that we have explained immediately after formulation (3.2) that the set of the optimal solutions to problem (3.2) is nonempty and bounded under Assumption 3.1. Thus, we may restrict the variables λ_i , $i = 1, \dots, q$ in problem (3.8) to take finite values. Specifically, we assume that there exists a positive constant C_0 such that

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

Note also that problem (3.8) is non-convex in that R is a non-convex 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 computational schemes for the problem. Here, we concentrate on approximation of (3.8) to (3.2) 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 3.2. *Let \mathcal{F} and $\mathcal{F}(\gamma)$ denote the feasible sets of (3.2) and (3.8) respectively, let ϑ^* and $\vartheta(\gamma)$ denote the corresponding optimal values⁴. 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 non-decreasing and $\vartheta(\gamma) \leq \vartheta^*$;
- (iv) if, in addition, (a) Assumptions 2.1 and 3.1 hold, (b) there exists a positive number ϵ such that⁵

$$(3.9) \quad \min_{x \in X} \mathbb{E}[(Y(\xi) - f(x, \xi))_+] \geq \epsilon,$$

then both ϑ^* and $\vartheta(\gamma)$ are finite, and the set of optimal solutions to (3.8) is nonempty and bounded.

Proof. Part (i). Compare constraints of problems (3.2) and (3.8), the only difference is the second constraint. Since $e_\gamma(-R(w, \xi)) \leq \sup_{\xi \in \Xi} R(w, \xi)$, then $\mathcal{F} \subset \mathcal{F}(\gamma)$. Part (ii). The monotonicity follows from the fact that for any fixed w , $e_\gamma(-R(w, \xi))$ increases in 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 $\mathbb{E}[e^{\gamma R(x, \lambda, \tau, \xi)}] \geq e^{\gamma \mathbb{E}[R(x, \lambda, \tau, \xi)]}$. Through the definition of R in (3.3), 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, it follows by (3.9),

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

⁴By 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 .

⁵(3.9) is similar to condition (d) of Assumption 2.1. If $\mathbb{E}[\cdot]$ is taking w.r.t a probability P contained in \mathcal{P} , this condition is implied by Assumption 2.1.

Furthermore, since $|\lambda_i| \leq C_0$ and $\psi_i(\cdot)$ is continuous on Ξ , for $i = 1, \dots, q$, then the inequality above means $\mathbb{E}[R(x, \lambda, \tau, \xi)] \rightarrow +\infty$ as $\tau \rightarrow +\infty$, and hence

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

The discussions above show that a large τ value would violate the constraint of (3.2) regardless of the value of γ , which means τ must be bounded at its optimum.

The non-emptiness and boundedness of the set of optimal solutions to (3.8) 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. \square

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 3.3. *There exist positive constants C and δ such that*

$$(3.10) \quad d(w, \mathcal{F}) \leq C \left(\sup_{\xi \in \Xi} R(w, \xi) \right)_+$$

for any $w \in W$ satisfying $d(w, \mathcal{F}) \leq \delta$. Note that for $w \in \mathcal{F}$, $\sup_{\xi \in \Xi} R(w, \xi) \leq 0$. Thus (3.10) holds trivially. The quantity $(\sup_{\xi \in \Xi} R(w, \xi))_+$ describes the significance of constraint violation when $w \notin \mathcal{F}$, and (3.10) establishes a relationship between deviation of w from \mathcal{F} and the quantity $(\sup_{\xi \in \Xi} R(w, \xi))_+$. In the literature of parametric programming, the relationship is known as error bound condition which plays an important role in stability analysis of the optimal value and optimal solutions, we refer interested readers the survey papers [3, 32] for more details in this regard. A sufficient condition for (3.10) is that for every $w \in \mathcal{F}$, the constraint qualification due to Borwein [10] holds, that is, for any $\alpha \in [0, +\infty)$, $0 \in \alpha \partial(\sup_{\xi \in \Xi} R(w, \xi)) + \mathcal{N}_{\mathcal{F}}(w)$ implies $\alpha = 0$, where ∂ denotes the Clarke subdifferential in w , and $\mathcal{N}_{\mathcal{F}}(w)$ denotes the normal cone, see [10, Theorem 3.2] for details.

We are now ready to discuss the stability of problem (3.8) against variation of the parameter γ .

THEOREM 3.3. *Assume: (a) \mathcal{F} is a compact set; (b) ψ_i , $i = 1, \dots, q$, is continuous, (c) 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*

$$\mathbb{H}(\mathcal{F}(\gamma), \mathcal{F}) \leq \epsilon, \forall \gamma \in [\gamma^*, +\infty);$$

(ii) *if, in addition, Assumption 3.3 holds, then there exists positive constants C and γ^* such that*

$$(3.11) \quad \mathbb{H}(\mathcal{F}(\gamma), \mathcal{F}) \leq C \sup_{w \in W} \Delta_\gamma(w), \forall \gamma \in [\gamma^*, +\infty),$$

where

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

(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.,

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

Proof. By Proposition 3.2 (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). The proof is similar to that of [53, Lemma 4.2 (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 (c), 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 3.3, 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 $w \in \mathcal{F}(\gamma)$,

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

where the second inequality follows from the fact $(a)_+ \leq (a - b)_+$ for any $b \leq 0$; the equality follows from the fact $\sup_{\xi \in \Xi} R(w, \xi) \geq e_\gamma(-R(w, \xi))$ for any $w \in W$. 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 [25, Theorem 1]. Here we include a proof for completeness. Let w^* and w_γ be an optimal solution of problems (3.2) and (3.8) respectively. Let τ^* and τ_γ be the

corresponding second 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. \square

Under condition (c), it follows by Lemma 3.1 that $\Delta_\gamma(w)$ goes to 0 uniformly for all $w \in W$ as $\gamma \rightarrow +\infty$. Theorem 3.3 states that $|\vartheta(\gamma) - \vartheta^*| \rightarrow 0$ 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 [14] 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 in [14]). 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:

$$(4.1) \quad \mathcal{P}(\bar{\mu}, \bar{\Sigma}, \gamma_1, \gamma_2) := \left\{ P \in \mathcal{P} : \begin{array}{l} \mathbb{E}_P[\xi - \bar{\mu}]^T \bar{\Sigma}^{-1} \mathbb{E}_P[\xi - \bar{\mu}] \leq \gamma_1 \\ \mathbb{E}_P[(\xi - \bar{\mu})(\xi - \bar{\mu})^T] \preceq \gamma_2 \bar{\Sigma} \end{array} \right\},$$

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 is particularly helpful when $\bar{\mu}$ and $\bar{\Sigma}$ are estimated through empirical data.

The first constraint in (4.1) 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 semi-definite cone defined with matrix inequality. A significant advantage of this particular way to construct the ambiguity set is that through a simple duality formulation, the resulting distributionally robust optimization problem can be converted into a tractable convex semi-definite program.

In this section, we consider a variation of (4.1). Instead of using ellipsoid constraints for the mean and semi-definite constraints for the covariance, we propose simple box constraints for each component of these quantities. Specifically we define the ambiguity set as follows:

$$(4.2) \quad \mathcal{P}^* := \left\{ P \in \mathcal{P} : \begin{array}{l} -\epsilon \leq (\mathbb{E}_P[\xi] - \bar{\mu})_i \leq \epsilon, \quad i = 1, \dots, m \\ \|\mathbb{E}_P[(\xi - \bar{\mu})(\xi - \bar{\mu})^T] - \bar{\Sigma}\|_* \leq \sigma \end{array} \right\},$$

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 [51] apparently were

among the first to consider box constraints of the mean and covariance to define a set of means and covariance matrices in portfolio optimization. They developed a minimax robust Markowitz's mean-variance model where the worst mean and covariance matrix are selected from the set.

Our motivations here for this particular way of construction of the ambiguity set are three-fold: (a) the semi-definite 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 (4.2) as

$$(4.3) \quad \mathcal{P}^* := \left\{ P \in \mathcal{P} : \begin{array}{l} -\epsilon \leq \mathbb{E}_P[\psi_i(\xi)] \leq \epsilon, \quad i = 1, \dots, k \\ -\sigma \leq \mathbb{E}_P[\psi_j(\xi)] \leq \sigma, \quad j = k+1, \dots, q \end{array} \right\},$$

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

Analogous to the arguments in [14], in practice we may use samples to construct an estimate of the true mean and covariance. Let ξ^1, \dots, ξ^N be an independent and identically distributed sample of ξ and

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

Then we may consider the following sample based ambiguity set

$$(4.4) \quad \mathcal{P}^N := \left\{ P \in \mathcal{P} : \begin{array}{l} -\epsilon \leq \mathbb{E}_P[\xi] - \mu^N \leq \epsilon, \\ \|\mathbb{E}_P[(\xi - \mu^N)(\xi - \mu^N)^T] - \Sigma^N\|_* \leq \sigma \end{array} \right\},$$

or equivalently

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

where k , q are defined as above, $\psi_i^N(\xi) = \xi_i - \mu_i^N$ and $\psi_j^N(\xi)$, $j = k+1, \dots, q$, 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. 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 convergence analysis or stability analysis of the distributionally robust reward risk ratio problem w.r.t. change of sample data.

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

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

where

$$\mathcal{M} := \{h : \mathbb{R}^k \rightarrow \mathbb{R} \mid h \text{ is } \mathcal{B} \text{ measurable, } \sup_{\xi \in \Xi} |h(\xi)| \leq 1\}.$$

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 \mathcal{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 \mathcal{P} .

PROPOSITION 4.2. *Suppose that Ξ is a compact set. Then $\mathbb{H}_{TV}(\mathcal{P}^N, \mathcal{P}^*) \rightarrow 0$ as $N \rightarrow +\infty$.*

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 [49, Lemma 2] that there exists a positive constant L such that

$$\begin{aligned} \mathbb{D}_{TV}(Q, \mathcal{P}^*) &\leq L (\|(\mathbb{E}_Q[\Psi_I(\xi)] - \epsilon e)_+\| + \|(-\mathbb{E}_Q[\Psi_I(\xi)] - \epsilon e)_+\|) \\ &\quad + L (\|(\mathbb{E}_Q[\Psi_J(\xi)] - \sigma e)_+\| + \|(-\mathbb{E}_Q[\Psi_J(\xi)] - \sigma e)_+\|) \\ (4.5) \quad &\leq L (\|(\mathbb{E}_Q[\Psi_I(\xi) - \Psi_I^N(\xi)]_+\| + \|(\mathbb{E}_Q[\Psi_J(\xi) - \Psi_J^N(\xi)]_+\|) \\ &\quad + L (\|(\mathbb{E}_Q[\Psi_I^N(\xi) - \Psi_I(\xi)]_+\| + \|(\mathbb{E}_Q[\Psi_J^N(\xi) - \Psi_J(\xi)]_+\|), \end{aligned}$$

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 denotes the vector with each component being 1 and for a vector a , $(a)_+ := \max(0, a)$ with the maximum being taken componentwise. The second inequality in (4.5) follows from the fact that $P \in \mathcal{P}^N$ and $(a)_+ \leq (a - b)_+$ if $b \leq 0$. Likewise, for any given $Q \in \mathcal{P}^*$,

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

Since $\mu^N \rightarrow \bar{\mu}$ and $\Sigma^N \rightarrow \bar{\Sigma}$, 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 $N \rightarrow +\infty$. Thus both $\mathbb{D}_{TV}(Q, \mathcal{P}^*)$ and $\mathbb{D}_{TV}(Q, \mathcal{P}^N)$ converge to zero as N goes to infinity, which implies $\mathbb{H}_{TV}(\mathcal{P}^N, \mathcal{P}^*) \rightarrow 0$. The proof is complete. \square

Next, we address question (b). We need some intermediate technical results which can be easily established by Shawe-Taylor and Cristianini's theorems (see [46]).

LEMMA 4.3. *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 the samples of ξ ,*

$$\|\mathbb{E}[\xi] - \mu^N\|_* \leq \frac{\rho}{\sqrt{N}} \left(2 + \sqrt{2 \ln \frac{1}{\delta}} \right),$$

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

$$\|\mathbb{E}[(\xi - \mu^N)(\xi - \mu^N)^T] - \Sigma^N\|_* \leq \frac{3\rho^2}{\sqrt{N}} \left(2 + \sqrt{2 \ln \frac{1}{\delta}} \right).$$

Proof. Part (i) follows straightforwardly from [46, Theorem 3]. We only prove Part (ii). Let

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

Then

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

Note that for any two events A_1 and A_2 , Bonferroni's inequality states that

$$P(A_1 A_2) \geq P(A_1) + P(A_2) - 1.$$

Consequently, we can use [46, Corollary 5] and Part (i) to claim that with probability at least $1 - 2\delta$

$$\begin{aligned} \|\mathbb{E}[(\xi - \mu^N)(\xi - \mu^N)^T] - \Sigma^N\|_* &\leq \frac{\rho^2}{\sqrt{N}} \left(2 + \sqrt{2 \ln \frac{1}{\delta}} \right) + \frac{2\rho^2}{\sqrt{N}} \left(2 + \sqrt{2 \ln \frac{1}{\delta}} \right) \\ &= \frac{3\rho^2}{\sqrt{N}} \left(2 + \sqrt{2 \ln \frac{1}{\delta}} \right). \end{aligned}$$

The proof is complete. \square

With Lemmas 4.3, 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 $1 - 3\delta$ for the given small positive number δ .

THEOREM 4.4. *Suppose that ξ is essentially bounded by a positive number ρ and the parameters ϵ and σ are chosen as follows:*

$$(4.6) \quad \epsilon_N := \frac{\rho}{\sqrt{N}} \left(2 + \sqrt{2 \ln \frac{1}{\delta}} \right), \quad \sigma_N := \frac{3\rho^2}{\sqrt{N}} \left(2 + \sqrt{2 \ln \frac{1}{\delta}} \right).$$

Then with probability at least $1 - 3\delta$ over the choice of the samples of ξ , the true distribution of ξ lies in the ambiguity set \mathcal{P}^N .

Proof. Theorem 4.4 follows from Bonferroni's inequality and Lemma 4.3 directly.

□

Note that Shawe-Taylor and Cristianini presented a similar result when the matrix norm is replaced by the Frobenius norm. Theorem 4.4 is an analogue of their result under the new matrix norm $\|\cdot\|_*$.

Using the ambiguity set \mathcal{P}^N , we can derive the dual formulation of problem (2.12) coupled by the entropic approximation as follows:

$$(4.7) \quad \begin{aligned} \min_{x \in X, \tau \in \mathbb{R}, \lambda \in \mathbb{R}^p \times \mathbb{R}_+^{q-p}} \quad & -\tau \\ \text{s.t.} \quad & e_\gamma(-R^N(x, \lambda, \tau, \xi)) \leq 0, \end{aligned}$$

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

Let $\mathcal{F}, \mathcal{F}(\gamma), \vartheta^*$ and $\vartheta(\gamma)$ be defined as in Proposition 3.2. Let $\mathcal{F}^N(\gamma)$ and $\vartheta^N(\gamma)$ denote the set of feasible solutions and the optimal value of problem (4.7), $S^N(\gamma)$, $S(\gamma)$ and S^* denote the optimal solution of problems (4.7), (3.8) and (3.2) respectively. Let $\mathcal{F}^s(\gamma)$ denote the set of strictly feasible solutions of problem (3.8). The following theorem summarizes the convergence of problem (4.7) to problems (3.8) and (3.2) in terms of the optimal value and the optimal solutions for fixed γ as $N \rightarrow +\infty$.

THEOREM 4.5. *Assume: (a) D is a compact set and $D \cap S(\gamma) \neq \emptyset$, $D \cap S^N(\gamma) \neq \emptyset$; (b) $\text{cl}\mathcal{F}^s(\gamma) \cap S(\gamma) \neq \emptyset$. Then*

- (i) $\limsup_{N \rightarrow +\infty} S^N(\gamma) \cap D \subseteq S(\gamma) \cap D$ and $\lim_{N \rightarrow +\infty} \vartheta^N(\gamma) = \vartheta(\gamma)$.
- (ii) if, in addition, Assumption 3.3 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

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

Condition (a) requires the set of optimal solutions of problems (4.7) and (3.8) to have at least one finite optimal solution. Condition (b) means that if problem (3.8) has at least one optimal solution which falls into the closure of the interior of its feasible set, the condition is fulfilled by the Slater constraint qualification, see similar conditions to be used in [27].

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 $\limsup_{N \rightarrow +\infty} \mathcal{F}^N(\gamma) \subseteq \mathcal{F}(\gamma)$. By taking a subsequence if necessary we may assume for the simplicity of notation that $(x^N, \tau^N, \lambda^N) \rightarrow (x^*, \tau^*, \lambda^*) \in \mathcal{F}(\gamma)$, which implies

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

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

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

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

Part (ii). By Part (i) and Theorem 3.3, 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 3.3. For any $w \in \mathcal{F}^N(\gamma)$, it follows by (3.10)

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

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 (3.2) and problem (4.7) 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. \square

Compared to Theorem 3.3, Theorem 4.5 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 $\Delta_\gamma^N(w) \rightarrow 0$ as $N, \gamma \rightarrow +\infty$. 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 by Lemma 3.1 and the second term tends to zero as $N \rightarrow +\infty$ by [42, Proposition 7, Chapter 6].

5. Implicit Dinkelbach method. In this section, we propose an iterative scheme for solving problem (3.8). As we discussed in Section 2, the problem is non-convex. 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 (3.8) in a slightly neater form

$$(5.1) \quad \begin{array}{ll} \min_{x \in X, \tau \in \mathbb{R}, \lambda \in \Lambda} & -\tau \\ \text{s.t.} & e_\gamma(-R(x, \lambda, \tau, \xi)) \leq 0, \end{array}$$

where $\Lambda := \mathbb{R}^p \times \mathbb{R}_+^{q-p}$. The following algorithm presents an iterative scheme for solving (5.1).

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:

$$(5.2) \quad e_\gamma(-R(x_k, \lambda_k, \tau, \xi)) = 0.$$

Step 3. For given τ_k , solve

$$(5.3) \quad \begin{array}{ll} \min & e_\gamma(-R(x, \lambda, \tau_k, \xi)) \\ \text{s.t.} & x \in X, \\ & \lambda \in \Lambda, \end{array}$$

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.

REMARK 5.2. Note that for fixed x and λ , $e_\gamma(-R(x, \lambda, \tau, \xi))$ is monotonically increasing in τ . This allows use to use the bisection method for finding the root of equation (5.2). Moreover, for fixed τ , (5.3) is a convex optimization problem.

We call the algorithm *implicit Dinkelbach function method* as the computational scheme resembles the algorithmic procedures of the Dinkelbach method in Steps 2 and 3 where we update τ_k and (x_k, λ_k) . The word *implicit* is used to reflect the fact that τ_k is embedded nonlinearly in the equation. There are two main differences from the standard Dinkelbach method: (a) the objective function in the minimization problem (5.3) is nonlinear in parameter τ or in (x, λ) , (b) τ_k is determined in Step 2 by solving a nonlinear equation.

Kapsos et al. [24, Algorithm 1] proposed a similar algorithm for solving a robust omega ratio problem where the uncertainty in probability is characterized through

mixture distributions. Different from Algorithm 5.1, their algorithm updates the ratio τ_k by a fixed increment and then solve a tractable LP. In our algorithm, even if f is linear in x and ξ , we would still have to solve a nonlinear convex program (5.3).

PROPOSITION 5.1. Let τ_k be generated by Algorithm 5.1 and $\Delta(\tau)$ be defined as the optimal value of program (5.3). Then

- (i) $\Delta(\tau)$ is continuous in τ ;
- (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 in (x, λ) . By [40, Theorem 1.17], the optimal value function is continuous in 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 $R(x, \lambda, \tau, \xi)$ does, see Proposition 2.3 and (3.3). Therefore (5.2) has a unique solution. Moreover, τ_k is the optimal value of problem (5.1) 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 (5.1) 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 (5.1). \square

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 (5.1).

THEOREM 5.2. 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 (5.1).

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 Proposition 5.1, 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 (5.3) 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 (5.3) 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 $\tau_k \rightarrow \tau^*$ 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

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

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

6. Numerical tests. We have carried out a number of numerical experiments on the robust ratio portfolio optimization model (2.1) by evaluating its performances under various scenarios. The experiments are based on the application of the model in portfolio optimization where certain assets are invested in stock markets. Specifically, we use 10 stocks (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 1000 records of historical stock returns (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

$$f(x, \xi) := r_1x_1 + r_2x_2 + \cdots + r_{10}x_{10},$$

where r_i denotes the random return of stock i .

TABLE 1
Daily return

Strategy	L	H	A	Down	Up
Our DRO Model	0.9717	1.0541	1.005	235	265
Popescu's DRO model	0.9666	1.062	1.005	233	267
SP model	0.9532	1.0664	1.003	235	265
EW model (benchmark)	0.9717	1.0541	1.004	241	259

In implementing the proposed robust ratio portfolio optimization model (2.1) and the subsequent computational scheme, we use the ambiguity set defined in (4.4). The parameter δ for the confidence interval in Theorem 4.4 is fixed at 0.01 which means that with 97% the true probability is contained in the ambiguity set. We can figure

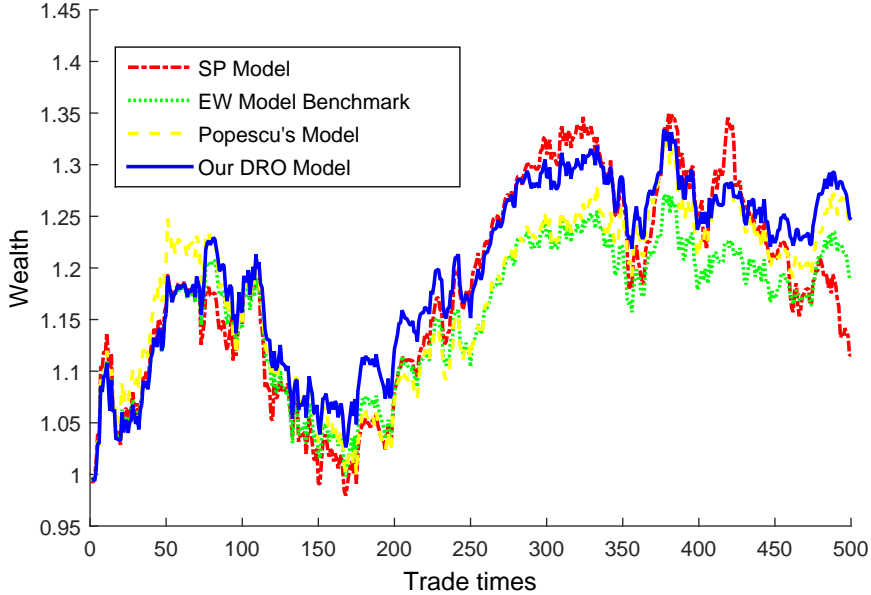


FIG. 1. Wealth evolution with the trading times.

out the corresponding parameters $\epsilon_{500} = 0.2621$ and $\sigma_{500} = 0.9155$. Moreover, we use the equally weighted portfolio as a benchmark strategy $Y(\xi)$ and set the small positive number satisfying (2.11) by 0.01. The entropic approximation parameter γ is fixed at 7000⁶. Note that if we set the parameters $\epsilon_{500} = 0$ and $\sigma_{500} = 0$, then the ambiguity set collapses to Popescu’s exact mean-covariance distributionally robust model [38]. In the numerical experiments, we compare our model with Popescu’s model and 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 2014 installed in a PC with Windows 7 operating system. We use the built-in optimization solver *fmincon* to solve minimization problem (5.3).

Table 1 summarizes the daily returns generated by the three portfolio models, where “L”, “H” and “A” denote respectively the lowest, highest and average returns from day 501 to day 1000. We record the number of days when the overall portfolio return falls below 1 and exceeds (or equals to) 1, denote them respectively by “Down” and “Up”. We can see that our distributionally robust optimization model achieves comparable average daily return and displays more stable performance with a narrower range between the best and worst return curves. Moreover, in comparison with the benchmark strategy, all of the other three models generate more “Up” times. Figure 1 depicts the performances of the three portfolio models. The figure indicates that the cumulative wealth curve of our DRO model outperforms the SP model and Popescu’s model, and it lies well above the benchmark wealth curve. The SP wealth curve is very close to the benchmark curve but it falls below at the end of the investment horizon. Compared to the SP model, our model and Popescu’s model display higher average

⁶ We use SAA method with 500 historical data to calculate the expectation in $e_\gamma(Z) := \frac{1}{\gamma} \ln \mathbb{E}_P[e^{-\gamma Z}]$. $\gamma = 7000$ ensures $\frac{1}{7000} \ln(500) < 10^{-3}$.

daily return, cumulative portfolio values at the end of horizon and generate more stable daily returns over the time horizon. In terms of computation time, our model requires overall 37546 seconds for finding out optimal allocations which is around 75 seconds per decision making, and Popescu’s model requires 47577 seconds (95 seconds per decision making) whereas the SP model requires only 12 seconds which is far more efficient than the other two models. Indeed, when the number of assets n is large, our model may fail since the size of the resulting optimization problem may increase significantly.

TABLE 2
Sensitivity-Daily return

Our DRO Model	L	H	A	Down	Up
$(\gamma = 7000, \delta = 0.01)$ -R	0.9717	1.0541	1.0007	96	104
$(\gamma = 70000, \delta = 0.01)$ -R	0.9719	1.0544	1.0005	98	102
$(\gamma = 7000, \delta = 0.01)$ -F	0.9713	1.0553	1.0005	101	99
$(\gamma = 7000, \delta = 0.001)$ -R	0.9562	1.0571	1.0003	95	105

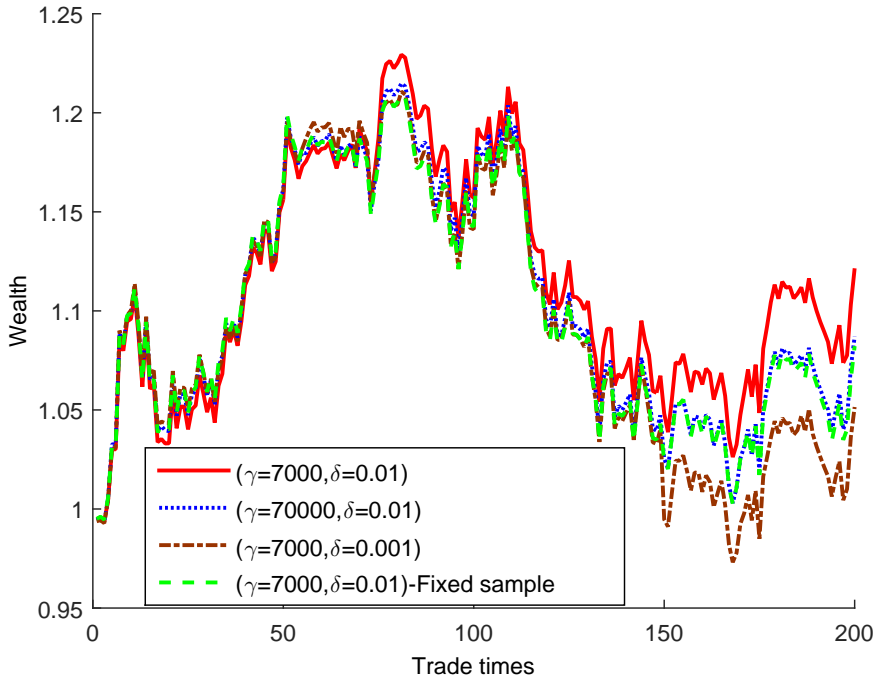


FIG. 2. Sensitivity-Wealth evolution with the trading times

We have also carried out tests on the sensitivity of our model w.r.t. the change of parameters including γ , ϵ_N and σ_N (which determine the size of the ambiguity set), as well as the sample size. Table 2 and Figure 2 display our findings. In Table 2, ‘R’ refers to the case where the historical data are rolling over 500 trading days and ‘F’ the case where the 500 historical data are fixed. The change of the ambiguity is made via parameter δ which in turn determines parameters ϵ_N and σ_N defined as in

(4.6), for example, $\delta = 0.01$ generates $\epsilon_N = 0.2621$, $\sigma_N = 0.9155$, and $\delta = 0.001$ gives $\epsilon = 0.2971$, $\sigma = 1.0395$. From the table, we can see that our DRO model is fairly stable w.r.t. the change of historical data (samples) and parameter γ . In contrast, it is more sensitive to the variation of δ . Figure 2 depicts the performance of the model w.r.t. the change of the parameters. The solid red curve is the base model with rolling sample and fixed parameters ($\gamma = 7000$, $\epsilon_N = 0.2621$, $\sigma_N = 0.9155$). The blue dotted curve, the green dashed curve and the red dash-dot curve correspond respectively to the perturbed parameters γ , the sample and (ϵ_N, σ_N) . The figure shows that our model is more sensitive to the change of the ambiguity set than to the change of historical data and parameter γ . The findings indicate the size of the ambiguity set has substantial effect on the performance of our model.

TABLE 3
Average return

Model	Freedom 3	Freedom 4	Freedom 5	Freedom 6
Our DRO model	36(0.9998)	2(1.0000)	2(0.9991)	20(1.0001)
Popescu's model	36(0.9993)	36(0.9997)	21(0.9989)	29(0.9994)
SP model	37(0.9989)	37(0.9996)	12(0.9988)	39(0.9995)
EW model (benchmark)	(1.0003)	(0.9993)	(0.9986)	(1.0000)

To further investigate stability of our model, we have also carried out tests with synthetic data for the ten stocks discussed above. Specifically, we randomly choose part of the historical data of the ten stocks typically of size 100 and use them as training data to figure out the optimal investment strategy for tomorrow (the next day of trading). We do so for each of the models to be tested. Let us use \hat{x} to denote such a strategy (model dependent). Next, we examine the performance of \hat{x} , that is, the portfolio return based on \hat{x} in the next trading day. To this end, we generate some independent test data with some t-distribution. Note that compared to normal distributions, t-distributions have been widely used to model return of financial assets whose distributions are usually fat-tailed. We refer the readers to Hu and Kercheval [23] and Platen and Sidorowicz [37] for recent developments and empirical evidences.

We use a t-distribution with specified degree of freedom to generate 1000 samples for the stock return rate tomorrow and scale them by multiplying each of the samples with the standard deviation of the training data. The modified data is then used to calculate the average return for tomorrow, that is, $\frac{1}{1000} \sum_{i=1}^{1000} \hat{x}^T \xi_i$, where ξ_1, \dots, ξ_{1000} denote the scaled samples.

We repeat this kind of experiment in total 200 times with 50 different training datasets and 4 different test datasets (generated by 4 t-distributions with different degrees of freedom and subsequently scaled as described above). The results are displayed in Table 3 and Figures 3-6. In Table 3, the last row displays the average return from the benchmark strategy. For each element $a(b)$ in the table, 'a' denotes the number of times where the average return falls below the benchmark (over the 50 experiments) and 'b' the average return. We can see that our DRO model achieves comparable average daily return. Figures 3-6 are plotted as follows. The horizontal axis depicts 50 different training datasets (they can be of any order but they have to be consistent for all models to be compared) and the vertical axis marks the average return (calculated with the test data). By connecting the 50 points, we obtain the average return curve for the specified strategy. From the figures, we can see that our proposed DRO model is more stable than the SP model and Popescu's model albeit it

does not necessarily achieve best return in every experiment. This is perhaps because we are using a relatively larger ambiguity set and hence the resulting optimal strategy is more resistant to the fluctuation of the data although no theoretical evidence can be established at this stage.

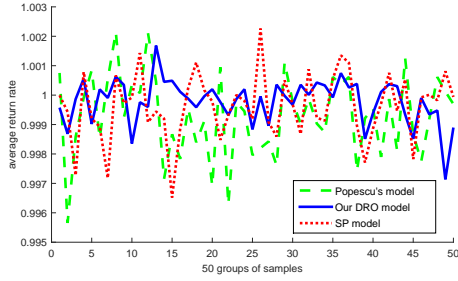


FIG. 3. Freedom 3

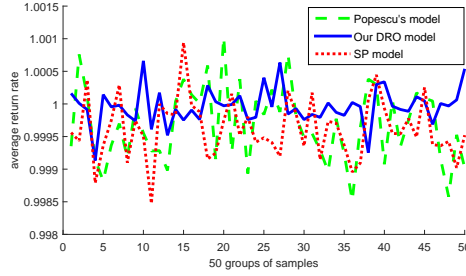


FIG. 4. Freedom 4

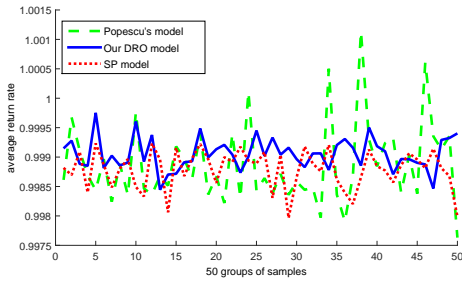


FIG. 5. Freedom 5

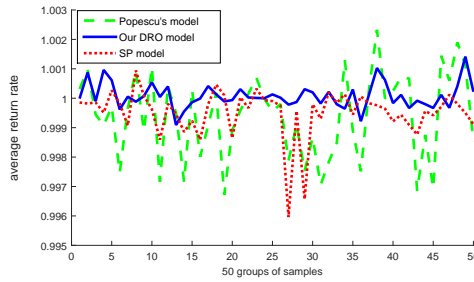


FIG. 6. Freedom 6

Acknowledgements. We would like to thank the associate editor Professor Darinka Dentcheva for organizing an effective review and two anonymous referees for insightful comments and constructive suggestions which help us significantly consolidate the paper. The research is supported by EPSRC grant EP/M003191/1 and NSFC grant 11571056. 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: Proc. 13 (2003), pp. 1-17.
- [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*, Oper. Res., 50 (2002), pp. 358-374.
- [6] D. BERTSIMAS AND I. POPESCU, *Optimal inequalities in probability theory: a convex optimization approach*, SIAM J. Optim., 15 (2005), pp. 780-804.

- [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*, Rev. Financ. Stud. 4 (1991), pp. 315-342.
- [8] A. BIGLOVA, S. ORTOBELLI, S. RACHEV AND S. STOYANOV, *Different approaches to risk estimation in portfolio theory*, J. Portfol. Manag., 31 (2004), pp. 103-112.
- [9] F. BLACK AND R. LITTERMAN, *Global portfolio optimization*, Financ. Analys. J., 48 (1992), pp. 28-43.
- [10] J. M. BORWEIN, *Stability and regular points of inequality systems*, J. Optim. Theory Appl., 48 (1986), pp. 9-52.
- [11] M. BROADIE, *Computing efficient frontiers using estimated parameters*, Annal. Oper. Res., 45 (1993), pp. 21-58.
- [12] L. CHEN, S. HE AND S. ZHANG, *Tight bounds for some risk measures, with applications to robust portfolio selection*, Oper. Res., 59 (2011), pp. 847-865.
- [13] V. K. CHOPRA AND W. T. ZIEMBA, *The effects of errors in means, variances, and covariance on optimal portfolio choice*, J. Portfol. Manag., 19 (1993), pp. 6-11.
- [14] E. DELAGE AND Y. YE, *Distributionally robust optimization under moment uncertainty with application to data driven problems*, Oper. Res., 58 (2010), pp. 595-612.
- [15] J. DUPAČOVÁ, *Uncertainties in minimax stochastic programs*, Optimization, 60 (2011), pp. 10-11.
- [16] L. EL-GHAOU, L. M. OKS AND F. OUSTRY, *Worst-case Value-at-Risk and robust portfolio optimization: A conic programming approach*, Oper. Res., 51 (2003), pp. 889-909.
- [17] H. FÖLLMER AND A. SCHIED, *Stochastic Finance: An Introduction in Discrete Time*, Walter de Gruyter, Berlin, 2004.
- [18] H. FÖLLMER AND T. KNISPEL, *Entropic risk measures: coherence vs. convexity, model ambiguity, and robust large deviations*, Sto. Dynam., 11 (2011), pp. 333-351.
- [19] D. GOLDFARB AND G. IYENGAR, *Robust portfolio selection problems*, Math. Oper. Res., 28 (2003), pp. 1-38.
- [20] B. L. GORISSEN, *Robust fractional programming*, J. Optim. Theory Appl., 166 (2015), pp. 508-528.
- [21] S. GUO, H. XU AND L. ZHANG, *Convergence analysis for mathematical programs with distributionally robust chance constraint*. To appear in *SIAM J. Optimization*.
- [22] B. V. HALLDÓRSSON AND R. H. TÛTÛNCÛ, *An interior-point method for a class of saddle-point problems*, J. Optim. Theory Appl., 116 (2003), pp. 559-590.
- [23] W. HU AND A.N. KERCHEVAL, *Portfolio optimization for student t and skewed t returns*, Quant. Finan., 10 (2010), 91-105.
- [24] M. KAPSOS, N. CHRISTOFIDES AND B. RUSTEM, *Worst-case robust Omega ratio*, Eur. J. Oper. Res., 243 (2014), 499-507.
- [25] 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.
- [26] J. B. LASSERRE, *Global optimization with polynomials and the problem of moments*, SIAM J. Optim., 11 (2011), pp. 796-817.
- [27] Y. LIU AND H. XU, *Entropic approximation for mathematical programs with robust equilibrium constraints*, SIAM J. Optim., 24 (2014), pp. 933-958.
- [28] Z. LU, *A new cone programming approach for robust portfolio selection*, Tech. Rep., Department of Mathematics, Simon Fraser University, Burnaby, BC, 2006.
- [29] Z. LU, *Robust portfolio selection based on joint ellipsoidal uncertainty set*, Optim. Meth. Soft., 26 (2011), pp. 89-104.
- [30] H. M. MARKOWITZ, *Portfolio Selection*, J. Financ., 1 (1952), pp. 77-91.
- [31] K. NATARAJAN, D. PACHAMANOVA AND M. SIM, *Constructing risk measures from uncertainty sets*, Oper. Res., 57 (2009), pp. 1129-1141.
- [32] J-S. PANG, *Error bounds in mathematical programming*, Math. Program., 79 (1997), pp. 299-332.
- [33] P. PARRILO, *Structured semidefinite programs and semialgebraic geometry methods in robustness and optimization*, PhD Thesis, California Institute of Technology, 2000.
- [34] G. Ch. Pflug and W. Römisch, *Modelling, Measuring and Managing Risk*, World Scientific, London, 2007.
- [35] G. C. PFLUG AND A. PICHLER, *Multistage Stochastic Optimization*, Springer Series in Operations Research and Financial Engineering, 2011.
- [36] M. PINAR AND R. H. TÛTÛNCÛ, *Robust profit opportunities in risky financial portfolios*, Oper. Res. Lett., 33 (2005), pp. 331-340.
- [37] E. PLATEN AND R. SIDOROWICZ, *Empirical evidence on student t log-returns of diversified world stock indices*, J. Stat. Theory Prac., 2 (2008), 233-251.

- [38] I. POPESCU, *Robust mean-covariance solutions for stochastic optimization*, Oper. Res., 55 (2007), pp. 98-112.
- [39] 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.
- [40] R. T. ROCKAFELLAR AND R. J-B. WETS, *Variational analysis*, Springer, Berlin, 1998 (3rd printing 2009).
- [41] N. RUJEERAPAIBOON, D. KUHN AND W. WIESEMANN, *Robust growth-optimal portfolios*, Manag. Scien., 62 (2016), pp. 2090-2109.
- [42] A. RUSZCZYŃSKI AND A. SHAPIRO, *Stochastic Programming*, Elsevire, Amsterdam, 2003.
- [43] 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.
- [44] W. F. SHARPE, *Mutual funds performance*, J. Busin., 39 (1966), pp. 119-138.
- [45] W. F. SHARPE, *The Sharpe ratio*, J. Portfolio Manag. 21 (1994), pp. 4958.
- [46] J. SHAWE-TAYLOR AND N. CRISTIANINI, *estimating the moments of a random vector with applications*, Proc. GRETSI 2003 Conference, pp. 47-52.
- [47] T. J. STIELTJES, *Recherches sur les fractions continues*, Annales de la Faculte des Sciences de Toulouse, 8 (1894), pp. 1-122.
- [48] S. STOYANOV, S. RACHEV AND F. FABOZZI, *Optimal financial portfolios*, Appl. Math. Financ., 14 (2007), pp. 403-438.
- [49] H. SUN AND H. XU, *Asymptotic convergence analysis for distributional robust optimization and equilibrium problems*, Math. Oper. Res., 41 (2016), pp. 377-401.
- [50] X. TONG AND F. WU, *Robust reward-risk ratio optimization with application in allocation of generation asset*, Optimization, 63 (2014), pp. 1761-1779.
- [51] R. H. TÜTÜNÇÜ AND M. KOENIG, *Robust asset allocation*, Annal. Oper. Res., 132 (2004), pp. 157-187.
- [52] W. WIESEMANN, D. KUHN AND M. SIM, *Distributionally robust convex optimization*, Oper. Res., 62 (2014), pp. 1358-1376.
- [53] H. Xu, *Uniform Exponential Convergence of Sample Average Random Functions under General Sampling with Applications in Stochastic Programming*, em Journal of Mathematical Analysis and Applications, Vol. 368, pp. 692-710, 2010.
- [54] H. XU, Y. LIU AND H. SUN, *Distributionally robust optimization with matrix moment constraints: lagrange duality and cutting plane methods*, optimization online, May, 2015.
- [55] S. ZHU AND M. FUKUSHIMA, *Worst-case conditional value-at-risk with application to robust portfolio management*, Oper. Res., 57 (2009) pp. 1155-1168.
- [56] S. ZHU, D. LI AND S. Y. WANG, *Robust portfolio selection under downside risk measures*, Quant. Financ., 9 (2009), pp. 869-885.
- [57] S. ZHU, M. FAN AND D. LI, *Robust portfolio selection via learning with mixture model*, February, 2013.
- [58] S. ZYMLER, D. KUHN AND B. RUSTEM, *Worst-case value at risk of nonlinear portolios*, Manag. Sci., 59 (2013), pp. 172-188.