Data-Driven Optimization of Reward-Risk Ratio Measures

Ran Ji
Department of Systems Engineering and Operations Research, George Mason University, rji2@gmu.edu

Miguel A. Lejeune
Department of Decision Sciences, The George Washington University, mlejeune@gwu.edu

We investigate a class of fractional distributionally robust optimization problems with uncertain probabilities. They consist in the maximization of ambiguous fractional functions representing reward-risk ratios and have a semi-infinite programming epigraphic formulation. We derive a new fully parameterized closed-form to compute a new bound on the size of the Wasserstein ambiguity ball. We design a data-driven reformulation and solution framework. The reformulation phase involves the derivation of the support function of the ambiguity set and the concave conjugate of the ratio function. We design modular bisection algorithms which enjoy the finite convergence property. This class of problems has wide applicability in finance and we specify new ambiguous portfolio optimization models for the Sharpe and Omega ratios. The computational study shows the applicability and scalability of the framework to solve quickly large, industry-relevant size problems, which cannot be solved in one day with state-of-the-art MINLP solvers.

Key words: Data-Driven Optimization, Distributionally Robust Optimization, Reward-Risk Ratio, Wasserstein Metric, Fractional Programming

History: This Version: Feb 2020.

1. Introduction

1.1. Problem Overview

Stochastic Programming (SP) is extensively used to solve optimization problems under uncertainty. SP models usually require the knowledge of the probability distribution of the random variables. However, the full information about the distribution is not always available in practice and, often, can only be approximated from data. To alleviate the difficulty of specifying the probability distribution, an alternative modeling approach, Distributionally Robust Optimization (DRO) (Delage and Ye 2010, Wiesemann et al. 2014), has recently received extensive attention. The DRO paradigm is designed for optimization under uncertainty where the random variable is governed by a probability distribution that is itself subject to uncertainty. The distribution belongs to an ambiguity set that contains all underlying distributions compatible with the optimizer's prior information.

This paper investigates a class of distributionally robust ratio (**DRR**) optimization problems which maximizes a fractional function representing a reward-risk ratio and in which the ambiguity set includes the true probability distribution with a large specified probability. This class of problems and formulations is very broad and encompassing. It can for example be applied in (i) financial optimization, in which asset returns are random variables whose probability distribution is difficult to elicit, and investment strategies are frequently determined using risk-adjusted return metrics (see, e.g., Bacon (2012), Cheridito and Kromer (2013)), (ii) supply chain optimization, when the fill rate service level defined as the fraction of uncertain demand satisfied from available inventory is required to exceed a certain value (Goh and Sim 2011), (iii) resource repartition problems, in which one maximizes the efficiency of the utilization of resources (i.e., average output per unit of time) subject to uncertain input units (see Pan et al. (2015) for an example in reservoir management), (iv) forestry management in which forest owners seek to maximize the reliability-to-stability revenue ratio (Lejeune and Kettunen 2017), or (v) assortment problems with mixed multinomial logit demand (Sen et al. 2018). This study focuses on the financial sector and proposes a series of novel **DRR** portfolio optimization problems implementing risk-adjusted return performance measures.

The generic formulation for the class of problems studied here is:

$$\max_{x \in \mathcal{X}} \inf_{\mathbb{P} \in \mathcal{D}} \frac{\mu(x,\xi)}{\rho(x,\xi)} \ , \tag{1}$$

where $x \in \mathbb{R}^M_+$ is a vector of decision variables and $\xi \in \mathbb{R}^M$ denotes the vector of uncertain variables with unknown true probability distribution $\mathbb{Q}$. The random variable ξ is defined on a probability space $(\Omega, \sigma(\Omega), \mathbb{P})$, in which $\sigma(\Omega)$ is the σ-algebra of Ω. The function $\mu(\cdot,\cdot) : \mathbb{R}^M \times \mathbb{R}^M \to \mathbb{R}$ in the numerator of the fractional function represents a reward measure, while the function $\rho(\cdot,\cdot) :$ $\mathbb{R}^M \times \mathbb{R}^M \to \mathbb{R}_{++}$ is a risk measure taking positive values. The set $\mathcal{X}$ is a polytope representing the convex feasible area defined by the deterministic constraints. Problem (1) maximizes the lower bound of the fractional function with respect to the worst-case distribution in the ambiguity set.

Let β be an auxiliary non-negative decision variable. The equivalent epigraphic formulation **DRR**

$$\textbf{DRR:} \quad \max \quad \beta \tag{2}$$

$$s.t. \quad \inf_{\mathbb{P} \in \mathcal{D}} \frac{\mu(x,\xi)}{\rho(x,\xi)} \geq \beta \tag{3}$$

$$x \in \mathcal{X} \tag{4}$$

$$\beta \in \mathbb{R}_+ \tag{5}$$

of (1) maximizes the worst-case reward-risk ratio value β which is upper-bounded by the ambiguous expression in the left-hand side of (3).

In this paper, the ambiguity set $\mathcal{D}$ is constructed via a data-driven approach using a finite number of data points. A key objective is to solve **DRR** efficiently, without any further assumptions about the structure of the true probability distribution. Problem **DRR** is in general nonconvex and NP-hard. The difficulty to solve **DRR** stems from two main reasons. First, **DRR** is a semi-infinite optimization problem, since (3) has infinitely many constraints. Second, the right-hand side β of the stochastic fractional inequality within constraint (3) is a decision variable and not a fixed parameter. Sections 3 and 4 describe how this further complexifies the solution process and in particular the derivation of a tractable counterpart for (3).

DRO dates back to the minimax modeling approach of Žáčková (1966). We give next a succinct review of modeling and algorithmic approaches presented in the DRO literature. Erdoğan and Iyengar (2006) propose the first ambiguous chance-constrained programming formulation in which the distance between probability distributions is measured with the Prohorov metric. They use the Strassen-Dudley Representation Theorem to derive a robust sampled problem which approximates the ambiguous problem. Ben-Tal et al. (2013) study constraints involving general convex functions and use ϕ-divergence to account for ambiguity. They show that tractable robust counterparts can be obtained for several specific ϕ-divergence (e.g., Kullback-Leibler, Burg entropy) functions. Wiesemann et al. (2014) propose a general framework for distributionlly robust convex optimization problems with expectation constraints. They derive tractable reformulations under the assumptions that the expectation function is bilinear in the decision and random variables and that the ambiguity set admits a conic representation. Ben-Tal et al. (2015) propose a framework to derive robust counterparts of the constraints that are nonlinear in the probability vector. They use the concepts of conjugate functions as well as Fenchel and conic duality to come up with tractable robust counterparts. Using the same framework, Postek et al. (2016) derive robust counterparts for a variety of risk measures and ambiguity sets. Esfahani and Kuhn (2017) study data-driven distributionally robust linear expectation problems with ambiguity sets based on the Wasserstein distance metric and derive reformulations under convexity assumptions. Gao and Kleywegt (2016) investigate the same DRO problem and consider Wasserstein distances of any order (i.e., ≥ 1) without imposing any restrictions on the form of the reference distribution.

1.2. Risk-Adjusted Return Performance Measures

The **DRR** class of optimization problems is suited for the derivation of risk-adjusted return investment policies. The earliest reward-risk ratio measure dates back to 1950s with the safety-first measure (Roy 1952) followed, shortly after, by the Sharpe ratio (Sharpe 1966). A flurry of performance ratio measures, such as the Sortino-Satchel (Sortino and Satchel 2001), or Omega (Keating and Shadwick 2002) ratios, have been introduced since. Bacon (2012) and Cheridito and Kromer (2013) provide in-depth analyses of ratio performance measures.

The evaluation of the value of any reward-risk ratio is affected by the knowledge and assumptions made on the distribution of the asset returns. Typically, a one-point estimate (i.e., mean return vector) is used for the vector of random asset returns (Sharpe 1966). This leads to the well-documented estimation risk problem. To mitigate the estimation risk, SP, Robust Optimization(RO), and DRO models have been proposed. SP models (e.g., VaR-based models (Gaivoronski and Pflug 2005), CVaR-based models (Rockafellar and Uryasev 2000), probabilistic Markowitz (Bonami and Lejeune 2009)) explicitly model asset returns as random variables which yield a probability distribution that must be specified. RO models (e.g., Goldfarb and Iyengar (2003)) allow the uncertain parameters to take any value within the uncertainty set taking usually a polytopic, box, or ellipsoidal form. Calafiore (2007) study some DRO portfolio optimization models under the assumption that the true return distribution is within some distance - measured with the Kullback-Leibler divergence - from a reference distribution. Pflug and Wozabal (2007) and Wozabal (2012, 2014) conduct a series of DRO studies with the Wasserstein distance metric and present financial illustrations. Pflug and Wozabal (2007) model the ambiguity set with the Wasserstein distance metric under the assumption that the probability space has a finite support with fixed atoms. While they focus on the CVaR metric, we strive in this study to maximize several worst-case risk-adjusted return performance measures.

Very few SP, RO or DRO models have been proposed for reward-risk ratio performance measures. Bailey and de Prado (2012) consider the probabilistic Sharpe ratio measuring the probability that the ratio of excess return to standard deviation exceeds a certain value. Deng et al. (2013) propose the VaR-adjusted Sharpe ratio measure to construct a portfolio with the largest worst-case scenario Sharpe ratio within a given confidence interval. They show that the original Sharpe ratio and the probabilistic Sharpe ratio are both special cases of the VaR-adjusted Sharpe ratio. Kapsos et al. (2014) study the worst-case Omega ratio with three types of uncertainties (mixture distribution, box (hyper-rectangle) uncertainty, ellipsoidal uncertainty).

Liu et al. (2017) investigate a distributionally robust reward-risk ratio optimization model with moment-based ambiguity set in which the means and covariances of distributions are restricted to intervals. They leverage the Lagrangian dualization to reformulate the DRO problem as a non-convex semi-infinite programming problem and approximate the semi-infinite constraints with the entropic risk measure, then develop an iterative method to solve the resulting reformulation. As will be shown later, compared to Liu et al. (2017), our models are constructed with a Wasserstein ambiguity set, which allows us to reformulate the semi-infinite constraints into a finite dimensional space and gives a convex programming problem if the worst-case reward-risk ratio value β is fixed. The reformulated problem can then be solved exactly via efficient bisection-type algorithms.

This study proposes new DRO portfolio problems that implement risk-adjusted return performance measures and belong to the class of **DRR** problems. In terms of model features, our study is closest to Postek et al. (2016)'s work which considers a Sharpe ratio constraint. A key difference with our study is that their model includes a constraint that requires the attainment of a *fixed*, pre-specified Sharpe ratio value, while we build the portfolio that maximizes the ambiguous Sharpe ratio value (see (1)). In the epigraphic reformulation, the right-hand side of the distributionally robust constraint (3) is not a fixed parameter, but a decision variable.

1.3. Contributions

The main contributions of this paper are manifold. First, we propose a new class of DRO problems in which we maximize fractional stochastic functions defining reward-risk ratios and ensure that the true probability distribution is within the ambiguity set with a large probability. This class of ambiguous semi-infinite problems has wide applicability, such as in finance, supply chain management, forestry, or resource allocation problems to name a few, and has not been studied before to our knowledge. Second, we derive in Theorem 2 (Section 2) a new bound on the radius of the Wasserstein ball. The bound offers a probabilistic guarantee on the inclusion of the unknown true distribution into the ambiguity set and is valid for discrete and continuous distributions. The bound defines the admissible size of the Wasserstein ambiguity set for a given number of data points characterizing the nominal distribution and permits to obtain a probabilistic lower bound on the optimal value of the true reward-risk ratio. We provide a closed-form, fully parameterized, formula for computing the value of the bound, which makes its very easy to use for computational tests. Third, we develop a computationally efficient reformulation (Section 3) and algorithmic (Section 5) framework when the distribution $\mathbb{P}$ to be elicited has fixed atoms and uncertain probabilities. We reformulate the semi-infinite programming problem **DRR** into a finite dimensional space by deriving robust counterparts for ambiguous reward-risk ratio constraints. We design bisection algorithms that solve efficiently problems of form **DRR**. The algorithms are modular, solve a sequence of convex problems, implement a descent-type approach, and converge finitely. While the idea underlying the bisection algorithms is simple, their computational efficiency is undeniable. They permit to solve problems that could not be solved otherwise in 24 hours with top optimization solvers, such as Baron. Fourth, we focus on the portfolio optimization context and propose a series of new portfolio optimization models based on risk-adjusted return measures expressed as ratios (Section 4), and show the applicability and scalability of our framework via an extended computational study (Section 6). We solve industry-relevant problems larger (i.e., 4 times more assets) than those in the DRO literature. The solution time does not increase with the number of securities, which is a critical feature permitting to handle larger portfolios if necessary. The out-of-sample analysis illustrates the robustness of the **DRR** Sharpe ratio model and its superior performance compared to the $1/N$ equal allocation strategy and the ambiguity-free Sharpe ratio model.

2. Specification of Ambiguity Set

The literature distinguishes two main types of ambiguity sets. Moment-based ambiguity sets are defined with respect to some moments of the probability distribution (see Delage and Ye (2010), Wiesemann et al. (2014), and the references therein). Statistical-based ambiguity sets include all underlying probability distributions that fall within a certain specified statistical distance from a reference distribution, which is an approximation of the true probability distribution. Popular statistical distance measures, such as Prohorov (Erdoğan and Iyengar 2006), Kullback-Leibler divergence (Jiang and Guan 2015), and Wasserstein (Pflug and Wozabal 2007, Wozabal 2012, 2014, Gao and Kleywegt 2016, Esfahani and Kuhn 2017), have been used to estimate the "closeness" between distributions. Esfahani and Kuhn (2017), and Wiesemann et al. (2014) observe that tractable conic reformulations are in general easier to derive for models with moment-based ambiguity sets.

For moment-based ambiguity sets, the information about the distribution of the uncertain parameters is inferred from the available data points. As their number increases, the values of the estimated first and second moments of the distribution eventually converge. However, there can be an infinite number of distributions that have the same first and second moments (including the true distribution), which means that moment-based approaches do not guarantee convergence to the true distribution. In contrast, statistical-based ambiguity sets enjoy convergence properties. Let $d(\mathbb{P},\mathbb{P}_0)$ denote any statistical distance between the underlying distribution $\mathbb{P}$ and the reference distribution $\mathbb{P}_0$ and let θ be the largest admissible distance, the statistical-based ambiguity set is defined as: $\mathcal{D} = \{\mathbb{P} : d(\mathbb{P},\mathbb{P}_0) \leq \theta\}$. One attractive feature of statistical-based ambiguity sets is that the conservatism of the optimization problem can be controlled by adjusting the values assigned to the parameter θ defining the size of the ambiguity set. The DRO problem reduces to an ambiguity-free problem if $\theta = 0$. Three main criteria (Jiang and Guan (2015), Hanasusanto et al. (2015), Esfahani and Kuhn (2017)) prevail to select the ambiguity set $\mathcal{D}$. First, $\mathcal{D}$ must contain the true and unknown probability distribution $\mathbb{Q}$ with high probability. Second, the size of $\mathcal{D}$ should decrease when more data points (i.e., sample size goes to infinity) are used. Finally, $\mathcal{D}$ should allow for the derivation of tractable reformulations. Alternatively, Gao and Kleywegt (2016) propose to limit the probability of the Wasserstein distance between a given data-generating probability distribution and an empirical distribution (constructed with independent and identically distributed (IID) samples from the given distribution) exceeding the radius of the Wasserstein ball.

In this study, we use the Wasserstein metric to construct a statistical-based ambiguity set. We next define the Wasserstein metric, describe the reference distribution $\mathbb{P}_0$, and propose a constructive approach to size the ambiguity set (Theorem 2) and ensure that it contains the true unknown distribution with a pre-specified probability.

The Wassserstein metric is defined as a distance function between two probability distributions on a given probability space. We consider a Wasserstein ambiguity set with finite support (i.e., the support Ω includes $|\mathcal{N}|=N$ data points). Let ξ_j denote the j-th atom of the underlying distribution $\mathbb{P}$, ξ_i^0 be the i-th atom of the reference distribution $\mathbb{P}_0$, p_j (resp., p_i^0) denote the probability of the j-th (resp., i-th) atom of $\mathbb{P}$ (resp., $\mathbb{P}_0$), and π_{ij} be the bivariate probability mass function of ξ_j and ξ_i^0, $\forall i,j \in \mathcal{N}$.

Definition 1. (Wasserstein Metric). The Wasserstein distance $d_W(\mathbb{P},\mathbb{P}_0)$ of order 1 (also known as Kantorovich distance) between $\mathbb{P}$ and $\mathbb{P}_0$ over a finite support Ω is given by:

$$d_W(\mathbb{P},\mathbb{P}_0) = \inf_{\pi \geq 0} \left(\sum_{i,j \in \mathcal{N}} \pi_{ij} ||\xi_j - \xi_i^0|| : \begin{array}{l} \sum_{j \in \mathcal{N}} \pi_{ij} = p_i^0, \ \forall i \in \mathcal{N} \\ \sum_{i \in \mathcal{N}} \pi_{ij} = p_j, \ \forall j \in \mathcal{N} \end{array} \right) . \tag{6}$$

The Wasserstein distance (6) can be interpreted as the minimum transport cost of the probability mass distribution from $\mathbb{P}_0$ (i.e., supply or reference distribution) to $\mathbb{P}$ (i.e., demand or underlying distribution), where the unit cost between p_i^0 and p_j is $||\xi_j - \xi_i^0|| = |\xi_j - \xi_i^0|$.

The reference (or nominal) distribution $\mathbb{P}_0$ is an estimate of the true probability distribution $\mathbb{Q}$. We use in this study a nonparametric approach to build the reference probability distribution and define it by relying on a series of data points. The reference distribution $\mathbb{P}_0$ is defined as an empirical distribution constructed by N IID data points ξ_i^0, $i \in \mathcal{N}$, and is formulated as

$$\mathbb{P}_0 := \frac{1}{N} \sum_{i=1}^{N} \delta_{\xi_i^0} , \tag{7}$$

where $\delta_{\xi_i^0}$ represents the Dirac distribution concentrating unit mass at realization ξ_i^0.

Definition 2. (Wasserstein Ambiguity Set) Suppose that we have N independent samples $\xi_i^0, i = 1,\ldots,N$ from an unknown true distribution $\mathbb{Q}$ with compact support. Let $\mathcal{P}(\Omega)$ represent the space of all probability distributions $\mathbb{P}$ supported on Ω. The Wasserstein ambiguity set $\mathcal{D}_W$ is a ball of radius θ centered around the reference distribution $\mathbb{P}_0$:

$$\mathcal{D}_W := \{\mathbb{P} \in \mathcal{P}(\Omega) : d_W(\mathbb{P},\mathbb{P}_0) \leq \theta\} . \tag{8}$$

The next question concerns the size of the ambiguity set and of its radius θ so that the inclusion of the true distribution $\mathbb{Q}$ in the ambiguity set $\mathcal{D}_W$ is probabilistically guaranteed to be at least equal to q. We first introduce the q-valid radius concept.

Definition 3. (q-valid radius) The radius θ of the Wasserstein ball is thereafter said to be q-valid if its value ensures that $\mathbb{Q} \in \mathcal{D}_W$ with probability at least q.

Fournier and Guillin (2015) provide a measure concentration result (later adapted by Esfahani and Kuhn (2017)) giving a bound on the distance between a pair of probability distributions. Let $\mathbb{1}_A$ denote an indicator function where $\mathbb{1}_A = 1$ if event A occurs and $= 0$ otherwise.

THEOREM 1. *(Fournier and Guillin (2015), Theorem 2) Assume the following conditions hold:*

$$\exists \alpha > 1,\ \exists \gamma > 0,\ \exists \varepsilon_{\alpha,\gamma} = \int_{\Omega} e^{\gamma|\xi|\mathbb{P}(d\xi)} < \infty \tag{9}$$

Then for all $N \geq 1$, $M \neq 2$, and $\theta \geq 0$, we have

$$P(d_W(\mathbb{P}_0, \mathbb{Q}) \geq \theta) \leq C \exp(-cN\theta^{\max\{M,2\}}) \mathbb{1}_{\{\theta \leq 1\}} + C \exp(-cN\theta^{\alpha}) \mathbb{1}_{\{\theta > 1\}} \tag{10}$$

where the positive constants C and c depend on α, γ, $\varepsilon_{\alpha,\gamma}$ and the dimension M of ξ.

The above bound is however not easy to apply and use conveniently in a computational setting, since it requires the a priori determination of the parameters c and C in (10) to calculate the value of the bound (10) and subsequently infer the size of the q-valid radius. No particular guidance is offered by Fournier and Guillin (2015) to set these parameters. In a private email communication, Dr. Fournier acknowledged it and wrote that "*it is clearly 'surhuman' to compute the values of C and c in Theorem 2*" (Fournier 2018)[1].

In Theorem 2, we propose new bounds on 1) the probability that the Wasserstein distance between the reference and true distributions does not exceed a specified threshold and 2) the size of a q-valid radius. We provide closed-form formulae to compute the bounds that are self-contained and do not require the computation of extraneous parameters (such as C and c in Fournier and Guillin (2015)). The proof includes two main parts. Using the properties of the weighted total variation and the weighted Csiszar-Kullback-Pinsker inequality, Part 1 establishes an upper bound on the distance between two arbitrary distributions and a lower bound of the Kullback information. Part 2 connects the results of Part 1 with an asymptotic upper bound on the Wasserstein distance between probability distributions to derive a lower bound on the radius size of the ambiguity set.

THEOREM 2. *Assume there exist some $\xi_0 \in \Omega$ such that the following exponential integrability condition (defined by Wang et al. (2010))*

$$\Lambda(a) := \log \int_{\Omega} e^{a d_W(\xi,\xi_0)} \mathbb{Q}(d\xi) < +\infty \ , \quad \forall a > 0 \tag{11}$$

holds true for all $a > 0$. Let N be the number of data points, B be the diameter of the M-dimensional (i.e., $M \geq 1$) compact support Ω defined as:

$$B = \sup\{d_W(\xi_i, \xi_j) = ||\xi_i - \xi_j|| : \xi_i, \xi_j \in \Omega\} \ . \tag{12}$$

A lower bound on the probability that the Wasserstein distance between the reference distribution $\mathbb{P}_0$ and the true distribution $\mathbb{Q}$ does not exceed θ is given by:

$$P(d_W(\mathbb{P}_0, \mathbb{Q}) \leq \theta) \geq 1 - e^{\left(-N\left(\frac{\sqrt{4\theta(4B+3)+(4B+3)^2}}{4B+3}-1\right)^2\right)} \ . \tag{13}$$

[1] Theorem 2 in Fournier and Guillin (2015) is Theorem 1 in our manuscript.

Furthermore, if

$$\theta \geq \left(B+\frac{3}{4}\right)\left(-\frac{1}{N}\log(1-q)+2\sqrt{-\frac{1}{N}\log(1-q)}\right) \ , \tag{14}$$

then

$$P(d_W(\mathbb{P}_0,\mathbb{Q}) \leq \theta) \geq q \quad \Rightarrow \quad \mathbb{Q} \in \mathcal{D}_W \textit{ with probability at least } q. \tag{15}$$

Proof: Part 1: Let $\mathbb{P}_1$ and $\mathbb{P}_2$ be two arbitrary probability measures and $\xi_0 \in \Omega$ with diameter B. The Wasserstein distance between $\mathbb{P}_1$ and $\mathbb{P}_2$ is upper bounded by the weighted total variation (see Villani (2008), Theorem 6.15):

$$d_W(\mathbb{P}_1,\mathbb{P}_2) \leq \int d|\xi-\xi_0| \cdot d|\mathbb{P}_1-\mathbb{P}_2|(\xi) \ , \tag{16}$$

According to the weighted Csiszar-Kullback-Pinsker inequality (see Theorem 2.1 in Bolley and Villani (2005)), the total variation term in the right side of (16) is related to the Kullback information $H(\mathbb{P}_1|\mathbb{P}_2)$ as follows:

$$\int d|\xi-\xi_0| \cdot d|\mathbb{P}_1-\mathbb{P}_2|(\xi) \leq \left(\frac{3}{2}+\log\int e^{2||\xi-\xi_0||}d\mathbb{P}_2(\xi)\right)\left(\sqrt{H(\mathbb{P}_1|\mathbb{P}_2)}+\frac{1}{2}H(\mathbb{P}_1|\mathbb{P}_2)\right) \tag{17}$$

with the Kullback information of $\mathbb{P}_1$ with respect to $\mathbb{P}_2$ given by:

$$H(\mathbb{P}_1|\mathbb{P}_2) = \int \mathbb{P}_1(\xi)\cdot\log\frac{\mathbb{P}_1(\xi)}{\mathbb{P}_2(\xi)}d\mathbb{P}_2(\xi) \ . \tag{18}$$

Since $||\xi-\xi_0||$ is upper bounded by the diameter B of Ω, we have:

$$\log\int e^{2||\xi-\xi_0||}d\mathbb{P}_2(\xi) \leq \log e^{2B} = 2B \ . \tag{19}$$

Therefore, using (17) and (19), we obtain:

$$d_W(\mathbb{P}_1,\mathbb{P}_2) \leq \left(\frac{3}{2}+2B\right)\cdot\left(\sqrt{H(\mathbb{P}_1|\mathbb{P}_2)}+\frac{1}{2}H(\mathbb{P}_1|\mathbb{P}_2)\right) \ . \tag{20}$$

Setting $d_W(\mathbb{P}_1,\mathbb{P}_2) \geq \theta$, we then have:

$$\theta \leq \left(\frac{3}{2}+2B\right)\cdot\left(\sqrt{H(\mathbb{P}_1|\mathbb{P}_2)}+\frac{1}{2}H(\mathbb{P}_1|\mathbb{P}_2)\right) \ . \tag{21}$$

After rearrangement, the function $H(\mathbb{P}_1|\mathbb{P}_2)$ can be expressed in terms of θ and B:

$$H(\mathbb{P}_1|\mathbb{P}_2) \geq \left(\frac{\sqrt{4\theta(4B+3)+(4B+3)^2}}{4B+3}-1\right)^2 \ . \tag{22}$$

Part 2: Sanov's theorem states that the probability that a sequence of IID random variables (e.g., an empirical distribution $\mathbb{P}_0$) is near its limit (i.e., the true distribution $\mathbb{Q}$) satisfies a large deviation principle on a space equipped with weak convergence topology with the rate function defined by the relative entropy. Sanov's theorem provides an asymptotic form that quantifies the convergence rate of an empirical distribution to the true distribution, as follows:

$$P\left(d_W(\mathbb{P}_0, \mathbb{Q} \geq \theta)\right) \simeq e^{-N\alpha(\theta)} \ , \quad \text{as } N \to \infty \tag{23}$$

where θ denotes the Wasserstein distance between the empirical and true distributions and the rate function $\alpha(\theta)$ is given by

$$\alpha(\theta) := \inf\left\{H(\mathbb{P}_0|\mathbb{Q}) : d_W(\mathbb{P}_0, \mathbb{Q}) \geq \theta\right\} \ . \tag{24}$$

Wang et al. (2010) (Theorem 1.1, page 506) expands Sanov's theorem from the classical weak convergence topology to the Wasserstein metric, if and only if the exponential integrability condition (11) is satisfied. Due to $H(\mathbb{P}_0|\mathbb{Q}) \geq 0$ and (22), together with Sanov's theorem, we obtain:

$$-\inf H(\mathbb{P}_0|\mathbb{Q}) \leq \max -H(\mathbb{P}_0|\mathbb{Q}) \ \leq -\left(\frac{\sqrt{4\theta(4B+3)+(4B+3)^2}}{4B+3} - 1\right)^2 \tag{25}$$

Combining (24) and (25) along with the increasing monotonicity property of the exponential function, we can rewrite (23) as follows:

$$P\left(d_W(\mathbb{P}_0, \mathbb{Q} \geq \theta)\right) \lesssim e^{-N\left(\frac{\sqrt{4\theta(4B+3)+(4B+3)^2}}{4B+3} - 1\right)^2} \ . \tag{26}$$

Taking the complement of the probability in the left side of (26), we obtain (13) by approximately replacing the $\lesssim$ with $\leq$. We recall that N and B are prespecified parameters with know values. Since we want a lower bound on the size of the radius θ such that $P\left(d_W(\mathbb{P}_0, \mathbb{Q}) \leq \theta\right) \geq q$, we set

$$1 - e^{-N\left(\frac{\sqrt{4\theta(4B+3)+(4B+3)^2}}{4B+3} - 1\right)^2} \geq q \ , \tag{27}$$

and solve for θ, which gives $\theta \geq \left(B + \frac{3}{4}\right)\left(-\frac{1}{N}\log(1-q) + 2\sqrt{-\frac{1}{N}\log(1-q)}\right)$ and completes the proof. □

Note that in the exponential integrability condition (11) of Wang et al. (2010), the integral part takes the form of a moment generating function. The condition (11) essentially requires the first moment of the empirical distribution ξ_0 to be finite ($< +\infty$) on the Wasserstein metric space.

As shown by the proof, the bound is valid regardless of whether the (unknown) true distribution is discrete or continuous. It follows from (14) that, for any given diameter B and probability level q, the value of θ decreases monotonically with the number of data points N. This implies the weak convergence property of the Wasserstein ambiguity set (8). When N becomes sufficiently large (e.g., $+\infty$), θ converges to zero and the Wasserstein ambiguity set reduces to the singleton of the reference distribution $\mathbb{P}_0$.

The diameter B defines the maximal Wasserstein distance between any pair of atoms in Ω. For any given N, M, and B defined by (12), Theorem 2 sets an one-on-one projection of q on θ. In other words, by properly setting the value of θ (i.e., using (14)), Theorem 2 establishes a probabilistic guarantee on the inclusion of the true probability distribution $\mathbb{Q}$ in the Wasserstein ambiguity set $\mathcal{D}_W$. The ambiguity set can then be understood as a confidence set containing the true probability distribution with confidence level q. Hereafter we call θ_q the tight q-valid radius, and the ball $\mathcal{D}_{W_q}$ with radius θ_q the q-valid Wasserstein ambiguity set that contains the true probability distribution $\mathbb{Q}$ with probability at least q: $P(\mathbb{Q} \in \mathcal{D}_{W_q}) \geq q$. The next corollary follows.

Corollary 1. *For any $N > 1$ and any arbitrary value for $q \in [0,1], M$, and B, the true distribution $\mathbb{Q}$ is contained within the Wasserstein ambiguity set $\mathcal{D}_W$ with probability at least q if the radius θ_q of the ball $\mathcal{D}_W$ is such that: $\theta_q \geq \theta$, with θ defined by (14).*

We next establish a bound on the value of the reward-risk ratio problem with known true probability distribution. Consider the epigraph formulation **DRRQ** of the ambiguity-free reward-risk ratio problem:

$$\begin{aligned} \textbf{DRRQ:} \quad & \max \quad && \beta_{\mathbb{Q}} && (28)\\ & s.t. && \frac{\mu(x,\xi)}{\rho(x,\xi)} \geq \beta_{\mathbb{Q}} && (29)\\ & && (4)-(5) \end{aligned}$$

where the optimal value $\beta^*_{\mathbb{Q}}$ represents the largest attainable ratio under the true distribution $\mathbb{Q}$.

Let $q = 0.95$. The set $\mathcal{D}_{W_{0.95}}$ is such that $P(\mathbb{Q} \in \mathcal{D}_{W_{0.95}}) \geq 0.95$. There is thus a 95% probability that the optimal value of **DRRQ** exceeds or is equal to that of **DRR**. Additionally, if x is feasible for all distributions in $\mathcal{D}_{W_q}$, the probability that x is feasible under the true probability distribution $\mathbb{Q}$ is also 95% or larger. The above observations lead to Corollary 2 and Corollary 3.

Corollary 2. *(Probabilistic lower bound) Let β^* be the optimal value of problem* **DRR** *with q-valid ambiguity set $\mathcal{D}_{W_q}$ and let $\beta^*_{\mathbb{Q}}$ be the optimal value of the ambiguity-free problem* **DRRQ**. *The optimal value of the ambiguous problem* **DRR** *provides a probabilistic lower bound on the value of the true reward-ratio: $P(\beta^*_{\mathbb{Q}} \geq \beta^*) \geq q$.*

Corollary 3. *(Feasibility Guarantee) Let x be a feasible solution for problem* **DRR** *with q-valid ambiguity set $\mathcal{D}_{W_q}$. The probability of x feasible for problem* **DRRQ** *is at least equal to q.*

3. Reformulation Framework

In this section, we develop a reformulation framework for the class of **DRR** reward-risk ratio optimization models with Wasserstein ambiguity, in which the underlying probability distribution $\mathbb{P}$ has the same N atoms as the reference distribution $\mathbb{P}_0$, but the probabilities $p_1, p_2, \ldots, p_N$ of the atoms are decision variables whose values must be uncovered. In other words, we purport to "transport" the reference probability measure $\mathbb{P}_0$ to a new one $\mathbb{P}$ that sits on the same data points. Since the distances $||\xi_j - \xi_i^0||$ are fixed, the representation (6) of the Wasserstein distance is linear in π and the constraint set is polyhedral. We denote by $\mathcal{D}'_W$ the Wasserstein ambiguity set including the distributions with uncertain probabilities and sitting on the same atoms as the reference distribution. We use the notation $\alpha(x,p)$ instead of $\alpha(x,\xi)$ since the atoms of the distribution are fixed but their probabilities $p_i, i = 1,\dots,N$ are uncertain. For the same reason, we use p to represent the underlying distribution $\mathbb{P}$ and write $p \in \mathcal{D}$. We first utilize the definition of the Wasserstein ambiguity set (8) and the probabilistic guarantee results (Theorem 2) to reformulate **DRR**.

THEOREM 3. *Consider the q-valid Wasserstein ambiguity ball $\mathcal{D}'_{W_q}$ with radius θ_q defined by (14). The distributionally robust reward-risk ratio optimization problem* **DRR** *can be rewritten as:*

$$\begin{aligned} \textbf{DRRW:} \quad \max \quad & \beta \\ s.t. \quad & \alpha(x,p) \geq \beta, \qquad \forall p \in \mathcal{D}'_{W_q} \qquad (30) \\ & (4)-(5) \end{aligned}$$

where the q-valid Wasserstein ambiguity ball $\mathcal{D}'_{W_q}$ is defined by $\mathcal{D}'_{W_q} := \left\{ p : \exists \pi \in Z_{\mathcal{D}'_{W_q}} \right\}$ with

$$Z_{\mathcal{D}'_{W_q}} := \left\{ \begin{array}{ll} \sum_{i,j\in\mathcal{N}} \pi_{ij} ||\xi_j - \xi_i^0|| \leq \theta_q & (31) \\ \sum_{j\in\mathcal{N}} \pi_{ij} = p_i^0, \quad \forall i \in \mathcal{N} & (32) \\ \sum_{i\in\mathcal{N}} \pi_{ij} = p_j, \quad \forall j \in \mathcal{N} & (33) \\ \pi_{ij} \geq 0, \quad \forall i,j \in \mathcal{N} & (34) \end{array} \right\}$$

Proof: Since $\mathcal{D}'_{W_q}$ is q-valid, Theorem 2 implies that $P(d_W(\mathbb{P}_0, \mathbb{Q}) \leq \theta) \geq q$. Thus, (3) can be rewritten as

$$\inf_{p\in\mathcal{D}'_{W_q}} \left[\frac{\mu(x,p)}{\rho(x,p)} \right] \geq \beta \ ,$$

equivalent to (30). Finally, (31)-(33) come from the definition of the Wasserstein distance (6) and ambiguity set (8) and imply that any underlying probability distribution p belongs to $\mathcal{D}'_{W_q}$. □

Next, we reformulate the semi-infinite DRO problem **DRRW** in a form amiable to its exact numerical solution. It is important to note that **DRRW** maximizes the worst-case reward-risk ratio value β (in the right-hand side of the distributionally robust inequality) defined as a decision variable. As shown later, defining β as a decision variable further complexifies the solution process. We focus now on the challenges posed by (30). We introduce the concepts of conjugate and support functions to derive robust counterparts for the reward-risk ratio constraints (30).

3.1. Conjugate and Support Functions

We use the conjugate and support function concepts (see also Ben-Tal et al. (2015), Postek et al. (2016)) to derive robust counterparts for distributionally robust reward-risk ratio constraints (30) with Wasserstein ambiguity set including a finite set of fixed atoms. This section is decomposed into two independent reformulation stages involving respectively the support function of the ambiguity set (Section 3.2), and the concave conjugate of the ratio function (Section 3.3).

The *concave conjugate* $f_*(\cdot)$ of a function $f:\mathbb{R}^N_+ \to \mathbb{R}$ is defined as: $f_*(v) = \inf_{p\in\mathbb{R}^N_+}\{v^T p - f(p)\}$. The function $f_*(x,v)$ denotes the partial concave conjugate of $f(x,p)$ with respect to the second variable p. The *convex conjugate* $g^*(\cdot)$ of a function $g:\mathbb{R}^N_+ \to \mathbb{R}$ is given by: $g^*(v) = \sup_{p\in\mathbb{R}^N_+}\{v^T p - g(p)\}$. The *indicator function* $\delta(\cdot|\mathcal{D})$ of a nonempty ambiguity set $\mathcal{D}$ is defined as: $\delta(p|\mathcal{D}) = 0$, if $p\in\mathcal{D}$; $+\infty$, otherwise. Taking the convex conjugate of $\delta(p|\mathcal{D})$, we obtain the *support function* $\delta^*(\cdot|\mathcal{D})$: $\delta^*(v|\mathcal{D}) = \sup_{p\in\mathcal{D}} v^T p$.

Theorem 4 (see also Postek et al. (2016)) is instrumental to derive the robust counterparts of the ambiguous reward-risk ratio constraints. We denote by $ri(\mathcal{D})$ the relative interior of the set $\mathcal{D}$, $\delta^*(v|\mathcal{D})$ the support function of $\mathcal{D}$, and $f_*(x,v)$ the concave conjugate function of $f(x,p)$ with respect to the second component p.

Theorem 4. *Let $\eta\in\mathbb{R}_+$ and $f:\mathbb{R}^M\times\mathbb{R}^N\to\mathbb{R}$ be a function such that $f(x,p)$ is closed and concave for each fixed $x\in\mathbb{R}^M$ and assume that $ri(\mathcal{D})\cap\mathbb{R}^N_{++}\neq\emptyset$. The constraints*

$$f(x,p)\le\eta, \quad \forall p\in\mathcal{D} \tag{35}$$

hold true for any x if and only if:

$$\exists v: \delta^*(v|\mathcal{D}) - f_*(x,v) \le \eta\ . \tag{36}$$

If a solution v exists for (36), a semi-infinite programming problem with constraints (35) can be equivalently reformulated in a finite dimensional space. We refer to Postek et al. (2016) for a detailed proof of Theorem 4. Theorem 4 supports the decomposition of the reformulation process into two independent stages described in Sections 3.2 and 3.3. Note that we could have used linear and conic programming duality theory to derive the dual form of constraint (30) under Wasserstein ambiguity set. We use the framework based on support and conjugate functions because of its generality and its applicability to various types of ambiguity sets and concave uncertain functions.

3.2. Support Function of Wasserstein Ambiguity Set

In this section, we propose a linear programming formulation to derive the support function of the Wasserstein ambiguity set.

THEOREM 5. *The support function $\delta^*(v|\mathcal{D}'_W)$ of the Wasserstein ambiguity set $\mathcal{D}'_W$ is the optimal value of the linear programming problem* **WA-D**

$$\begin{aligned}
\textbf{WA-D}:\ \delta^*(v|\mathcal{D}'_W) = \inf_{\gamma,y} \quad & \gamma\theta + \sum_{i\in\mathcal{N}} y_i p_i^0 && (37)\\
s.t. \quad & ||\xi_j - \xi_i^0||\gamma + y_i \geq v_j, \qquad \forall i,j\in\mathcal{N} && (38)\\
& \gamma \geq 0\ , && (39)
\end{aligned}$$

with decision variables γ, y_i, where $y_i, i\in\mathcal{N}$ are unrestricted in sign.

Proof: Following the definition, we obtain the support function of $\mathcal{D}'_W$:

$$\begin{aligned}
\delta^*(v|\mathcal{D}'_W) = \sup_{\pi,p} \quad & \sum_{j\in\mathcal{N}} v_j p_j && (40)\\
s.t. \quad & (31)-(34)
\end{aligned}$$

with decision variables π_{ij} and p_j. Owing to (33), we substitute $\sum_{i\in\mathcal{N}} \pi_{ij}$ for $p_j, j\in\mathcal{N}$ in the objective function, which gives:

$$\begin{aligned}
\textbf{WA-P}:\ \delta^*(v|\mathcal{D}'_W) = \sup_{\pi} \quad & \sum_{j\in\mathcal{N}} v_j \left(\sum_{i\in\mathcal{N}} \pi_{ij}\right) && (41)\\
s.t. \quad & (31)-(32); (34)
\end{aligned}$$

Let γ and y be the dual variables for the constraints (31) and (32), respectively. The dual of **WA-P** is the linear problem **WA-D**, which provides the desired result, since strong duality holds for LP. $\square$

3.3. Reformulation of Reward-Risk Ratio Constraints

The above results allow us to reformulate the reward-risk ratio constraints (30). We recall that we use a Wasserstein ambiguity set with fixed atoms, in which the probabilities of the atoms are decision variables. Recall that the risk function $\rho(x,p)$ is assumed to be positive. Therefore, β is positive and we have:

$$-\alpha(x,p) = -\frac{\mu(x,p)}{\rho(x,p)} \leq -\beta \ \Leftrightarrow\ \rho(x,p) - \frac{1}{\beta}\mu(x,p) \leq 0, \qquad \forall p\in\mathcal{D}'_W\ .$$

Let $f(x,p) = \rho(x,p) - \frac{1}{\beta}\mu(x,p)$. For any arbitrary positive β, the constraint

$$f(x,p) = \rho(x,p) - \frac{1}{\beta}\mu(x,p) \leq 0, \qquad \forall p\in\mathcal{D}'_W \tag{42}$$

is in the form of (35) with $\eta = 0$. We derive the conjugate function of $f(x,p)$ to reformulate the reward-risk constraints (42) in a finite dimensional constraint space. We study two classes of function $f(x,p)$ that differ in their dependency (i.e., linear or nonlinear) on the probability vector p. This makes our reformulation approach applicable to various forms of reward and risk functions.

3.3.1. Linear Uncertainty We first consider the case of functions $f(x,p)$ that are linear in p. Both the reward $\mu(x,p)$ and risk $\rho(x,p)$ functions are linear in p:

$$\mu(x,p) = \sum_{j\in\mathcal{N}} p_j \mu_j(x) \tag{43}$$

$$\rho(x,p) = \sum_{j\in\mathcal{N}} p_j \rho_j(x) \tag{44}$$

The reward $\mu(x,p)$ (resp. risk $\rho(x,p)$) has possible outcomes $\mu_j(x)$ (resp. $\rho_j(x)$) with probability p_j. With (43) and (44), the modified reward-risk ratio function is

$$f^L(x,p) = \sum_{j\in\mathcal{N}} p_j \rho_j(x) - \frac{1}{\beta}\sum_{j\in\mathcal{N}} p_j \mu_j(x), \quad \forall p \in \mathcal{D}'_W \ . \tag{45}$$

For any $\beta > 0$, $f^L(x,p)$ is linear in p for every $x \in \mathbb{R}^M$ and its concave conjugate function is:

$$\begin{aligned} f^L_*(x,v) &= \inf_p \left\{ v^T p - f^L(x,p) \right\} = \inf_p \left\{ \sum_{j\in\mathcal{N}} v_j p_j - \left(\sum_{j\in\mathcal{N}} p_j \rho_j(x) - \frac{1}{\beta}\sum_{j\in\mathcal{N}} p_j \mu_j(x) \right) \right\} \\ &= \begin{cases} 0, & \text{if} \quad v_j = \rho_j(x) - \frac{1}{\beta}\mu_j(x), \forall j \in \mathcal{N} \\ -\infty, & \text{otherwise} \end{cases} . \end{aligned} \tag{46}$$

3.3.2. Nonlinear Uncertainty We now consider functions $f(x,p)$ that are nonlinear and concave in p. More precisely, the reward function $\mu(x,p)$ (43) is linear in p, while the risk function $\rho(x,p)$ is concave in p:

$$\rho(x,p) = \sqrt{\sum_{j\in\mathcal{N}} p_j \rho_j^2(x)} \ . \tag{47}$$

The modified reward-risk ratio function reads

$$f^N(x,p) = \sqrt{\sum_{j\in\mathcal{N}} p_j \rho_j^2(x)} - \frac{1}{\beta}\sum_{j\in\mathcal{N}} p_j \mu_j(x), \quad \forall p \in \mathcal{D}'_W \ , \tag{48}$$

and its concave conjugate is:

$$f^N_*(x,v) = \sup_w \quad -\frac{w}{4} \tag{49}$$

$$\text{s.t.} \quad \left\| \begin{bmatrix} 2\rho_j(x) \\ v_j + \frac{1}{\beta}\mu_j(x) - w \end{bmatrix} \right\|_2 \le v_j + \frac{1}{\beta}\mu_j(x) + w, \qquad \forall j \in \mathcal{N} \tag{50}$$

$$v_j + \frac{1}{\beta}\mu_j(x) \ge 0, \qquad \forall j \in \mathcal{N} \tag{51}$$

$$w \ge 0 \ . \tag{52}$$

Observe that, for any positive β, the function $f^N(x,p)$ with reward and risk functions taking form (43) and (47) is concave in p.

3.4. Reformulations in Finite Dimensional Constraint Space

The results presented in the previous subsection allow us to rewrite the constraints (35) in a finite dimensional constraint space using (36). The support function **WA-D** of the Wasserstein ambiguity set is obtained through the solution of the linear programming problem given in Theorem 5, and the concave conjugates of the modified reward-risk ratio functions are given in Section 3.3.

We first consider the constraints corresponding to the modified reward-risk ratio function (45) that is linear in p.

THEOREM 6. *The robust constraint*

$$\sum_{j\in\mathcal{N}} p_j\rho_j(x) - \frac{1}{\beta}\sum_{j\in\mathcal{N}} p_j\mu_j(x) \le 0, \quad \forall p \in \mathcal{D}'_W$$

corresponding to the modified reward-risk ratio function $f^L(x,v)$ (45) *can be reformulated into the following finite dimensional constraint space:*

$$\begin{aligned} \textbf{DRR-L}: \quad & \gamma\theta + \sum_{i\in\mathcal{N}} y_i p_i^0 \le 0 && (53)\\ & v_j + \frac{1}{\beta}\mu_j(x) = \rho_j(x), \qquad \forall j\in\mathcal{N} && (54)\\ & (38)-(39)\ . \end{aligned}$$

Proof: It follows from (42) that the optimal value of problem **WA-D** must be nonpositive, which is enforced via (53) and (38)-(39). Constraint (54) is due to the form of the concave conjugate $f^L_*(x,v)$ (46) of $f^L(x,v)$. □

We proceed similarly for the constraints corresponding to the modified reward-risk ratio function (48) that is nonlinear in p.

THEOREM 7. *The robust constraint*

$$f^N(x,p) = \sqrt{\sum_{j\in\mathcal{N}} p_j\rho_j^2(x)} - \frac{1}{\beta}\sum_{j\in\mathcal{N}} p_j\mu_j(x), \quad \forall p\in\mathcal{D}'_W$$

corresponding to the modified reward-risk ratio function $f^N(x,v)$ (48) *can be reformulated into the following finite dimensional constraint space:*

$$\begin{aligned} \textbf{DRR-N}: \quad & \gamma\theta + \sum_{i\in\mathcal{N}} y_i p_i^0 + \frac{w}{4} \le 0 && (55)\\ & (38)-(39); (50)-(52)\ . \end{aligned}$$

Proof: The optimal value of **WA-D** is nonpositive due to (42), which is ensured with (55) and (38)-(39). Constraints (50)-(52) are due to the form of $f^N_*(x,v)$. □

4. Risk-Adjusted Return Financial Performance Measures

The class of problem **DRR** has direct applications in finance when risk-adjusted return measures are used to build or rebalance a portfolio of securities. We assume here that the underlying probability measures have known atoms (identical to those of the reference distribution) whose probabilities are unknown. One may argue about this assumption given that any future observation (i.e., realized return) is likely to differ from historical observations. Some DRO financial studies (see, e.g., Pflug and Wozabal (2007)) justify this assumption by arguing that the atoms of the reference distribution are sufficiently dense (i.e., the sample size of the historical data is large enough) to represent all possible realizations of ξ. The so-called fixed support assumption has been used in multiple previous DRO studies, such as, Pflug and Wozabal (2007) with Wasserstein distance, Jiang and Guan (2015) with ϕ-divergence, and Postek et al. (2016) with various combinations of ambiguity sets and risk measures. A similar setting has also been extensively studied in stochastic programming (see, e.g., Gaivoronski and Pflug (2005) for Value-at-Risk and Rockafellar and Uryasev (2002) for Conditional Value-at-Risk).

We now consider some of the most commonly used reward-risk performance ratios in finance. We provide new models that permit to construct the portfolio generating the highest worst-case reward-risk level. In the financial context, the polytope $\mathcal{X}$ defining the feasible deterministic set is: $\mathcal{X} = \left\{x \in \mathbb{R}^M_+ : \sum_{m\in\mathcal{M}} x_m = 1\right\}$. The decision variable x_m is the proportion of capital invested in security m, and M is the number of securities. The above constraint ensures that the entire capital is invested, while the nonnegativity of x prevents short-selling. Next, we study the DRO problems with ambiguous Sharpe (Section 4.1) and Omega (Section 4.2) ratios used as performance measures.

4.1. Sharpe Ratio

The Sharpe ratio (Sharpe 1966), also called reward-to-variability ratio, is the quotient of the excess portfolio return (i.e., with respect to the fixed return of the risk-free asset) to the standard deviation of the portfolio return. Without loss of generality, we set the return of the risk-free asset equal to 0 in order to ease the notations. We denote by Y^j_m the return of asset m in scenario j, and define

$$R_j(x) = \sum_{m\in\mathcal{M}} x_m Y^j_m \tag{56}$$

as the portfolio return $R_j(x)$ in scenario j. The robust constraint on the ambiguous Sharpe ratio is:

$$\alpha^S(x,p) = \frac{\sum\limits_{j\in\mathcal{N}} p_j R_j(x)}{\sqrt{\sum\limits_{j\in\mathcal{N}} p_j \left(R_j(x) - \sum\limits_{j'\in\mathcal{N}} p_{j'} R_{j'}(x)\right)^2}} \geq \beta, \quad \forall p \in \mathcal{D}'_W \ . \tag{57}$$

Let β be fixed to some arbitrary positive value. The constraint corresponding to the modified Sharpe ratio function $f^S(x,p)$ derived from (42) reads

$$f^S(x,p) = \sqrt{\sum_{j\in\mathcal{N}} p_j \left(R_j(x) - \sum_{j'\in\mathcal{N}} p_{j'} R_{j'}(x)\right)^2} - \frac{1}{\beta}\sum_{j\in\mathcal{N}} p_j R_j(x) \leq 0, \qquad \forall p \in \mathcal{D}'_W$$

$$= \inf_{\kappa\in\mathbb{R}} \sqrt{\sum_{j\in\mathcal{N}} p_j \left(R_j(x) - \kappa\right)^2} - \frac{1}{\beta}\sum_{j\in\mathcal{N}} p_j R_j(x) \leq 0, \qquad \forall p \in \mathcal{D}'_W \tag{58}$$

where κ is an auxiliary decision variable. The minimum of $\sqrt{\sum_{j\in\mathcal{N}} p_j \left(R_j(x) - \kappa\right)^2}$ is the standard deviation and obtained when κ equals the mean return $\sum_{j\in\mathcal{N}} p_j R_j(x)$. Additionally, the inf expression in (58) can be dropped since it appears on the left-hand side of a stochastic inequality with "$\leq$" form, which gives:

$$f^S(x,p) = \sqrt{\sum_{j\in\mathcal{N}} p_j \left(R_j(x) - \kappa\right)^2} - \frac{1}{\beta}\sum_{j\in\mathcal{N}} p_j R_j(x) \leq 0, \quad \forall p \in \mathcal{D}'_W \ . \tag{59}$$

The modified reward-risk function $f^S(x,p)$ is nonlinear and concave in p for any $\beta > 0$; $f^S(x,p)$ takes the form of (48) by setting $\mu_j(x) = R_j(x)$ and $\rho_j(x) = R_j(x) - \kappa$. Using **DRR-N** and with appropriate notational substitutions, we provide an equivalent reformulation

$$\textbf{DRR-S}: \quad \left\| \begin{bmatrix} 2(R_j(x) - \kappa) \\ v_j + \frac{1}{\beta}R_j(x) - w \end{bmatrix} \right\|_2 \leq v_j + \frac{1}{\beta}R_j(x) + w, \qquad \forall j \in \mathcal{N} \tag{60}$$

$$v_j \geq -\frac{1}{\beta}R_j(x), \qquad \forall j \in \mathcal{N} \tag{61}$$

$$(38) - (39); (52); (55) - (56)$$

of the feasible set of the ambiguous Sharpe ratio constraint. The decision variables are $\gamma, y, w, v, \kappa, x$. The inequality (61) is obtained by replacing $\mu_j(x)$ by $R_j(x)$ in (51).

Next, we derive in Lemma 1 cuts and valid inequalities to tighten the formulation of **DRR-S**. As shown in Section 6, the strengthening approach significantly speeds up the solution of **DRR-S**.

LEMMA 1. *Let* $R^L_j := \min_{m\in\mathcal{M}} Y^j_m$ *and* $R^U_j := \max_{m\in\mathcal{M}} Y^j_m$ *denote the minimal and maximal portfolio return in scenario* j, $\forall j \in \mathcal{N}$. *For any* $\beta > 0$, *the inequalities*

$$-\frac{R^U_j}{\beta} \leq v_j \leq -\frac{R^L_j}{\beta}\ , \quad \forall j \in \mathcal{N} \tag{62}$$

$$\sum_{i\in\mathcal{N}} y_i \leq 0 \tag{63}$$

tighten the feasible region of **DRR-S**.

Proof: Since $0 \leq x_m \leq 1, \forall m \in \mathcal{M}$, the portfolio return $R_j(x)$ is bounded: $R_j(x) \in [R^L_j, R^U_j]$. Replacing $R_j(x)$ by its upper bound in (61), we obtain the valid inequality $v_j \geq -\frac{R^U_j}{\beta}$, which is the left side of (62). The largest value that v_j can be forced to take by (61) is $-\frac{R_j^L}{\beta}$. Besides (61), (38) is the only constraint in which v_j appears and is such that the lower the value of v_j, the easier (38) holds. It follows that the constraint $v_j \leq -R_j^L/\beta$ in (62) is a cut that reduce the domain of v_j but does not discard any optimal solution.

Consider now (55). Since θ is a nonnegative constant, $\gamma, \omega \geq 0$, and $p_i^0 = 1/N > 0, \forall i \in \mathcal{N}$, the left-hand side of (55) can only be nonpositive if (63) holds, which defines a valid inequality. $\square$

4.2. Omega Ratio

The Omega ratio (Keating and Shadwick 2002) belongs to the group of upside / downside ratio performance measures. Let τ be the target return value set by the investor and with respect to which the portfolio upside (i.e., excess return) and downside (i.e., return shortfall) are computed. The ambiguous Omega ratio takes the form:

$$\alpha^O(x,p) = \frac{\sum\limits_{j\in\mathcal{N}} p_j \cdot (\max\{R_j(x) - \tau, 0\})}{\sum\limits_{j\in\mathcal{N}} p_j \cdot (\max\{\tau - R_j(x), 0\})} \geq \beta, \quad \forall p \in \mathcal{D}'_W \ . \tag{64}$$

The robust constraints for the Omega ratio read:

$$\sum_{j\in\mathcal{N}} p_j \left(\max\{\tau - R_j(x), 0\}\right) - \frac{1}{\beta}\left(\sum_{j\in\mathcal{N}} p_j \left(\max\{R_j(x) - \tau, 0\}\right)\right) \leq 0, \forall p \in \mathcal{D}'_W \ . \tag{65}$$

Let d_j^+ and d_j^- respectively denote the above- and below-target deviations: $d_j^+ = \max\{R_j(x) - \tau, 0\}$ and $d_j^- = \max\{\tau - R_j(x), 0\}, \forall j \in \mathcal{N}$. We can rewrite (65) as:

$$f^O(x,p) = \sum_{j\in\mathcal{N}} p_j d_j^- - \frac{1}{\beta}\left(\sum_{j\in\mathcal{N}} p_j d_j^+\right) \leq 0, \qquad \forall p \in \mathcal{D}'_W \tag{66}$$

$$d_j^+ \geq R_j(x) - \tau, \qquad \forall j \in \mathcal{N} \tag{67}$$

$$d_j^- \geq \tau - R_j(x), \qquad \forall j \in \mathcal{N} \tag{68}$$

$$d_j^+, d_j^- \geq 0, \qquad \forall j \in \mathcal{N} \ . \tag{69}$$

The modified Omega ratio function $f^O(x,p)$ is linear in p for β fixed to an arbitrary positive value and is in the form of (45) with $\mu_j(x) = d_j^+ = \max\{R_j(x) - \tau, 0\}, \forall j \in \mathcal{N}$ and $\rho_j(x) = d_j^- = \max\{\tau - R_j(x), 0\}, \forall j \in \mathcal{N}$. Using **DRR-L**, we obtain the inequalities

$$\textbf{DRR-O}: \quad ||\xi_j - \xi_i^0||\gamma + y_i \geq d_j^- - \frac{1}{\beta} d_j^+, \qquad \forall i, j \in \mathcal{N} \tag{70}$$

$$(39); (53); (56); (67) - (69)$$

that define a feasible set equivalent to the one of the ambiguous Omega ratio constraint. The decision variables are $\gamma, y, d^+, d^-, x, \beta$, and v_j is replaced by $d_j^- - \frac{1}{\beta} d_j^+$. As for the Sharpe ratio, we now tighten the feasible set of **DRR-O** via the introduction of valid inequalities.

Lemma 2. *Let $R_j^L := \min_{m \in \mathcal{M}} Y_m^j$ and $R_j^U := \max_{m \in \mathcal{M}} Y_m^j$ denote the minimal and maximal portfolio return in scenario j, $\forall j \in \mathcal{N}$. The following inequalities are valid for* **DRR-O**:

$$d_j^+ \geq R_j^L - \gamma, \qquad \forall j \in \mathcal{N} \tag{71}$$

$$d_j^- \geq \gamma - R_j^U, \qquad \forall j \in \mathcal{N} \tag{72}$$

$$(63)\ .$$

The proof is similar to the one presented for Lemma 1.

5. Bisection Algorithmic Methods

As above-mentioned, since β is a decision variable, the reformulations of the **DRR** reward-risk ratio problems are nonconvex and it would be very challenging (see Section 6.2) to solve them directly. Instead, we develop bisection algorithms that provide the exact optimal value via the solution of a sequence of convex programming problems.

Consider a maximization problem with a continuous objective function defined on $[\beta_L^0, \beta_U^0]$. A bisection algorithm divides at each iteration t the incumbent interval $[\beta_L^t, \beta_U^t]$ in two equal parts. It computes the value of the objective function at the midpoint $\beta^t = (\beta_L^t + \beta_U^t)/2$ of the interval. If the problem is feasible at β^t, there is potentially a better feasible solution larger than β^t. The interval $[\beta_L^t, \beta^t)$ can be discarded and the search continues on $[\beta^t, \beta_U^t]$. Otherwise, if the problem is infeasible at β^t, any value larger than or equal to β^t can be dropped and the next iterations focus on $[\beta_L^t, \beta^t)$. The process continues until the interval is sufficiently small. We implement this idea for the problem at hand. We first present in Section 5.1 the details of the *standard bisection algorithm (SBA)*. Building on that, we propose in Section 5.2 the *bisection algorithm with interval compaction (BAIC)* that improves the efficiency of the search by compacting the search intervals.

5.1. Standard Bisection Algorithm

For any arbitrary value assigned to β, it is easy to see that the feasible sets **DRR-L** and **DRR-N** of the generic formulations are convex. Additionally, the objective function of **DRR** is nondecreasing in β. These are key properties that motivated us to resort to a bisection search procedure. To describe the algorithmic procedure, we consider the case of $f(x,p)$ nonlinear and concave in p. The exact same procedure applies to the case of $f(x,p)$ linear in p (with slightly different notations).

Problem **DRR** with $f(x,p)$ nonlinear in p has feasible set **DRR-N** and can be recast as:

$$\begin{aligned} \mathbf{DRR} : \max \quad & \beta \\ s.t. \quad & (x,v,y,w,\gamma,\beta) \in \mathbf{DRR\text{-}N} \\ & (4) - (5)\ . \end{aligned}$$

Let β^* be the optimal solution of **DRR** and $[\beta_L^0, \beta_U^0], \beta_L^0 \geq 0$ be the domain on which β is initially defined. For the algorithm to converge, we must set β_L^0 and β_U^0 so that $\beta^* \in [\beta_L^0, \beta_U^0]$. We denote by $[\beta_L^t, \beta_U^t]$ the updated interval on which β can take value at the incumbent iteration t.

The bisection algorithm solves at each iteration t a convex feasibility problem in which β is fixed to the midpoint β^t of the incumbent interval $[\beta_L^t, \beta_U^t]$. The feasibility problem $\textbf{DRR-F}_t$ at t is:

$$\textbf{DRR-F}_t : \text{Find } (x, v, y, w, \gamma) \tag{73}$$

$$s.t. \qquad (x, v, y, w, \gamma) \in \textbf{DRR-N} \tag{74}$$

$$\beta = \beta^t \tag{75}$$

$$(4)\ .$$

The only goal is to check whether $\textbf{DRR-F}_t$ admits a feasible solution. The tolerance level for the stopping criterion is ε. The iterative procedure stops if $\beta_U^t - \beta_L^t \leq \varepsilon$. Let s_t be a variable storing the best value known at iteration t.

A few comments are in order. First, the determination of suitable values for β_L^0 and β_U^0 ensures that the optimal value and solution β^* belongs to each of the successive intervals $[\beta_L^t, \beta_U^t]$. Second, the size of the updated interval $[\beta_L^t, \beta_U^t]$ shrinks at each iteration by a factor of $(\beta_U^t - \beta_L^t)/2$, and the algorithm therefore converges to the optimal value. Finally, this is a descent algorithm: the sequence $\{s_t\}$ is monotone increasing. The next theorem summarizes these features.

THEOREM 8. *Let $\beta^* \in [\beta_L^0, \beta_U^0]$. The standard bisection algorithm finds the optimal solution β^* of* **DRR** *in finitely many (i.e., at most* $\log_2((\beta_U^0 - \beta_L^0)/\varepsilon)$ *iterations with precision level ε.*

The proof of Theorem 8 is provided in Electronic Companion EC.1.

5.2. Bisection Algorithm with Interval Compaction

The computational efficiency of the standard bisection algorithm is highly impacted by the length of the search intervals $[\beta_L^t, \beta_U^t], t = 0, \dots, T$. We develop a bisection algorithm that integrates two schemes to compact the search interval. The *a priori* interval compaction procedure (Section 5.2.1) reduces the size of the initial interval, while the *iterative* interval compaction procedure (Section 5.2.2) is carried out on each of the successive intervals. The proposed *a priori* and *iterative* interval compaction methods are independent and *modular*; they can be implemented individually or jointly.

5.2.1. A Priori Interval Compaction The objective is to find a small initial search interval that contains with certainty the optimal value of β. The lower bound β_L^0 can be set to zero due to the non-negativity of the reward and risk measures. The upper bound β_U^0 can be set to an arbitrary large positive number (e.g., a high reward-risk ratio value that can not be achieved in practice) to ensure that the interval $[\beta_L^0, \beta_U^0]$ contains the optimal value of β. The a priori interval compaction method determines a tight upper bound β_U for β, thereby initializing the bisection algorithm with a narrower search region, which in turn will reduce the number of iterations.

Naturally, the reward-risk ratio $\mu(x,p)/\rho(x,p)$ is upper bounded by the maximal reward value divided by the minimal risk value:

$$\beta \leq \frac{\mu(x,p)}{\rho(x,p)}, \forall p \in \mathcal{D}'_W \implies \quad \beta \leq \frac{\max\limits_{p \in \mathcal{D}'_W} \mu(x,p)}{\min\limits_{p \in \mathcal{D}'_W} \rho(x,p)} = \frac{\max\{\beta_\mu | \exists p \in \mathcal{D}'_W, \text{ such that } \mu(x,p) \geq \beta_\mu\}}{\min\{\beta_\rho | \exists p \in \mathcal{D}'_W, \text{ such that } \rho(x,p) \leq \beta_\rho\}} = \beta_U \ ,$$

where β_μ and β_ρ are non-negative and denote the largest possible reward and the smallest risk.

The computation of the upper bound β_U on β can be carried out through the solution of two sub-problems: (1) Maximizing the robust reward measure (i.e., portfolio return) without restriction on the risk measure; (2) Minimizing the robust risk measure (i.e., standard deviation) without restriction on the reward.

Next, we solve the two corresponding DRO problems by adapting the reformulation framework presented in Section 3. We only need to adjust the reward and risk functions appropriately into formulation (35) and derive the corresponding conjugate functions, while the support function of the Wasserstein ambiguity set remains the same as in **WA-D**. We give below the formulations of the two problems and give the details of their reformulation in Electronic Companion EC.2. The generic problem maximizing the robust reward measure with Wasserstein ambiguity is:

$$\begin{aligned}
\textbf{DRR-}\boldsymbol{\mu}\textbf{:} \ \max \quad & \beta_\mu && && (76)\\
s.t. \quad & -\mu(x,p) \leq -\beta_\mu, && \forall p \in \mathcal{D}'_W && (77)\\
& \beta_\mu \in \mathbb{R}_+ && && (78)\\
& (4); (31)-(34) \ .
\end{aligned}$$

The two sides of (77) are written with negative sign in order to fit the form of (35) in Theorem 4.

The generic problem minimizing the risk measure with Wasserstein ambiguity is:

$$\begin{aligned}
\textbf{DRR-}\boldsymbol{\rho}\textbf{:} \ \min \quad & \beta_\rho && && (79)\\
s.t. \quad & \rho(x,p) \leq \beta_\rho, && \forall p \in \mathcal{D}'_W && (80)\\
& \beta_\rho \in \mathbb{R}_+ && && (81)\\
& (4); (31)-(34) \ .
\end{aligned}$$

The solution of problems **DRR-Rμ** and **DRR-Rρ** provides the maximal robust reward value β^*_μ and the minimal robust risk value β^*_ρ, and the value of the upper bound $\beta_U = \frac{\beta^*_\mu}{\beta^*_\rho}$.

5.2.2. Iterative Interval Compaction The objective of the bisection algorithm with iterative interval compaction method is to derive a better lower bound β_L^t for β at each iteration t, thereby narrowing down the search region at each iteration, reducing the number of iterations, and speeding up convergence. Similar to the standard bisection algorithm, the bisection algorithm with iterative compaction starts with the solution of the convex feasibility problem **DRR-F**$_t$ where β is fixed equal to $\beta^t = (\beta_L^t + \beta_U^t)/2$. If the problem is feasible, the bisection algorithm with iterative interval compaction calls a procedure to check whether the feasibility of **DRR-F**$_t$ can be maintained if β is set to a value larger than β^t.

Assume that (x, v, y, w, γ) with $\beta = \beta^t$ is feasible for **DRR-F**$_t$. The algorithm will then verify whether β can take a value larger than β^t by examining the status (i.e., binding or not) of the constraints in which β appears (i.e., (50) and (51)) in the constraint set **DRR-N**. If at least one of these constraints is binding, then setting $\beta > \beta^t$ would violate the binding constraint(s). However, if all inequality constraints involving β are not binding, β can be set to a larger value β'^t (than β^t) without violating any constraints. This compresses the search region $[\beta'^t, \beta_U^t] \subset [\beta^t, \beta_U^t]$.

Assume that at any arbitrary t, (x, v, y, w, γ) is feasible for **DRR-F**$_t$ and that no inequality constraint involving β is binding. We can then derive a better lower bound for β by solving

$$
\begin{aligned}
\max \quad & \beta' && && (82)\\
\text{s.t.} \quad & \left\| \begin{bmatrix} 2\rho_j(x) \\ v_j + \frac{1}{\beta'}\mu_j(x) - w \end{bmatrix} \right\|_2 \leq v_j + \frac{1}{\beta'}\mu_j(x) + w, && \forall j \in \mathcal{N} && (83)\\
& v_j + \frac{1}{\beta'}\mu_j(x) \geq 0, && \forall j \in \mathcal{N} && (84)
\end{aligned}
$$

where β' is a decision variable and v_j, $\mu_j(x)$, w, $\rho_j(x)$ are parameters computed according to the obtained feasible solution (x, v, y, w). The above problem (82)-(84), in which the decision variable β' appears in the denominator of the left-hand side of constraints (83), is in general nonconvex. However, setting $\alpha' = 1/\beta'$ gives the equivalent convex programming problem **DRR-RF**$_t$:

$$
\begin{aligned}
\textbf{DRR-RF}_t : \min \quad & \alpha' && && (85)\\
\text{s.t.} \quad & \left\| \begin{bmatrix} 2\rho_j(x) \\ v_j + \alpha'\mu_j(x) - w \end{bmatrix} \right\|_2 \leq v_j + \alpha'\mu_j(x) + w, && \forall j \in \mathcal{N} && (86)\\
& v_j + \alpha'\mu_j(x) \geq 0, && \forall j \in \mathcal{N} && (87)
\end{aligned}
$$

We only solve **DRR-RF**$_t$ if none of constraints involving β is binding at the incumbent solution of **DRR-F**$_t$. The pseudo-code of the bisection algorithm with interval compaction is provided in Electronic Companion EC.3.

The bisection algorithm with interval compaction converges to the optimal value of **DRR-N** in a finite numner of iterations with precision level ϵ. It involves the solution of the two optimization problems **DRR-F**$_t$ and **DRR-RF**$_t$ at each iteration if none of the inequality constraints of **DRR-F**$_t$ in which β appears is binding. In the next section, we analyze the computational efficiency of several variants of the bisection algorithm approach in which the a priori and iterative interval compaction methods are implemented separately and jointly.

6. Computational Tests

We carried out computational tests based on real financial data and problems of industry-relevant size to test the applicability and scalability of the reformulation and algorithmic framework. We assessed the efficiency of the bisection algorithms and derived insights pertaining to the impact of the number of assets and number of data points on the scalability of the method. Finally, an out-of-sample analysis was conducted to test the performance of the ambiguous Sharpe ratio portfolios. We compared the portfolios constructed with the ambiguous Sharpe ratio model, the $1/N$ portfolio strategy (DeMiguel et al. 2009b), and the ambiguity-free Sharpe ratio model.

6.1. Data and Experimental Design

Our dataset comprised weekly prices of 400 securities included in the Standard & Poor's 500 (S&P 500) index from 01/07/2000 to 12/30/2005 and for which there was no missing data. The securities which were included/delisted during that period were removed. We created 24 problem instances by specifying three levels for M (i.e., $25, 100, 400$) and four levels for N (i.e. $24, 60, 120, 180$), and two levels for ε (0.01 and 0.001). We solved the 24 corresponding DRO Sharpe ratio problems with the proposed bisection algorithms. We used the ambiguous Sharpe ratio model (Section 4.1) for two main reasons. First, the Sharpe ratio is the performance ratio most commonly used in the finance industry. Second, among the ambiguous reward-risk ratios in Section 4, the Sharpe ratio problem is the more complex one, and hence more suitable, to test the efficiency of our approach.

We used two schemes to set the initial upper bound β_U^0 for β. First, we used an arbitrary (and unrealistically) large value for the upper bound β_U^0 and the initial interval $[\beta_L^0, \beta_U^0]$ was set to be $[0, 5]$. Setting the lower bound to zero was due to the assumed non-negativity of the reward and risk measures. Second, we used the a priori compaction approach presented in Section 5.2.1 and in which the ambiguous Sharpe ratio was upper-bounded by the ratio of the maximal robust mean return to the minimal robust standard deviation. We incorporated the bounds and valid inequalities presented in Lemma 1 within each bisection algorithm. We used AMPL to formulate the problems. Each problem instance was solved with Cplex 12.8.0.0 solver on a 64-bit desktop with Intel(R) Core(TM) i5-6500 processor, running at 3.2GHz CPU and with 8GB RAM.

6.2. Computational Assessment

We assessed the scalability and computational efficiency of the algorithmic framework. All tests were conducted with a probability level q equal to 0.95. We reported for each algorithm the optimal values of β, the number of iterations (labeled # Iter), the CPU time (labeled TotalCPU) in seconds, and the average CPU time per iteration (labeled AveCPU) in seconds. We denoted by *SBA* the standard bisection algorithm, by *BAIC-IC* the bisection algorithm that used the iterative interval compaction method, by *BAIC-PC* the one that used the a priori interval compaction method, and by *BAIC* the one that used both the a priori and iterative interval compaction methods. We had also tried to solve directly the ambiguous Sharpe ratio model **DRR** with the MINLP solver Baron 17.1.2. However, we could not solve any of the 24 problem instances to optimality within 24 hours.

We first derived some insights regarding the impact of the number of assets M and the number of data points N on the solution times. Figure 1 displays for each algorithm the average CPU time in terms of the number of assets. With the algorithms *SBA* and *BAIC-PC*, an increase in the number M of assets led to an increase in the solution time. For example, with *SBA*, the average time for the instances with 400 assets was respectively 84.97% and 17.24% larger than the one for the instances with 25 and 100 assets. For the algorithms *BAIC-PC* and *BAIC* using a priori compaction, the solution time increased when the number of assets M increased from 25 to 100; however, moving from 100 to 400 assets did not trigger a time increase, which allows these two methods to handle larger portfolios if necessary. The minor impact of the size of the asset universe on the computational time follows from the features of the reformulation. Indeed, the number of assets M has a linear impact on the number of decision variables x_m (i.e., one decision variable for each asset) and the number of constraints in the overall reformulation is invariant with the number of assets. The remarkable scalability of our algorithmic procedure allows for the efficient solution of problems with asset universe comprising up to 400 securities. The size of these instances is much larger than the numerical financial illustrations that have so far been presented in the distributionally robust literature, in which - to our knowledge - the largest problem instances consider 100 (Wozabal 2014) or 80 securities (Xu et al. 2018). Most DRO financial illustrations consider no more than 15 assets (see, e.g., Pflug and Wozabal (2007), Wozabal (2012), Esfahani and Kuhn (2017), Postek et al. (2016). This is even more striking given that our reformulation is nonconvex, while the reformulations of the above-mentioned earlier studies are convex. The importance of the scalability of our approach with respect to the asset universe size is further illustrated in Section 6.3.

The solution time increased with the number of data points N. For example, with *SBA*, switching from 60 to 120 (resp., 180) data points, raised the average time per iteration from 0.24 to 3.15

(resp., 385.03) seconds. This observation could be extended to the other algorithms as well. The impact of N on the solution time was due to the number of constraints which was quadratic in N (see constraint (60)). As expected, the solution time decreases with ϵ and becomes larger when the precision level is smaller. More detailed results are reported in Table EC.4.

We used Figures 2, 3, and 4 to compare the four bisection algorithm variants and to study the benefits of two (a priori and iterative) compaction schemes. In particular, we checked the impact on the number of iterations, time per iteration, and solution time. Figure 2 illustrates the effect of the a priori compaction method on the value of the initial upper bound of the search interval. Figure 3 and 4 display the average CPU time per iteration and the average total CPU time for each algorithm. The detailed results underlying Figures 2, 3, and 4 are in Table EC.4.

The a priori compaction method decreased both the number of iterations and the average time per iteration, while the iterative compaction method reduced the number of iterations but increased the iteration time. In general, *BAIC-IC* took more time per iteration, but involved less iterations than *SBA*, which led to an overall smaller solution time. The reduction in the number of iterations more than compensated for the increase in iteration time. The following example illustrates this observation. For the instance with $M = 100, N = 100$ and $\epsilon = 0.001$, the standard bisection algorithm found the optimal solution after 5025.66 seconds (12 iterations with 418.80s per iteration on average), while it took 4724.51 seconds for *BAIC-IC* (11 iterations with 429.50 seconds per iteration). The *BAIC-IC* algorithm eventually saved 1 iterations (11 vs. 12) and 5.99% of the total time, although the average time per iteration increased by 2.55% compared to *SBA*.

As shown in Figure 4, each algorithm incorporating one or more interval compaction methods was faster than the standard bisection algorithm *SBA*. The fastest algorithm *BAIC* used both (a priori and iterative) compaction methods. On average, it was 6.75% quicker than the second fastest algorithm *BAIC-PC* and 29.94% faster than *SBA*. *BAIC* solved all but one of the 24 instances in less than one hour and obtained the optimal solution in less than one minute for all instances with up to 120 data points. The incorporation of the bounds and valid inequalities in Lemma 1 played a significant role in reducing the computational times. To give an example, for the instance $M = 100$, $N = 180$, and $\epsilon = 0.01$, the solution time with *SBA* was 7497 seconds without the bounds and valid inequalities and 3237 seconds with them, which corresponded to a 56.65% time saving.

6.3. Out-of-Sample Performance Evaluation

We conducted an empirical analysis to assess the out-of-sample performance of three portfolios constructed with different Sharpe ratio based models. The first portfolio $\mathbf{P}_{q=0.95}$ was obtained by solving model **DRRW** with probability level $q = 0.95$. The second one $\mathbf{P}_{\mathbb{Q}=\mathbb{P}_0}$, called nominal portfolio, was obtained by solving the ambiguity-free model **DRRQ** that assumes that the reference

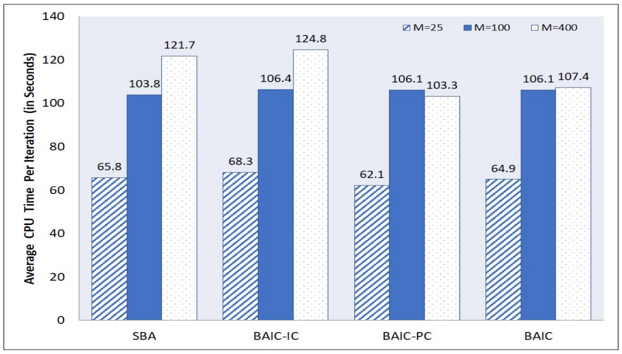

Figure 1 Average Time and Number of Assets M

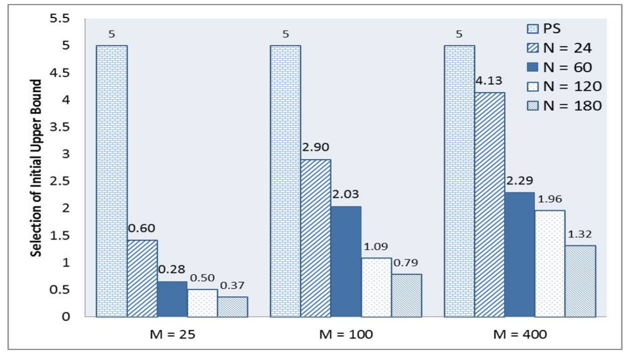

Figure 2 Initial Upper Bound of Each Problem Instance

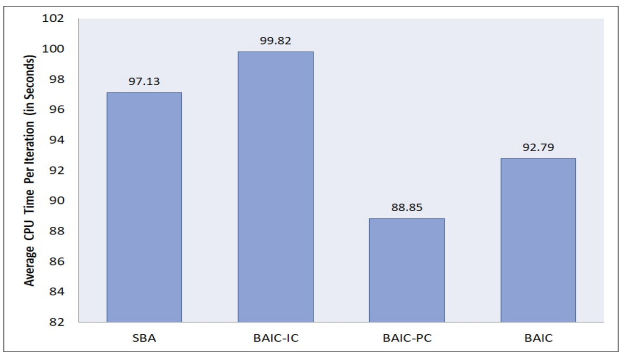

Figure 3 Average Time per Iteration

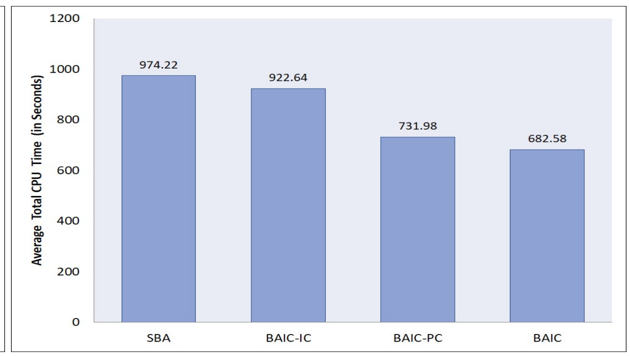

Figure 4 Average Total Time

distribution is the true one. The third one, called $1/N$ portfolio, was constructed using the $1/N$ equally weighted allocation strategy DeMiguel et al. (2009b). The nominal and the $1/N$ portfolios served as benchmarks for the distributionally robust portfolio $\mathbf{P}_{q=0.95}$. As in DeMiguel et al. (2009a), we employed a rolling-horizon procedure to rebalance the portfolio on a weekly basis. First, we constructed three datasets including $M = 25, 100, 400$ assets and corresponding to small, medium, and large asset universe, covering the weekly returns from 01/07/2000 to 12/30/2005 with a total number of periods T equal to 313. We chose a window of training periods $\zeta = 52$ corresponding to one year data points. Second, using the return data over the training periods ζ, we built the three portfolios (i.e., $\mathbf{P}_{q=0.95}$, $\mathbf{P}_{\mathbb{Q}=\mathbb{P}_0}$, $1/N$). Third, we repeated this rolling-horizon procedure for the next week by adding the data for the next week and dropping the data for the earliest week in the testing period. We continued this procedure until the end of the testing period (12/30/2005). At the end of this procedure, we had generated $T - \zeta = 261$ portfolios for each problem instance ($M = 25, 100, 400$) and model ($\mathbf{P}_{q=0.95}$, $\mathbf{P}_{\mathbb{Q}=\mathbb{P}_0}$ and $1/N$).

Let $r_{mt}, m = 1, \dots, M$ and $t = 1, \dots, T$ denote the return of asset m in testing period t. The out-of-sample portfolio return is $R_t = \sum_{m=1}^{M} r_{mt} x_m$. We used the time series of portfolio returns to compute the following metrics: (i) out-of-sample Sharpe ratio $SR = \frac{\mu}{\sigma}$; (ii) out-of-sample cumulative return $CR = \prod_{t=\zeta}^{T-1}(1 + R_{t+1}) - 1$; (iii) percentage of time the distributionally robust Sharpe ratio portfolio beat the benchmark portfolio: $NB = \frac{1}{T-\zeta}\sum_{t=\zeta}^{T-1} s_t$, where $s_t = 1$ if the cumulative return CR_t of $\mathbf{P}_{q=0.95}$ was higher than the cumulative return of the benchmark portfolio.

Figures 5 and 6 plot the time-series cumulative returns of the three portfolios for asset universes of size $M = 100$ and $M = 400$, respectively. Table 1 summarizes the out-of-sample performance of the distributionally robust Sharpe ratio portfolio ($\mathbf{P}_{q=0.95}$), the nominal Sharpe ratio portfolio ($\mathbf{P}_{\mathbb{Q}=\mathbb{P}_0}$), and the equally weighted portfolio $1/N$ with respect to the above three metrics.

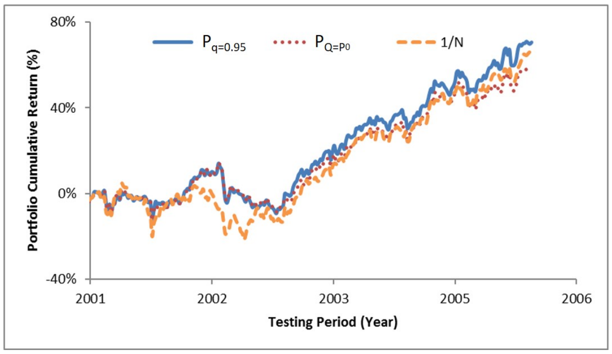

Figure 5 Time-Series Cumulative Return ($M = 100$)

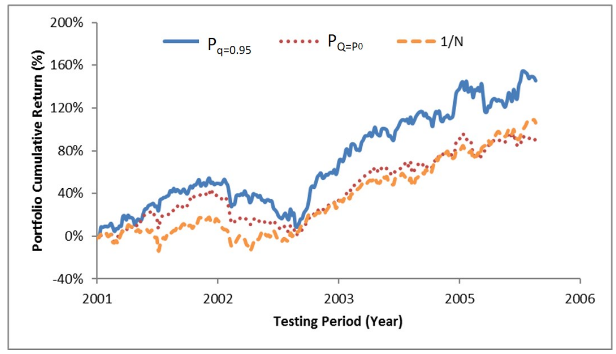

Figure 6 Time-Series Cumulative Return ($M = 400$)

Table 1 Out-of-Sample Portfolio Performance

Model Performance	$M = 25$			$M = 100$			$M = 400$		
	$\mathbf{P}_{q=0.95}$	$\mathbf{P}_{\mathbb{Q}=\mathbb{P}_0}$	$1/N$	$\mathbf{P}_{q=0.95}$	$\mathbf{P}_{\mathbb{Q}=\mathbb{P}_0}$	$1/N$	$\mathbf{P}_{q=0.95}$	$\mathbf{P}_{\mathbb{Q}=\mathbb{P}_0}$	$1/N$
SR	0.0935	0.0864	0.0712	0.1179	0.1020	0.0966	0.1460	0.1289	0.1311
TCR	61.05%	104.87%	41.61%	70.50%	57.41%	63.54%	145.35%	89.93%	105.99%
NB	–	35.35%	80.84%	–	75.48%	91.57%	–	100.00%	99.62%

For all datasets and asset universe sizes, the distributionally robust portfolio $\mathbf{P}_{q=0.95}$ outperformed both the nominal $\mathbf{P}_{\mathbb{Q}=\mathbb{P}_0}$ and the $1/N$ portfolios with respect to the out-of-sample Sharpe ratio (SR). The same observation prevailed for the other two metrics TCR and NB at the exception of the small asset universe (M=25) for which the nominal portfolio $\mathbf{P}_{\mathbb{Q}=\mathbb{P}_0}$ had the best terminal cumulative return. The $\mathbf{P}_{q=0.95}$ portfolio beat the $1/N$ one most of the time, i.e., in respectively 80.84%, 91.57%, and 99.62% of the cases for the small, medium, and large asset universes. Similarly, $\mathbf{P}_{q=0.95}$ beat $\mathbf{P}_{\mathbb{Q}=\mathbb{P}_0}$ 75.48% and 100% of the time for the medium and large asset universes. The **DRR** portfolio $\mathbf{P}_{q=0.95}$ dominated the nominal $\mathbf{P}_{\mathbb{Q}=\mathbb{P}_0}$ and the $1/N$ portfolios almost surely for large asset universes. In addition, $\mathbf{P}_{q=0.95}$ provided an out-of-sample terminal cumulative return TCR which was 46.72% (resp., 10.95% and 37.14%) larger than the TCR provided by the $1/N$ approach for asset universe of size 25 (resp., 40 and 100). Similarly, $\mathbf{P}_{q=0.95}$ provided an out-of-sample TCR which was 13.09% (resp., 55.42%) larger than the TCR of the $\mathbf{P}_{\mathbb{Q}=\mathbb{P}_0}$ portfolio for the small (resp., large) asset universe. The $\mathbf{P}_{q=0.95}$ portfolio generally dominated the other two investment strategies and that this superiority was increasingly marked as the size of the asset universe increases. This highlights the benefits of our algorithmic method which scales well and for which the solution time is not an increasing function of the number of assets (see Section 6.2).

7. Conclusion

We investigate a new class of distributionally robust optimization problems with stochastic fractional functions representing reward-risk ratios. We use a data-driven approach and derive a new fully parameterized closed-form expression that lower-bounds the size of the Wasserstein ambiguity ball. We develop an efficient and scalable reformulation and algorithmic framework. The reformulation phase involves the derivation of the support function of the ambiguity set and of the concave conjugate of the reward-risk function. We reformulate the semi-infinite problem **DRR** in a finite dimensional constraint space. We design a series of bisection algorithms that solve exactly and efficiently industry-sized problems and in which two types of compaction schemes can be integrated.

The computational study attests the great scalability of the method. We solve portfolio optimization problems comprising up to 400 securities, which is four times more then the number of assets considered in distributionally robust financial problems (see, e.g., Wozabal (2014), Xu et al. (2018)). The fastest algorithm solves 23 of the 24 instances in less than one hour, while no instance can be solved to optimality in one day with the Baron MINLP solver. Additionally, the solution time does not increase with the number of securities, which makes the approach applicable for portfolio problems with larger asset universe. The out-of-sample analysis illustrates the robustness of the **DRR** Sharpe ratio model and its superior performance compared to the $1/N$ equal allocation strategy and the ambiguity-free Sharpe ratio

References

Bacon CR (2012) *Practical Risk-adjusted Performance Measurement* (John Wiley & Sons).

Bailey DH, de Prado M (2012) The Sharpe ratio efficient frontier. *Journal of Risk* 15(2):3–44.

Ben-Tal A, Den Hertog D, De Waegenaere A, Melenberg B, Rennen G (2013) Robust solutions of optimization problems affected by uncertain probabilities. *Management Science* 59(2):341–357.

Ben-Tal A, Den Hertog D, Vial JP (2015) Deriving robust counterparts of nonlinear uncertain inequalities. *Mathematical Programming* 149(1-2):265–299.

Bolley F, Villani C (2005) Weighted Csiszár-Kullback-Pinsker inequalities and applications to transportation inequalities. *Annales de la Faculté des Sciences de Toulouse: Mathématiques*, volume 14, 331–352.

Bonami P, Lejeune MA (2009) An exact solution approach for portfolio optimization problems under stochastic and integer constraints. *Operations Research* 57(3):650–670.

Calafiore GC (2007) Ambiguous risk measures and optimal robust portfolios. *SIAM Journal on Optimization* 18(3):853–877.

Cheridito P, Kromer E (2013) Reward-risk ratios. *Journal of Investment Strategies* 3:1–16.

Delage E, Ye Y (2010) Distributionally robust optimization under moment uncertainty with application to data-driven problems. *Operations Research* 58(3):595–612.

DeMiguel V, Garlappi L, Nogales FJ, Uppal R (2009a) A generalized approach to portfolio optimization: Improving performance by constraining portfolio norms. *Management Science* 55(5):798–812.

DeMiguel V, Garlappi L, Uppal R (2009b) Optimal versus naive diversification: How inefficient is the 1/n portfolio strategy? *Review of Financial Studies* 22(5):1915–1953.

Deng G, Dulaney T, McCann C, Wang O (2013) Robust portfolio optimization with value-at-risk-adjusted Sharpe ratios. *Journal of Asset Management* 14(5):293–305.

Erdoğan E, Iyengar G (2006) Ambiguous chance constrained problems and robust optimization. *Mathematical Programming* 107(1-2):37–61.

Esfahani PM, Kuhn D (2017) Data-driven distributionally robust optimization using the Wasserstein metric: Performance guarantees and tractable reformulations. *Mathematical Programming* 1–52.

Fournier N (2018) Personal/private email communication (Jan 13th, 2018) .

Fournier N, Guillin A (2015) On the rate of convergence in Wasserstein distance of the empirical measure. *Probability Theory and Related Fields* 162(3-4):707–738.

Gaivoronski AA, Pflug G (2005) Value-at-risk in portfolio optimization: Properties and computational approach. *Journal of Risk* 7(2):1–31.

Gao R, Kleywegt AJ (2016) Distributionally robust stochastic optimization with Wasserstein distance, Working Paper: Available at Optimization Online.

Goh J, Sim M (2011) Robust optimization made easy with ROME. *Operations Research* 59(4):973–985.

Goldfarb D, Iyengar G (2003) Robust portfolio selection problems. *Mathematics of Operations Research* 28(1):1–38.

Hanasusanto GA, Roitch V, Kuhn D, Wiesemann W (2015) A distributionally robust perspective on uncertainty quantification and chance constrained programming. *Mathematical Programming* 151(1):35–62.

Jiang R, Guan Y (2015) Data-driven chance constrained stochastic program. *Mathematical Programming* 158:291–327.

Kapsos M, Christofides N, Rustem B (2014) Worst-case robust Omega ratio. *European Journal of Operational Research* 234(2):499–507.

Keating C, Shadwick WF (2002) A universal performance measure. *Journal of Performance Measurement* 6(3):59–84.

Lejeune MA, Kettunen J (2017) Managing reliability and stability risks in forest harvesting. *Manufacturing & Service Operations Management* 19(4):620–638.

Liu Y, Meskarian R, Xu H (2017) Distributionally robust reward-risk ratio optimization with moment constraints. *SIAM Journal on Optimization* 27(2):957–985.

Pan L, Housh M, Liu P, Cai X, Chen X (2015) Robust stochastic optimization for reservoir operation. *Water Resources Research* 51(1):409–429.

Pflug GC, Wozabal D (2007) Ambiguity in portfolio selection. *Quantitative Finance* 7(4):435–442.

Postek K, Den Hertog D, Melenberg B (2016) Computationally tractable counterparts of distributionally robust constraints on risk measures. *SIAM Review* 58(4):603–650.

Rockafellar RT, Uryasev S (2000) Optimization of conditional Value-at-Risk. *Journal of Risk* 2:21–42.

Rockafellar RT, Uryasev S (2002) Conditional Value-at-Risk for general loss distributions. *Journal of Banking & Finance* 26(7):1443–1471.

Roy AD (1952) Safety first and the holding of assets. *Econometrica* 431–449.

Sen A, Atamturk A, Kaminsky P (2018) A conic integer optimization approach to the constrained assortment problem under the mixed multinomial logit model. *Operations Research* 66(4):994–1003.

Sharpe WF (1966) Mutual fund performance. *Journal of Business* 119–138.

Sortino FA, Satchel S (2001) *Managing Downside Risk in Financial Markets* (Butterworth-Heinemann).

Villani C (2008) *Optimal Transport: Old and New*, volume 338 (Springer Science & Business Media).

Wang R, Wang X, Wu L (2010) Sanov's theorem in the Wasserstein distance: A necessary and sufficient condition. *Statistics & Probability Letters* 80(5-6):505–512.

Wiesemann W, Kuhn D, Sim M (2014) Distributionally robust convex optimization. *Operations Research* 62(6):1358–1376.

Wozabal D (2012) A framework for optimization under ambiguity. *Annals of Operations Research* 193:21–47.

Wozabal D (2014) Robustifying convex risk measures for linear portfolios: A nonparametric approach. *Operations Research* 62(6):1302–1315.

Xu H, Liu Y, Sun H (2018) Distributionally robust optimization with matrix moment constraints: Lagrange duality and cutting plane methods. *Mathematical Programming* 169(2):489–529.

Žáčková J (1966) On minimax solutions of stochastic linear programming problems. *Časopis pro pěstování matematiky* 91(4):423–430.

EC.1. Proof for Theorem 8

Theorem 8. Let $\beta^* \in [\beta_L^0, \beta_U^0]$. The standard bisection algorithm finds the optimal solution β^* of **DRR** *in finitely many (i.e., at most $\log_2((\beta_U^0 - \beta_L^0)/\varepsilon)$)) iterations with precision level ε.*

Proof. At each iteration, let β^t be the midpoint of the incumbent interval $[\beta_L^t, \beta_U^t]$. Without loss of generality, we assume $\beta^* - \beta^t \neq 0$ for all $t \geq 1$. By construction, the bounds of the iterative intervals are as follows:

$$\beta_L^1 \leq \beta_L^2 \leq \ldots \leq \beta_L^t \leq \ldots \leq \ldots \leq \beta_U^t \leq \ldots \leq \beta_U^2 \leq \beta_U^1 \ ,$$

where the sequences $\{\beta_L^t\}$ and $\{\beta_U^t\}$ monotonically converge to the limits β_L^t and β_U^t $(\beta_L^t \leq \beta_U^t)$, respectively. It follows that $[\beta^* - \beta_L^t] \cdot [\beta^* - \beta_U^t] < 0$, and that there exists a value of β^* that satisfies: $\beta^* \in [\beta_L^t, \beta_U^t] \subset [\beta_L^0, \beta_U^0]$. At each iteration of the bisection algorithm, we split the incumbent interval $[\beta_L^t, \beta_U^t]$ into two equally-spaced intervals and calculates the midpoint $\beta^t = (\beta_L^t + \beta_U^t)/2$, from which it results that: $|\beta^t - \beta^*| \leq (\beta_U^t - \beta_L^t)/2$. Due to the monotonicity of $\{\beta_L^t\}$ and $\{\beta_U^t\}$, we have

$$\beta_U^t - \beta_L^t = \frac{\beta_U^{t-1} - \beta_L^{t-1}}{2} = \frac{\beta_U^{t-2} - \beta_L^{t-2}}{2^2} = \cdots = \frac{\beta_U^1 - \beta_L^1}{2^{n-1}} = \frac{\beta_U^0 - \beta_L^0}{2^{n-1}}$$

Therefore the precision level ε (i.e., absolute value deviation from β^t to β^*) is such that:

$$\varepsilon = |\beta^t - \beta^*| \leq \frac{\beta_U^0 - \beta_L^0}{2^t} \ , \quad t = 1, \ldots, \infty$$

Expressing the above equation with respect to t, we obtain the following bound on the number of iterations needed to reach the precision level ε

$$t = \log_2 \left(\frac{\beta_U^0 - \beta_L^0}{\varepsilon} \right),$$

which completes the proof. $\square$

EC.2. Bisection Algorithm with a Priori Interval Compaction – Problem Reformulations

EC.2.1. Maximizing Robust Reward

Consider problem **DRR-μ**. Suppose that the reward measure $\mu(x,p)$ is the portfolio's mean return, that is linear in p: $\mu(x,p) = \sum_{j \in \mathcal{N}} p_j R_j(x)$, where $R_j(x)$ represents the portfolio return in time period j. Let the function $f^\mu(x,p) = -\mu(x,p)$ denote the negative mean return. The corresponding conjugate function reads:

$$f_*^\mu(x,v) = \inf_p \left\{ v^T p - f(x,p) \right\} = \inf_p \left\{ v^T p + \mu(x,p) \right\} \tag{EC.1}$$

$$= \inf_p \left\{ \sum_{j \in \mathcal{N}} v_j p_j + \sum_{j \in \mathcal{N}} p_j R_j(x) \right\} = \begin{cases} 0, & \text{if} \quad v_j \geq -R_j(x), \forall j \in \mathcal{N} \\ -\infty, & \text{otherwise} \end{cases} . \tag{EC.2}$$

The reformulation of constraints (77) is based on Theorem 4 and involves the support function of the Wasserstein ambiguity set (**WA-D**) and the conjugate function $f_*^\mu(x,v)$ of the "risk" measure (negative mean return, $-\mu(x,p)$). Problem **DRR-**μ can be reformulated as:

$$
\begin{aligned}
\textbf{DRR-R}\boldsymbol{\mu}\textbf{:}\ \max\quad & \beta_\mu \\
s.t.\quad & \gamma\theta + \sum_{i\in\mathcal{N}} y_i p_i^0 \le -\beta_\mu && \text{(EC.3)} \\
& v_j \ge -R_j(x), \qquad \forall j\in\mathcal{N} && \text{(EC.4)} \\
& (4);(38)-(39);(78)\ .
\end{aligned}
$$

Problem **DRR-R**μ is linear with decision variables (x,y,β_μ,v,γ).

EC.2.2. Minimizing Robust Risk

Consider problem **DRR-**ρ and a risk measure $\rho(x,p)$ nonlinear in p. Let $f^\rho(x,p)=\rho(x,p)$. The conjugate function is

$$
\begin{aligned}
f_*^\rho(x,v) = \sup_w\quad & -\frac{w}{4} && \text{(EC.5)} \\
\text{s.t.}\quad & \left\| \begin{bmatrix} 2\rho_j(x) \\ (v_j - w) \end{bmatrix} \right\|_2 \le v_j + w, \qquad \forall j\in\mathcal{N} && \text{(EC.6)} \\
& v_j \ge 0, \qquad \forall j\in\mathcal{N} && \text{(EC.7)} \\
& w \ge 0\ . && \text{(EC.8)}
\end{aligned}
$$

The reformulation of constraints (77) is also based on Theorem 4, involving the support function of the Wasserstein ambiguity set (**WA-D**) and the conjugate function $f_*^\rho(x,v)$ of the risk measure $\rho(x,p)$. Problem **DRR-**ρ can be reformulated as

$$
\begin{aligned}
\textbf{DRR-R}\boldsymbol{\rho}\textbf{:}\ \min\quad & \beta_\rho \\
s.t.\quad & \gamma\theta + \sum_{i\in\mathcal{N}} y_i p_i^0 + \frac{w}{4} \le \beta_\rho && \text{(EC.9)} \\
& (4);(38)-(39);(81);(\text{EC.6})-(\text{EC.8})\ ,
\end{aligned}
$$

where the decision variables are $(x,y,\beta_\rho,v,\gamma,w)$.

EC.3. Pseudo-Code of Bisection Algorithm with Interval Compaction

Algorithm 1: Pseudo-Code of Bisection Algorithm with Interval Compaction

Initialization: Determine $[\beta_L^0, \beta_U^0]$ such that $\beta^* \in [\beta_L^0, \beta_U^0]$, where $\beta_U^0 = \beta_U$ is obtained using a priori compaction method; set $s^0 = \beta_L^0$.

Iterative Process: repeat

- **Step 1:** Set $\beta^t = (\beta_L^t + \beta_U^t)/2$;
- **Step 2:** Check if $\mathbf{DRR\text{-}F}_t$ admits a feasible solution;
- **if** $\mathbf{DRR\text{-}F}_t$ *is feasible,* **then**
  - check the binding status of all constraints involving β_t;
  - **if** *none of checking constraints is binding* **then**
    - solve problem $\mathbf{DRR\text{-}RF}_t$, obtain solution β'^t; set $s^t = \beta'^t$;
  - **else** set $s^t = \beta^t$;
  - update interval $[\beta_L^{t+1}, \beta_U^{t+1}]$: $\beta_L^{t+1} = s^t$ and $\beta_U^{t+1} = \beta_U^t$,
  - let $t = t + 1$;
- **else**
  - set $s^t = s^{t-1}$;
  - update interval $[\beta_L^{t+1}, \beta_U^{t+1}]$: $\beta_L^{t+1} = \beta_L^t$ and $\beta_U^{t+1} = \beta^t$,
  - let $t = t + 1$;
- **end**

until $\beta_U^t - \beta_L^t \le \varepsilon$*; the optimal value is:* $\beta^* = s^t$;

EC.4. Computational Results

The following insights were obtained from the analysis of the numerical results in Table EC.1:

1. *Reduction in number of iterations due to a priori compaction.* On average, the a priori compaction method reduced (Figure 2) the initial upper bound by 67.59% (i.e., 1.62 vs. 5). At $\epsilon = 0.01$, the average number of iterations was 6.58 with *BAIC-PC* and 8 with *SBA*, which corresponds to a drop of 17.75% in the number of iterations due to the a priori compaction method. Similarly, the average number of iterations with *BAIC* was 5.92 and 7.83 for *BAIC-IC*, i.e., a reduction of 24.47%. Similar reductions took place for ϵ=0.001.

2. *Reduction in number of iterations due to iterative compaction.* At $\epsilon = 0.01$, *BAIC-IC* and *SBA* required on average 7.83 and 8 iterations, respectively. Thus, the iterative compaction method used independently from the a priori compaction one reduced the number of iterations by 2.13%. On the other hand, *BAIC* and *BAIC-IC* respectively needed on average 5.92 and 6.58 iterations. The joint use of iterative and a priori compaction reduced the number of iterations by 26.04%.

3. *Impact of a priori compaction on average time per iteration.* Compared to *SBA*, the a priori compaction method algorithm *BAIC-PC* reduced the average time per iteration by 8.53% (88.85s vs. 97.13s). When added to the iterative compaction method, the a priori approach reduced the average time per iteration by 7.04% (i.e., 99.82s for *BAIC-IC* vs. 92.79s for *BAIC*).

4. *Impact of iterative compaction on average time per iteration.* Across the 24 instances, the average time per iteration with *BAIC-IC* was 2.77% higher than with the *SBA* one (i.e., 99.82 vs

97.13 seconds). The time increase is due to the fact that at each iteration, *BAIC-IC* solves two convex problems (i.e., $\mathbf{DRR\text{-}F}_t$ and $\mathbf{DRR\text{-}RF}_t$), while *SBA* only solves the first one.

Table EC.1 Computational Efficiency Results

			SBA				*BAIC-IC*			
		M	25				25			
ϵ	N	β^U	β	# Iter	TotalCPU	AveCPU	β	# Iter	TotalCPU	AveCPU
0.01	24	5.0	0.5957	8	0.77	0.10	0.6004	7	0.70	0.10
	60	5.0	0.2832	8	1.59	0.20	0.2863	8	1.92	0.24
	120	5.0	0.2637	8	25.17	3.15	0.2679	8	26.50	3.31
	180	5.0	0.2441	8	2046.52	255.81	0.2406	8	2062.91	257.86
0.001	24	5.0	0.6049	12	1.11	0.09	0.6050	10	1.25	0.13
	60	5.0	0.2881	12	2.86	0.24	0.2882	11	2.66	0.24
	120	5.0	0.2704	12	37.75	3.15	0.2701	10	39.98	4.00
	180	5.0	0.2454	12	3165.77	263.81	0.2455	11	3081.58	280.14
		M	100				100			
ϵ	N	β^U	β	# Iter	TotalCPU	AveCPU	β	# Iter	TotalCPU	AveCPU
0.01	24	5.0	0.9668	8	0.89	0.11	0.9699	7	0.92	0.13
	60	5.0	0.6348	8	1.77	0.22	0.6310	8	1.84	0.23
	120	5.0	0.4102	8	26.41	3.30	0.4122	8	28.36	3.54
	180	5.0	0.2734	8	3237.66	404.71	0.2736	8	3312.77	414.10
0.001	24	5.0	0.9692	12	1.44	0.12	0.9699	10	1.28	0.13
	60	5.0	0.6396	12	2.81	0.23	0.6394	11	2.52	0.23
	120	5.0	0.4144	12	38.91	3.24	0.4147	11	40.13	3.65
	180	5.0	0.2734	12	5025.66	418.80	0.2736	11	4724.51	429.50
		M	400				400			
ϵ	N	β^U	β	# Iter	TotalCPU	AveCPU	β	# Iter	TotalCPU	AveCPU
0.01	24	5.0	1.8066	8	1.00	0.13	1.8061	8	1.27	0.16
	60	5.0	0.7910	8	1.91	0.24	0.7869	8	2.17	0.27
	120	5.0	0.5371	8	23.09	2.89	0.5407	8	24.33	3.04
	180	5.0	0.3613	8	3820.51	477.56	0.3629	8	3880.13	485.02
0.001	24	5.0	1.8103	12	1.70	0.14	1.8129	9	1.28	0.14
	60	5.0	0.7910	12	3.36	0.28	0.7912	11	2.97	0.27
	120	5.0	0.5420	12	38.48	3.21	0.5424	11	37.02	3.37
	180	5.0	0.3662	12	5874.05	489.50	0.3656	11	5565.78	505.98

			BAIC-PC				*BAIC*			
		M	25				25			
ϵ	N	β^U	β	# Iter	TotalCPU	AveCPU	β	# Iter	TotalCPU	AveCPU
0.01	24	1.4127	0.6015	7	0.70	0.10	0.6030	6	0.75	0.13
	60	0.6482	0.2836	6	1.38	0.23	0.2884	5	1.11	0.22
	120	0.5039	0.2677	5	19.80	3.96	0.2683	5	20.08	4.02
	180	0.3653	0.2454	5	1117.88	223.58	0.2440	5	1132.56	226.51
0.001	24	1.4127	0.6050	10	0.97	0.10	0.6053	8	0.92	0.12
	60	0.6482	0.2880	9	1.92	0.21	0.2884	8	1.98	0.25
	120	0.5039	0.2697	8	34.27	4.28	0.2702	7	30.05	4.29
	180	0.3653	0.2454	8	2116.17	264.52	0.2453	7	1988.39	284.06
		M	100				100			
ϵ	N	β^U	β	# Iter	TotalCPU	AveCPU	β	# Iter	TotalCPU	AveCPU
0.01	24	2.9037	0.9641	8	0.98	0.12	0.9676	7	0.95	0.14
	60	2.0335	0.6355	7	1.69	0.24	0.6360	6	1.70	0.28
	120	1.0852	0.4070	6	23.73	3.96	0.4143	5	16.86	3.37
	180	0.7900	0.2716	6	2423.85	403.98	0.2715	5	2139.23	427.85
0.001	24	2.9037	0.9691	11	1.25	0.11	0.9695	10	1.27	0.13
	60	2.0335	0.6394	10	2.42	0.24	0.6397	9	2.05	0.23
	120	1.0852	0.4149	10	32.30	3.23	0.4149	9	32.05	3.56
	180	0.7900	0.2731	9	3571.64	396.85	0.2731	8	3304.90	413.11
		M	400				400			
ϵ	N	β^U	β	# Iter	TotalCPU	AveCPU	β	# Iter	TotalCPU	AveCPU
0.01	24	4.1336	1.8084	8	1.23	0.15	1.8099	7	0.98	0.14
	60	2.2880	0.7865	7	1.59	0.23	0.7897	7	1.84	0.26
	120	1.9647	0.5372	7	21.52	3.07	0.5416	7	22.03	3.15
	180	1.3174	0.3654	7	2845.79	406.54	0.3599	6	2578.63	429.77
0.001	24	4.1336	1.8100	12	1.78	0.15	1.8099	10	1.56	0.16
	60	2.2880	0.7915	11	2.83	0.26	0.7915	10	2.58	0.26
	120	1.9647	0.5420	10	30.22	3.02	0.5424	10	33.45	3.35
	180	1.3174	0.3660	10	4131.61	413.16	0.3661	9	3795.51	421.72