Probability in the Engineering and Informational Sciences (2024), **38:**2 318–340
doi:10.1017/S0269964823000141

CAMBRIDGE UNIVERSITY PRESS

RESEARCH ARTICLE

Worst-case Omega ratio under distribution uncertainty with its application in robust portfolio selection

Qiuyang Li[1] and Xinqiao Xie[2]

[1]School of Data Science, University of Science and Technology of China, Hefei, Anhui, China
[2]Department of Finance and Statistics, School of Management, University of Science and Technology of China, Hefei, Anhui, China
Corresponding author: Xinqiao Xie; Email: xxqyzb@mail.ustc.edu.cn

Keywords: distributional robustness; moments; Omega ratio; portfolio selection; Wasserstein distance

Abstract
Omega ratio, a risk-return performance measure, is defined as the ratio of the expected upside deviation of return to the expected downside deviation of return from a predetermined threshold described by an investor. Motivated by finding a solution protected against sampling errors, in this paper, we focus on the worst-case Omega ratio under distributional uncertainty and its application to robust portfolio selection. The main idea is to deal with optimization problems with all uncertain parameters within an uncertainty set. The uncertainty set of the distribution of returns given characteristic information, including the first two orders of moments and the Wasserstein distance, can handle data problems with uncertainty while making the calculation feasible.

1. Introduction

Optimal investment theory is an important topic in quantitative risk management and involves determining how to allocate capital among available securities in order to maximize the expected return for a given level of risk or minimize the associated risk for a given level of return. Markowitz [13] proposed the first quantitative method for determining the optimal portfolio based on minimizing the portfolio variance for a given expectation, called mean-variance principle. In classical optimal investment theory, the distribution obeyed by the data is usually assumed to be known or an empirical distribution function is used instead of the unknown distribution, whereas in practice, the distribution obeyed by the data cannot be determined if the complete data is not available or if a small sample of data is obtained and an empirical distribution function cannot be used. This resulted in other samples performing poorly.

Motivated by this, distributionally robust optimization (DRO) has emerged as a paradigm aimed at finding a solution that is protected against sampling errors [6]. It seeks for a solution that performs the best with respect to the most adversarial distribution from a set of distributions, known as ambiguity set. The DRO identifies the most adversarial distribution, the worst-case distribution, from an ambiguity set, and makes an optimal decision, the optimal portfolio, which minimizes the cost induced by the worst-case distribution. The choice of a proper ambiguity set is crucial in DRO. Such DRO problem and its robust decision have been verified to exhibit some desirable properties such as finite-sample guarantee and computational tractability, which are in sharp contract with the classic stochastic optimization problem. Hence, DRO has been widely applied in data-driven problems arising from many operations research and machine learning applications. Among others, in portfolio selection, Blanchet et al. [3] consider DRO based on mean-variance principle. The same principle is also used in Calafiore and Ghaoui [4], Zhu et al. [22] and Popescu [17], in which a linear-chance-constrained problem, a minimax regret objective and a portfolio optimization problem are considered, respectively. However, mean-variance

principle has been criticized for that its measure of risk, variance, is not a monotone measure. Shadwick and Keating [18] introduced the Omega ratio to analyze information on the distribution of potential returns. Essentially, it is the ratio of the upside of an investment to the downside. Omega ratio has been verified to provide a new perspective in the performance evaluation of hedge funds [18] and build as a useful measure [1]. Therefore, in this paper, we focus on DRO based on the mean-Omega ratio principle in portfolio selection. Specifically, we first calculate the worst-case Omega ratio from the ambiguity set and then find the optimal portoflio, which minimizes the worst-case Omega ratio. While the current paper focuses on mean-Omega ratio principles, we should point out that recently several other works have explored other measures such as mean-VaR and mean-CVaR in distributionally robust portfolio selection (see [11, 12] and the references therein).

In addition to the principle employed in DRO, the choice of a proper ambiguity set is crucial in DRO. In particular, one common way of defining the set $\mathcal{F}$ is through specifying the moments of the distribution (see, e.g., [2, 6, 15, 17, 21]); DRO problems based on moments information has been proved to have a tractable form in applications. So we will first consider the worst-case Omega ratio based on the mean-variance information. Motivated by recent advances in data-driven DRO problem, we also consider that the uncertainty set is the Wasserstein ball centred at the empirical distribution. The Wasserstein ball that is centered at the empirical distribution has become popular as it can make full use of the data, and this ambiguity set has become an attractive ambiguity set adopted in DRO. To solve a DRO problem, identifying the most adversarial distribution, the worst-case distribution, from an ambiguity set is essential. We first study the worst-case distribution and calculate its Omega ratio. The worst-case Omega ratio is calculated under mean-variance uncertainty set and mean-Wasserstein uncertainty set, respectively. With aid of the projection result of the two ambiguity sets, we apply the results in the portfolio selection.

The rest of the paper is organized as follows. Section 2 presents the definition of the Omega ratio, gives explicit expression of the worst-case Omega ratio under mean-variance uncertainty set and aims to tackle the worst-case Omega ratio under mean-Wasserstein uncertainty set. Section 3 applies the worst-case Omega ratio under mean-variance uncertainty set and mean-Wasserstein uncertainty set to the portfolio. In Section 4, we present the simulation and empirical results derived from portfolios comprised of our strategies, comparing them to those of various alternative models.

Throughout the paper, let $d \in \mathbb{N}$ and $\mathbb{R}^d$ be the d-dimensional Euclidean space. Let $\mathcal{P}(\mathbb{R}^d)$ be the set of all distributions on $\mathbb{R}^d$ and denote $\mathcal{P}_p(\mathbb{R}^d)$ by the subset of $\mathcal{P}(\mathbb{R}^d)$ with finite pth moment for $p \in [1, \infty)$. For a random variable (vector) $\mathbf{X} \in \mathcal{P}(\mathbb{R}^d)$, denote $\mathbb{E}[\mathbf{X}]$ by the mean (vector) of the random variable (vector) $\mathbf{X}$. For a law-invariant[1] real-valued mapping ρ on $\mathcal{P}(\mathbb{R}^d)$, the notation $\rho^F(\mathbf{X})$ means the value of $\rho(\mathbf{X})$, where $\mathbf{X}$ has the distribution F, and we omit F for simplicity. Furthermore, we denote by $x_+ = \max(0, x)$ and $x_- = \max(0, -x)$.

2. Worst-case Omega ratio

In this section, we only discuss the case of $d = 1$. The case of $d > 1$ is discussed in Section 3, where the corresponding portfolio selection problem is investigated. One popular measure of investment performance in finance is Omega ratio introduced by Shadwick and Keating [18]: the ratio of the upside of an investment to the downside.

Definition 2.1. (Shadwick and Keating [18]). *For a payoff X with distribution F and a sure payoff c, Omega ratio is defined as*

$$\Omega_X(c) := \frac{\mathbb{E}\left[(X-c)_+\right]}{\mathbb{E}\left[(c-X)_+\right]}.$$

[1] We say a mapping ρ on $\mathcal{P}(\mathbb{R}^d)$ is law-invariant if $\rho(\mathbf{X}) = \rho(\mathbf{Y})$ whenever $\mathbf{X}$ and $\mathbf{Y}$ has the same distribution.

https://doi.org/10.1017/S0269964823000141 Published online by Cambridge University Press

2.1. Mean-variance uncertainty set

We assume that the distribution of X lies in a distribution set where first and second moments are fixed. Based on this assumption, we consider the following DRO problem for Omega ratio:

$$\underline{\Omega}_X(c,\mu,\sigma) = \inf_{F\in\mathcal{S}_1} \Omega_X^F(c), \qquad \overline{\Omega}_X(c,\mu,\sigma) = \sup_{F\in\mathcal{S}_1} \Omega_X^F(c), \tag{1}$$

where

$$\mathcal{S}_1 := \mathcal{S}(\mu,\sigma) = \left\{F \in \mathcal{P}_2(\mathbb{R}) : \int_{-\infty}^{\infty} x\,\mathrm{d}F(x) = \mu, \int_{-\infty}^{\infty} x^2\,\mathrm{d}F(x) = \mu^2+\sigma^2\right\}, \tag{2}$$

and $\Omega_X^F(c)$ represents the Omega ratio that is calculated under the constraint that the distribution of X is F. The following theorem gives the explicit solutions of the optimization problem (1).

Theorem 2.2. *Given $\mu \in \mathbb{R}$ and $\sigma > 0$, we have*

$$\underline{\Omega}_X(c,\mu,\sigma) = \begin{cases} 0, & \mu < c, \\ \frac{\sqrt{1+S^2}+S}{\sqrt{1+S^2}-S}, & \mu \geqslant c, \end{cases} \tag{3}$$

and

$$\overline{\Omega}_X(c,\mu,\sigma) = \begin{cases} \frac{\sqrt{1+S^2}+S}{\sqrt{1+S^2}-S}, & \mu < c, \\ \infty, & \mu \geqslant c, \end{cases} \tag{4}$$

where $S = \frac{\mu-c}{\sigma}$.

We first present a lemma from [10], which is subsequently used in the proof of Theorem 2.2.

Lemma 2.3. *Given $\mu \in \mathbb{R}$ and $\sigma > 0$, we have*

$$\max_{F\in\mathcal{S}_1} \mathbb{E}^F[(X-c)_+] = \frac{1}{2}\left[\sqrt{\sigma^2+(\mu-c)^2}+\mu-c\right],$$

where $\mathcal{S}_1$ is defined by Eq. (2).

Proof of Theorem 2.2. Since $\mathbb{E}^F[X-c] = \mathbb{E}^F[(X-c)_+] - \mathbb{E}^F[(X-c)_-] = \mathbb{E}^F[(X-c)_+] - \mathbb{E}^F[(c-X)_+]$, the Omega function can be rewritten as

$$\Omega_X^F(c) = \frac{\mathbb{E}^F[(X-c)_+]}{\mathbb{E}^F[(c-X)_+]} = \frac{\mathbb{E}^F[(X-c)_+]}{\mathbb{E}^F[(X-c)_+] - \mathbb{E}^F[(X-c)]} = \left\{1 - \frac{\mathbb{E}^F[X]-c}{\mathbb{E}^F[(X-c)_+]}\right\}^{-1}. \tag{5}$$

We first consider the case of $\underline{\Omega}_X(c,\mu,\sigma)$. If $\mu \geqslant c$, then we have

$$\begin{aligned}\underline{\Omega}_X(c,\mu,\sigma) &= \inf_{F\in\mathcal{S}_1} \Omega_X^F(c) = \inf_{F\in\mathcal{S}_1} \left\{1 - \frac{\mathbb{E}^F[X]-c}{\mathbb{E}^F[(X-c)_+]}\right\}^{-1} \\ &= \left\{1 - \frac{\mu-c}{\max_{F\in\mathcal{S}_1}\{\mathbb{E}^F[(X-c)_+]\}}\right\}^{-1}\end{aligned}$$

https://doi.org/10.1017/S0269964823000141 Published online by Cambridge University Press

$$= \left\{1 - \frac{\mu - c}{\frac{1}{2}[\sqrt{\sigma^2 + (\mu - c)^2} + \mu - c]}\right\}^{-1}$$
$$= \frac{\sqrt{1 + S^2} + S}{\sqrt{1 + S^2} - S},$$

where the second and fourth equalities follow from Eq. (5) and Lemma 2.3, respectively. If $\mu < c$, define a discrete random variable as follows:

$$X_\alpha = \begin{cases} \mu - \sigma\sqrt{\frac{1-\alpha}{\alpha}}, & \text{with probability } \alpha, \\ \mu + \sigma\sqrt{\frac{\alpha}{1-\alpha}}, & \text{with probability } 1 - \alpha. \end{cases} \tag{6}$$

Let $\alpha < \frac{(\mu-c)^2}{\sigma^2+(\mu-c)^2}$. One can check that $\mathbb{E}[X_\alpha] = \mu$, $\mathrm{Var}(X_\alpha) = \sigma^2$, which implies that $F_{X_\alpha} \in \mathcal{S}_1$. Also noting that $X_\alpha < c$ almost surely, we have

$$0 \leqslant \underline{\Omega}_X(c, \mu, \sigma) \leqslant \Omega_{X_\alpha}(c) = 0.$$

Hence, we conclude that $\underline{\Omega}_X(c, \mu, \sigma) = 0$ for $\mu < c$. This completes the proof of Eq. (3).

Next, we consider the case of $\overline{\Omega}_X(c, \mu, \sigma)$. If $\mu \leqslant c$, then we have

$$\overline{\Omega}_X(c, \mu, \sigma) = \sup_{F \in \mathcal{S}_1} \Omega_X^F(c) = \sup_{F \in \mathcal{S}_1} \left\{1 - \frac{\mathbb{E}^F[X] - c}{\mathbb{E}^F[(X - c)_+]}\right\}^{-1}$$
$$= \left\{1 + \frac{c - \mu}{\max_{F \in \mathcal{S}_1}\{\mathbb{E}^F[(X - c)_+]\}}\right\}^{-1}$$
$$= \left\{1 + \frac{c - \mu}{\frac{1}{2}[\sqrt{\sigma^2 + (\mu - c)^2} + \mu - c]}\right\}^{-1}$$
$$= \frac{\sqrt{1 + S^2} + S}{\sqrt{1 + S^2} - S},$$

where the second and fourth equalities follow from Eq. (5) and Lemma 2.3, respectively. If $\mu > c$, recall the definition of X_α in Eq. (6). Let $\alpha > \frac{(\mu-c)^2}{\sigma^2+(\mu-c)^2}$. One can check that $\mathbb{E}[X_\alpha] = \mu$, $\mathrm{Var}(X_\alpha) = \sigma^2$, which implies that $F_{X_\alpha} \in \mathcal{S}_1$. Also noting that $X_\alpha > c$ almost surely, we have

$$\overline{\Omega}_X(c, \mu, \sigma) \geqslant \Omega_{X_\alpha}(c) = +\infty.$$

Hence, we have $\underline{\Omega}_X(c, \mu, \sigma) = 0$ for $\mu > c$. This completes the proof of (4). □

Remark 1. The slope of a straight line tangent to the efficient frontier, consisting of both risky and risk-free assets, is called the Sharpe ratio [19]. That is, for a random variable X with mean μ and standard deviation σ, Sharpe ratio is defined as $S_X = \frac{\mathbb{E}[X]-r_0}{\sqrt{\mathrm{Var}(X)}}$, where Var represents the variance and r_0 is a reference return rate, typically taken as the risk-free return rate. Theorem 2.2 implies that the upper and lower bounds of the Omega ratio can be obtained from a given Sharpe ratio, which is included by the uncertainty set (2). Observing the forms of Eqs. (3) and (4), we find that the robust value increases with the given Sharpe ratio.

https://doi.org/10.1017/S0269964823000141 Published online by Cambridge University Press

2.2. Mean-Wasserstein uncertainty set

In this section, we present the worst-case Omega ratio based on Wasserstein uncertainty set. We first give the definition of Wasserstein Distance from [20].

Definition 2.5. (Wasserstein Distance). *The p-Wasserstein distance $W_p(G_1, G_2)$ between $G_1, G_2 \in \mathcal{P}_p(\mathbb{R}^d)$ is defined by*

$$W_p(G_1, G_2) := \inf_{\pi \in \mathcal{P}_p(\mathbb{R}^d \times \mathbb{R}^d)} \left\{ \left(\int_{\mathbb{R}^d \times \mathbb{R}^d} \|x - y\|^p \, \pi(\mathrm{d}x, \mathrm{d}y) \right)^{\frac{1}{p}} \Big| \begin{array}{c} \pi \in \mathcal{P}_p(\mathbb{R}^d \times \mathbb{R}^d) \ \ s.t. \\ \pi(\cdot \times \mathbb{R}^d) = G_1, \ \pi(\mathbb{R}^d \times \cdot) = G_2 \end{array} \right\},$$

where $\|\cdot\|$ is a norm on $\mathbb{R}^d$, whose dual norm is defined as $\|y\|_* := \sup_{\|x\| \leqslant 1} x^\top y$ for $y \in \mathbb{R}^d$.

The Wasserstein distance is a natural way to compare two distributions when one is obtained from the other by perturbation. For a fixed $F \in \mathcal{P}_p(\mathbb{R}^d)$ and $\varepsilon > 0$, the Wasserstein ball centered on F is defined as follows:

$$\mathscr{B}_\varepsilon^p(F) := \left\{ G \in \mathcal{P}_p(\mathbb{R}^d) \middle| W_p(G, F) \leqslant \varepsilon \right\}.$$

For $d = 1$, we are interested in the worst-case Omega ratio in scenarios where the expected return is greater than the risk-free return, that is, $\mathbb{E}[X] \geqslant c$, where c is the risk-free return. Hence, we consider the worst-case Omega ratio under the mean and p-Wasserstein uncertainty set:

$$\underline{\Omega}_X^F(c, \mu, \varepsilon) = \inf_{G \in \mathcal{S}_2} \Omega_X^G(c), \tag{7}$$

where $F \in \mathcal{P}_p(\mathbb{R})$,

$$\mathcal{S}_2 := \mathcal{S}_{p,\varepsilon}(\mu, F) = \left\{ G \in \mathscr{B}_\varepsilon^p(F) \middle| \int x \, \mathrm{d}G(x) = \mu \right\}$$

and $\Omega_X^G(c)$ represents the Omega ratio that is calculated under the constraint that the distribution of X is G.

We turn to the main result of this section, which indicates that the explicit solution of problem (7) is obtained for $p = 1$, and we simplify problem (7) for $p > 1$.

Theorem 2.6. *Given c, μ, $\varepsilon > 0$, $p \in [1, \infty)$ and $F \in \mathcal{P}_p(\mathbb{R})$, assume $\mu \geqslant c$. Let $\underline{\Omega}_X^F(c, \mu, \varepsilon)$ be the optimal value of problem (7) for a given distribution F, μ_F denote the random variable's mean under the distribution F. We have the following results.*

(i) For $p = 1$, $\mu_F^{c+} := \mathbb{E}^F[(X - c)_+]$, then

$$\underline{\Omega}_X^F(c, \mu, \varepsilon) = \begin{cases} 1, & \text{if } \mu \in \mathcal{A}_1 \cup \mathcal{B}_1, \\ \left[1 - \dfrac{2(\mu - c)}{\mu - \mu_F + \varepsilon} \right]^{-1}, & \text{if } \mu \in \mathcal{A}_2 \cup \mathcal{B}_2, \end{cases}$$

where $\mathcal{A}_1 = \{\mu : \mu > \mu_F - \mu_F^{c+}, \mu > \mu_F - 2\mu_F^{c+} + \varepsilon\}$, $\mathcal{A}_2 = \{\mu : \mu > \mu_F - \mu_F^{c+}, \mu \leqslant \mu_F - 2\mu_F^{c+} + \varepsilon\}$, $\mathcal{B}_1 = \{\mu : \mu \leqslant \mu_F - \mu_F^{c+}, \mu < \mu_F - \varepsilon\}$ and $\mathcal{B}_2 = \{\mu : \mu \leqslant \mu_F - \mu_F^{c+}, \mu \geqslant \mu_F - \varepsilon\}$.

https://doi.org/10.1017/S0269964823000141 Published online by Cambridge University Press

(i) Otherwise, for $p > 1$, we have

$$\underline{\Omega}_X^F(c,\mu,\varepsilon) = \left(1 - \frac{\mu - c}{T(\mu)}\right)^{-1},$$

where

$$T(\mu) := \inf_{\lambda \geqslant 0, \gamma \in \mathbb{R}} \left\{ \lambda \varepsilon^p + \gamma(\mu - \mu_F) + |\gamma| \left(1 - \frac{1}{p}\right) \left(\frac{|\gamma|}{\lambda p}\right)^{\frac{1}{p-1}} + \mathbb{E}^F[(X - c_1(\lambda,\gamma))_+] \right\}$$

and

$$c_1(\lambda,\gamma) := c + \left(1 - \frac{1}{p}\right) \left[\left(\frac{|\gamma|}{\lambda p}\right)^{\frac{1}{p-1}} |\gamma| - \left(\frac{|1-\gamma|}{\lambda p}\right)^{\frac{1}{p-1}} |1-\gamma| \right].$$

To simplify problem (7), we need to utilize the following lemma, the proof of which is directly derived from Theorem 1 of [8].

Lemma 2.7. *For $p \in [1,\infty)$, $F \in \mathcal{P}_p(\mathbb{R})$, $\mu \in \mathbb{R}$, $\varepsilon > 0$ and $\Psi : \mathbb{R} \to \mathbb{R}$, we have*

$$\sup_{G \in \mathcal{S}_2} \int_{\mathbb{R}} \Psi(\xi)\, \mathrm{d}G(\xi) = \inf_{\lambda \geqslant 0, \gamma \in \mathbb{R}} \left\{ \lambda \varepsilon^p + \int_{\mathbb{R}} \sup_{\xi \in \mathbb{R}} [\Psi(\xi) - \gamma\xi - \lambda|\xi - \zeta|^p]\, \mathrm{d}F(\zeta) + \gamma\mu \right\}.$$

Proof of Theorem 2.6. We know that if $\mathbb{E}^G[X] = \mu$, Omega ratio can be expressed as

$$\Omega_X^G(c) = \frac{\mathbb{E}^G[(X-c)_+]}{\mathbb{E}^G[(c-X)_+]} = \left\{1 - \frac{\mu - c}{\mathbb{E}^G[(X-c)_+]}\right\}^{-1}.$$

Denote $T(\mu) = \sup_{G \in \mathcal{S}_2} \mathbb{E}^G[(X-c)_+]$, for $\mu \geqslant c$, we can obtain that

$$\inf_{G \in \mathcal{S}_2} \frac{\mu - c}{\mathbb{E}^G[(X-c)_+]} = \frac{\mu - c}{T(\mu)}.$$

Hence, problem (7) can be reduced as follow:

$$\begin{aligned}
\underline{\Omega}_X(\mu,c,\varepsilon) &= \inf_{G \in \mathcal{S}_2} \Omega_X^G(c) = \inf_{G \in \mathcal{S}_2} \left\{1 - \frac{\mu - c}{\mathbb{E}^G[(X-c)_+]}\right\}^{-1} \\
&= \left\{1 - \inf_{G \in \mathcal{S}_2} \frac{\mu - c}{\mathbb{E}^G[(X-c)_+]}\right\}^{-1} \\
&= \left\{1 - \frac{\mu - c}{T(\mu)}\right\}^{-1} \\
&= \frac{T(\mu)}{T(\mu) - (\mu - c)}.
\end{aligned}$$

https://doi.org/10.1017/S0269964823000141 Published online by Cambridge University Press

Applying Lemma 2.7, denote by

$$H_{\lambda,\gamma}(y) = \sup_{x\in\mathbb{R}}\{(x-c)_+ - \gamma x - \lambda|x-y|^p\},$$

$T(\mu)$ can be reformulated as

$$T(\mu) = \sup_{G\in\mathcal{S}_2} \mathbb{E}^G[(X-c)_+] = \inf_{\lambda\geqslant 0,\gamma\in\mathbb{R}} \left\{\lambda\varepsilon^p + \mathbb{E}^F\left[H_{\lambda,\gamma}(Y)\right] + \gamma\mu\right\}. \tag{8}$$

To simplify $T(\mu)$, we calculate the RHS of Eq. (8) in the following steps.

1. Calculate $H_{\lambda,\gamma}(y)$.

   Let $h_y(x) = (x-c)_+ - \gamma x - \lambda|x-y|^p$. For $x\in\mathbb{R}\backslash\{c,y\}$, the derivative of $h_y(x)$ with respect to x is

$$\frac{\partial h_y(x)}{\partial x} = I(x>c) - \gamma - \operatorname{sgn}(x-y)\lambda p|x-y|^{p-1}.$$

   To calculate the maximum of $h_y(x)$, we consider the case of $p=1$ and $p>1$, respectively.

   (i) If $p=1$, for $x\in\mathbb{R}\backslash\{c,y\}$, the derivative of $h_y(x)$ with respect to x can be written as

$$\frac{\partial h_y(x)}{\partial x} = \begin{cases} 1-\gamma-\lambda, & x>\max\{c,y\},\\ I(x>c)-\gamma-\operatorname{sgn}(x-y)\lambda, & x\in(\min\{c,y\},\max\{c,y\}),\\ -\gamma+\lambda, & x<\min\{c,y\}. \end{cases}$$

   In the case of $\lambda\geqslant\frac{1}{2}$ and $1-\lambda\leqslant\gamma\leqslant\lambda$, if $\frac{\partial h_y(x)}{\partial x}\geqslant 0$ for $x<\min\{c,y\}$ and $\frac{\partial h_y(x)}{\partial x}\leqslant 0$ for $x>\max\{c,y\}$, then the maximizer to $h_y(x)$ lies in the interval $[\min\{c,y\},\max\{c,y\}]$. If $y<c$, as $\frac{\partial h_y(x)}{\partial x}\leqslant 0$ for $x\in(y,c)$, the maximizer to $h_y(x)$ is $x^*=y$ and $h_y(y)=-\gamma y$. If $y\geqslant c$, as $\frac{\partial h_y(x)}{\partial x}\geqslant 0$ for $x\in(c,y)$, the maximizer to $h_y(x)$ is $x^*=c$ and $h_y(c)=-\lambda y+(\lambda-\gamma)c$. Otherwise, it can be verified that

$$\sup_{x\in\mathbb{R}} h_y(x) = \max\{\lim_{x\to\infty} h_y(x), \lim_{x\to-\infty} h_y(x)\} = \infty.$$

   In summary, in the case of $p=1$, if $\lambda\geqslant\frac{1}{2}$ and $1-\lambda\leqslant\gamma\leqslant\lambda$, the value of $H_{\lambda,\gamma}(y)$ is

$$H_{\lambda,\gamma}(y) = \sup_{x\in\mathbb{R}} h_y(x) = \begin{cases} -\gamma y, & y<c,\\ -\lambda y+(\lambda-\gamma)c, & y\geqslant c; \end{cases}$$

   otherwise, the value of $H_{\lambda,\gamma}(y)$ is ∞.

   (ii) If $p>1$, for $y<c$, we have

$$\frac{\partial h_y(x)}{\partial x} = \begin{cases} 1-\gamma-\lambda p|x-y|^{p-1}, & x>c,\\ -\gamma-\lambda p|x-y|^{p-1}, & x\in(y,c),\\ -\gamma+\lambda p|x-y|^{p-1}, & x<y. \end{cases}$$

https://doi.org/10.1017/S0269964823000141 Published online by Cambridge University Press

Conversely, for $y \geqslant c$, then

$$\frac{\partial h_y(x)}{\partial x} = \begin{cases} 1-\gamma-\lambda p|x-y|^{p-1}, & x > y, \\ 1-\gamma+\lambda p|x-y|^{p-1}, & x \in (c, y), \\ -\gamma+\lambda p|x-y|^{p-1}, & x < c. \end{cases}$$

Therefore, it is necessary to discuss the classification of γ. For $\gamma \leqslant 0$, we consider the following three cases.

(a) For $y < c - \left(\frac{1-\gamma}{\lambda p}\right)^{\frac{1}{p-1}}$, solving $\frac{\partial h_y(x)}{\partial x} > 0$ yields $x < y + \left(\frac{-\gamma}{\lambda p}\right)^{\frac{1}{p-1}}$ and solving $\frac{\partial h_y(x)}{\partial x} \leqslant 0$ yields $x \geqslant y + \left(\frac{-\gamma}{\lambda p}\right)^{\frac{1}{p-1}}$. As a result, the maximizer to $h_y(x)$ is $x_1^* = y + \left(\frac{-\gamma}{\lambda p}\right)^{\frac{1}{p-1}}$ and

$$h_y(x_1^*) = -\gamma y - \gamma\left(1-\frac{1}{p}\right)\left(\frac{-\gamma}{\lambda p}\right)^{\frac{1}{p-1}}.$$

(b) For $c - \left(\frac{1-\gamma}{\lambda p}\right)^{\frac{1}{p-1}} \leqslant y \leqslant c - \left(\frac{-\gamma}{\lambda p}\right)^{\frac{1}{p-1}}$, we find that $\frac{\partial h_y(x)}{\partial x} > 0$ in the cases that $x < y + \left(\frac{-\gamma}{\lambda p}\right)^{\frac{1}{p-1}}$ or $c < x < y + \left(\frac{1-\gamma}{\lambda p}\right)^{\frac{1}{p-1}}$ and $\frac{\partial h_y(x)}{\partial x} \leqslant 0$ in the cases that $y + \left(\frac{-\gamma}{\lambda p}\right)^{\frac{1}{p-1}} \leqslant x \leqslant c$ or $x \geqslant y + \left(\frac{1-\gamma}{\lambda p}\right)^{\frac{1}{p-1}}$. Hence, the local maximizers to $h_y(x)$ are

$$x_1^* = y + \left(\frac{-\gamma}{\lambda p}\right)^{\frac{1}{p-1}} \quad \text{and} \quad x_2^* = y + \left(\frac{1-\gamma}{\lambda p}\right)^{\frac{1}{p-1}}.$$

In addition, the local maximum values of $h_y(x)$ are

$$h_y(x_1^*) = -\gamma y - \gamma\left(1-\frac{1}{p}\right)\left(\frac{-\gamma}{\lambda p}\right)^{\frac{1}{p-1}}$$

and

$$h_y(x_2^*) = y - \gamma y - c + (1-\gamma)\left(1-\frac{1}{p}\right)\left(\frac{1-\gamma}{\lambda p}\right)^{\frac{1}{p-1}}.$$

(c) For $y < c - \left(\frac{-\gamma}{\lambda p}\right)^{\frac{1}{p-1}}$, solving $\frac{\partial h_y(x)}{\partial x} > 0$ yields $x < y + \left(\frac{1-\gamma}{\lambda p}\right)^{\frac{1}{p-1}}$ and solving $\frac{\partial h_y(x)}{\partial x} \leqslant 0$ yields $x \geqslant y + \left(\frac{1-\gamma}{\lambda p}\right)^{\frac{1}{p-1}}$. As a result, the maximizer to $h_y(x)$ is $x_2^* = y + \left(\frac{1-\gamma}{\lambda p}\right)^{\frac{1}{p-1}}$ and

$$h_y(x_2^*) = y - \gamma y - c + (1-\gamma)\left(1-\frac{1}{p}\right)\left(\frac{1-\gamma}{\lambda p}\right)^{\frac{1}{p-1}}.$$

Let

$$c_0(\lambda, \gamma) = c + \left(1-\frac{1}{p}\right)\left[\left(\frac{-\gamma}{\lambda p}\right)^{\frac{1}{p-1}}(-\gamma) - \left(\frac{1-\gamma}{\lambda p}\right)^{\frac{1}{p-1}}(1-\gamma)\right],$$

https://doi.org/10.1017/S0269964823000141 Published online by Cambridge University Press

one can easily verify that $c-\left(\frac{1-\gamma}{\lambda p}\right)^{\frac{1}{p-1}} \leqslant c_0(\lambda,\gamma) \leqslant c-\left(\frac{-\gamma}{\lambda p}\right)^{\frac{1}{p-1}}$ by mean value theorem. In addition, $y < c_0(\lambda,\gamma)$ implies $h_y(x_1) > h_y(x_2)$. Due to the above discussion, if $\gamma \leqslant 0$, the maximum value of $h_y(x)$ is as follows:

$$H_{\lambda,\gamma}(y) = \sup_{x\in\mathbb{R}} h_y(x) = \begin{cases} -\gamma y - \gamma\left(1-\frac{1}{p}\right)\left(\frac{-\gamma}{\lambda p}\right)^{\frac{1}{p-1}}, & y < c_0(\lambda,\gamma), \\ y-\gamma y - c + (1-\gamma)\left(1-\frac{1}{p}\right)\left(\frac{1-\gamma}{\lambda p}\right)^{\frac{1}{p-1}}, & y \geqslant c_0(\lambda,\gamma). \end{cases}$$

Similarly, for $0 < \gamma < 1$ and $\gamma \geqslant 1$, we can obtain similar results with only the difference in the sign of the open root formula. Hence, for $\gamma \in \mathbb{R}$ and $\lambda \geqslant 0$, we obtain the value of $H_{\lambda,\gamma}(y)$ as

$$H_{\lambda,\gamma}(y) = \sup_{x\in\mathbb{R}} h_y(x) = \begin{cases} -\gamma y + |\gamma|\left(1-\frac{1}{p}\right)\left(\frac{|\gamma|}{\lambda p}\right)^{\frac{1}{p-1}}, & y < c_1(\lambda,\gamma), \\ y-\gamma y - c + |1-\gamma|\left(1-\frac{1}{p}\right)\left(\frac{|1-\gamma|}{\lambda p}\right)^{\frac{1}{p-1}}, & y \geqslant c_1(\lambda,\gamma), \end{cases}$$

where

$$c_1(\lambda,\gamma) := c + \left(1-\frac{1}{p}\right)\left[\left(\frac{|\gamma|}{\lambda p}\right)^{\frac{1}{p-1}}|\gamma| - \left(\frac{|1-\gamma|}{\lambda p}\right)^{\frac{1}{p-1}}|1-\gamma|\right].$$

2. Calculate $\mathbb{E}^F[H_{\lambda,\gamma}(Y)]$.

Denoted $f(\lambda,\gamma) = \mathbb{E}^F[H_{\lambda,\gamma}(Y)]$. For $p=1$, if $\lambda \geqslant \frac{1}{2}$ and $1-\lambda \leqslant \gamma \leqslant \lambda$, we have

$$\begin{aligned} f(\lambda,\gamma) &= \mathbb{E}^F[H_{\lambda,\gamma}(Y)] \\ &= \int_{-\infty}^{c} -\gamma y \,\mathrm{d}F(y) + \int_{c}^{+\infty} -\lambda y + (\lambda-\gamma)c \,\mathrm{d}F(y) \\ &= -\gamma\mu_F - (\lambda-\gamma)\int_{c}^{+\infty}(y-c)\,\mathrm{d}F(y) \\ &= -\gamma\mu_F - (\lambda-\gamma)\mu_F^{c+}. \end{aligned}$$

Otherwise, $f(\lambda,\gamma) = \infty$ because $H_{\lambda,\gamma}(y) = \infty$ for all $y \in \mathbb{R}$. For $p > 1$, we find that

$$\begin{aligned} f(\lambda,\gamma) &= \mathbb{E}^F[H_{\lambda,\gamma}(Y)] \\ &= \int_{-\infty}^{c_1(\lambda,\gamma)} -\gamma y + |\gamma|\left(1-\frac{1}{p}\right)\left(\frac{|\gamma|}{\lambda p}\right)^{\frac{1}{p-1}} \mathrm{d}F(y) \\ &\quad + \int_{c_1(\lambda,\gamma)}^{+\infty} y-\gamma y - c + |1-\gamma|\left(1-\frac{1}{p}\right)\left(\frac{|1-\gamma|}{\lambda p}\right)^{\frac{1}{p-1}} \mathrm{d}F(y) \\ &= -\gamma\mu_F + |\gamma|\left(1-\frac{1}{p}\right)\left(\frac{|\gamma|}{\lambda p}\right)^{\frac{1}{p-1}} + \int_{c_1(\lambda,\gamma)}^{+\infty} y - c_1(\lambda,\gamma)\,\mathrm{d}F(y) \\ &= -\gamma\mu_F + |\gamma|\left(1-\frac{1}{p}\right)\left(\frac{|\gamma|}{\lambda p}\right)^{\frac{1}{p-1}} + \mathbb{E}^F[(Y-c_1(\lambda,\gamma))_+]. \end{aligned}$$

https://doi.org/10.1017/S0269964823000141 Published online by Cambridge University Press

Ultimately,

$$T(\mu) = \sup_{G\in\mathcal{S}_2} \mathbb{E}^G[(X-c)_+]$$
$$= \begin{cases} \inf_{\lambda\geqslant\frac{1}{2},1-\lambda\leqslant\gamma\leqslant\lambda} \left\{\lambda(\varepsilon-\mu_F^{c_+})+\gamma(\mu-\mu_F+\mu_F^{c_+})\right\}, & p=1, \\ \inf\limits_{\lambda\geqslant 0,\gamma\in\mathbb{R}} \Big\{\lambda\varepsilon^p+\gamma(\mu-\mu_F) \\ \qquad\qquad +|\gamma|\left(1-\frac{1}{p}\right)\left(\frac{|\gamma|}{\lambda p}\right)^{\frac{1}{p-1}}+\mathbb{E}^F[(Y-c_1(\lambda,\gamma))_+]\Big\}, & p>1. \end{cases}$$

3. Calculate $\underline{\Omega}_X^F(c,\mu,\varepsilon)$ in the case of $p=1$.

We denote $g(\lambda,\gamma) := \lambda(\varepsilon-\mu_F^{c_+})+\gamma(\mu-\mu_F+\mu_F^{c_+})$. Fixed λ, the partial derivative of $g(\lambda,\gamma)$ with respect to γ is

$$\frac{\partial g(\lambda,\gamma)}{\partial\gamma} = \mu-\mu_F+\mu_F^{c_+}.$$

Since $\mu > \mu_F-\mu_F^{c_+}$ yields $\frac{\partial g(\lambda,\gamma)}{\partial\gamma} > 0$, the minimizer to $g(\lambda,\gamma)$ is $\gamma^* = 1-\lambda$ and

$$g(\lambda,1-\lambda) = \lambda(\varepsilon+\mu_F-\mu-2\mu_F^{c_+})+(\mu-\mu_F+\mu_F^{c_+}).$$

One can verify that

$$\frac{\partial g(\lambda,1-\lambda)}{\partial\lambda} = \varepsilon+\mu_F-\mu-2\mu_F^{c_+}.$$

Noting $\mu > \varepsilon+\mu_F-2\mu_F^{c_+}$ yields $\frac{\partial g(\lambda,1-\lambda)}{\partial\lambda} < 0$, the minimizer to $g(\lambda,1-\lambda)$ is $\lambda^* = \infty$ and the minimum value of $g(\lambda,1-\lambda)$ is $-\infty$. Conversely, if $\mu \leqslant \varepsilon+\mu_F-2\mu_F^{c_+}$, as $\frac{\partial g(\lambda,1-\lambda)}{\partial\lambda} \geqslant 0$, the minimizer to $g(\lambda,1-\lambda)$ is $\lambda^* = \frac{1}{2}$ and $g\left(\frac{1}{2},\frac{1}{2}\right) = \frac{1}{2}(\mu-\mu_F+\varepsilon)$. On the other hand, if $\mu \leqslant \mu_F-\mu_F^{c_+}$, as $\frac{\partial g(\lambda,\gamma)}{\partial\gamma} \leqslant 0$, the minimizer to $g(\lambda,\gamma)$ is $\gamma^* = \lambda$ and

$$g(\lambda,\lambda) = \lambda(\mu-\mu_F+\varepsilon).$$

In addition, we have

$$\frac{\partial g(\lambda,\lambda)}{\partial\lambda} = \mu-\mu_F+\varepsilon.$$

Then $\mu \geqslant \mu_F-\varepsilon$ yields $\frac{\partial g(\lambda,\lambda)}{\partial\lambda} \geqslant 0$ and the minimizer to $g(\lambda,\lambda)$ is $\lambda^* = \frac{1}{2}$ and $g\left(\frac{1}{2},\frac{1}{2}\right) = \frac{1}{2}(\mu-\mu_F+\varepsilon)$. If $\mu < \mu_F-\varepsilon$, we have $\frac{\partial g(\lambda,\lambda)}{\partial\lambda} < 0$, then the minimizer to $g(\lambda,\lambda)$ is $\lambda^* = \infty$ and the minimum value of $g(\lambda,\lambda)$ is $-\infty$. To sum up,

$$T(\mu) = \inf_{\lambda\geqslant\frac{1}{2},1-\lambda\leqslant\gamma\leqslant\lambda} g(\lambda,\gamma) = \begin{cases} -\infty, & \text{if } \mu\in\mathcal{A}_1\cup\mathcal{B}_1, \\ \frac{1}{2}(\mu-\mu_F+\varepsilon), & \text{if } \mu\in\mathcal{A}_2\cup\mathcal{B}_2, \end{cases}$$

where $\mathcal{A}_1$, $\mathcal{A}_2$, $\mathcal{B}_1$ and $\mathcal{B}_2$ are defined in Theorem 2.6. As a result,

https://doi.org/10.1017/S0269964823000141 Published online by Cambridge University Press

$$\underline{\Omega}_X^F(c,\mu,\varepsilon) = \left(1 - \frac{\mu - c}{T(\mu)}\right)^{-1} = \begin{cases} 1, & \text{if } \mu \in \mathcal{A}_1 \cup \mathcal{B}_1, \\ \left[1 - \frac{2(\mu-c)}{\mu - \mu_F + \varepsilon}\right]^{-1}, & \text{if } \mu \in \mathcal{A}_2 \cup \mathcal{B}_2. \end{cases}$$ □

In Theorem 2.6, we provide an explicit solution to problem (7) and derive a more concise optimization problem for $p > 1$. In the next section, we will apply problems (1) and (7) in portfolio selection.

3. Application in portfolio selection

In this section, we explore interesting properties of the distributionally robust portfolio optimization problem under the mean-variance and Wasserstein uncertainty sets. Throughout this section, suppose that $\mathbb{W}$ is a subset of $\mathbb{R}^d$. Let $\mathbf{X} = (X_1, \ldots, X_d)^\top \in \mathcal{P}(\mathbb{R}^d)$ and $\mathbf{w} = (w_1, \ldots, w_d)^\top \in \mathbb{W}$, where X_i represents the return of the ith asset, w_i represents the investment weight of the ith asset. The distribution of $\mathbf{w}^\top \mathbf{X}$ is denoted by $F_{\mathbf{w}}$ when the distribution of $\mathbf{X}$ is F.

3.1. Mean-variance uncertainty set

We consider the distributionally robust portfolio optimization problem under mean-variance uncertainty set:

$$\sup_{\mathbf{w}\in\mathbb{W}} \inf_{F\in\mathcal{S}(\boldsymbol{\mu},\Sigma)} \Omega^F_{\mathbf{w}^\top\mathbf{X}}(c), \tag{9}$$

where

$$\mathcal{S}(\boldsymbol{\mu},\Sigma) := \{F \in \mathcal{P}_2(\mathbb{R}^d) : \mathbb{E}^F[\mathbf{X}] = \boldsymbol{\mu},\ \mathrm{Var}^F(\mathbf{X}) = \Sigma\},$$

and $\Omega^F_{\mathbf{w}^\top\mathbf{X}}(c)$ represents the Omega ratio of $\mathbf{w}^\top\mathbf{X}$ that is calculated under the constraint that the distribution of $\mathbf{X}$ is F.

In the following theorem, we solve the inner problem of (9) and give an equivalent optimization problem with the same optimal solution as problem (9).

Theorem 3.1. *Given $c \in \mathbb{R}$, $\boldsymbol{\mu} \in \mathbb{R}^d$ and Σ, a positive semidefinite matrix of $\mathbb{R}^{d\times d}$, assuming $\mathbf{w}^\top\boldsymbol{\mu} \geqslant c$ for $\mathbf{w} \in \mathbb{W}$, problem (9) is equivalent to the following optimization problem:*

$$\sup_{\mathbf{w}\in\mathbb{W}} \frac{\mathbf{w}^\top\boldsymbol{\mu} - c}{\sqrt{\mathbf{w}^\top\Sigma\mathbf{w}}}, \tag{10}$$

in the sense that two problems have the same optimal solution.

To prove Theorem 3.1, we need the following lemma from [5].

Lemma 3.2. *For $\mathbf{w} \in \mathbb{R}^d$, it holds that*

$$\mathcal{S}(\mathbf{w},\boldsymbol{\mu},\Sigma) = \mathcal{S}_{\mathbf{w}}(\boldsymbol{\mu},\Sigma),$$

where

$$\mathcal{S}(\mathbf{w},\boldsymbol{\mu},\Sigma) = \{F_{\mathbf{w}} \in \mathcal{P}_2(\mathbb{R}) : F \in \mathcal{S}(\boldsymbol{\mu},\Sigma)\ F_{\mathbf{w}} \text{ is the distribution of } \mathbf{w}^\top\mathbf{X} \text{ when the distribution of } \mathbf{X} \text{ is } F\}$$

https://doi.org/10.1017/S0269964823000141 Published online by Cambridge University Press

and

$$\mathcal{S}_{\mathbf{w}}(\boldsymbol{\mu}, \Sigma) = \{F \in \mathcal{P}_2(\mathbb{R}) : \mathbb{E}^F[\mathbf{X}] = \mathbf{w}^\top \boldsymbol{\mu}, \mathrm{Var}^F(\mathbf{X}) = \mathbf{w}^\top \Sigma \mathbf{w}\}.$$

Proof of Theorem 3.1 The core of the robust optimization problem in portfolio is solving for the internal worst-case Omega ratio under the uncertainty set $\mathcal{S}(\boldsymbol{\mu}, \Sigma)$. By Lemma 3.2, we have

$$\inf_{F \in \mathcal{S}(\boldsymbol{\mu}, \Sigma)} \Omega^F_{\mathbf{w}^\top \mathbf{X}}(c) = \inf_{F_{\mathbf{w}} \in \mathcal{S}(\mathbf{w}, \boldsymbol{\mu}, \Sigma)} \Omega^{F_{\mathbf{w}}}_{\mathbf{w}^\top \mathbf{X}}(c) = \inf_{G \in \mathcal{S}_{\mathbf{w}}(\boldsymbol{\mu}, \Sigma)} \Omega^G_Y(c).$$

Hence, for $\mathbf{w}^\top \boldsymbol{\mu} \geqslant c$, one can verify that

$$\underline{\Omega}_{\mathbf{X}}(\mathbf{w}, c, \boldsymbol{\mu}, \Sigma) := \inf_{F \in \mathcal{S}(\boldsymbol{\mu}, \Sigma)} \Omega^F_{\mathbf{w}^\top \mathbf{X}}(c) = \inf_{G \in \mathcal{S}_{\mathbf{w}}(\boldsymbol{\mu}, \Sigma)} \Omega^G_Y(c) = \min_{G \in \mathcal{S}_{\mathbf{w}}(\boldsymbol{\mu}, \Sigma)} \Omega^G_Y(c) = \frac{\sqrt{1 + S_{\mathbf{w}}^2} + S_{\mathbf{w}}}{\sqrt{1 + S_{\mathbf{w}}^2} - S_{\mathbf{w}}}$$

directly by Theorem 2.2, where $S_{\mathbf{w}} = \frac{\mathbf{w}^\top \boldsymbol{\mu} - c}{\sqrt{\mathbf{w}^\top \Sigma \mathbf{w}}}$. The above conclusion gives the worst-case Omega ratio with respect to the weight $\mathbf{w}$. The monotonic progression of $\underline{\Omega}_{\mathbf{X}}(\mathbf{w}, c, \boldsymbol{\mu}, \Sigma)$ with respect to $S_{\mathbf{w}}$ is obvious, implying that the worst-case Omega ratio problem in the portfolio is equivalent to the supremum of the Sharpe ratio $S_{\mathbf{w}}$. Thus, we complete the proof. □

Remark 2. We consider the situation where the expected return is greater than the risk-free return, that is, $\mathbf{w}^\top \boldsymbol{\mu} \geqslant c$. Therefore, in Theorem 3.1, we simplify problem (9) in the case of $\mathbf{w}^\top \boldsymbol{\mu} \geqslant c$. For $\mathbf{w}^\top \boldsymbol{\mu} < c$, by Theorem 2.2, $\underline{\Omega}_{\mathbf{X}}(\mathbf{w}, c, \boldsymbol{\mu}, \Sigma) = 0$, and thus the optimal value of problem (9) is 0.

Remark 3. Given $c \in \mathbb{R}$, $\boldsymbol{\mu}_1, \boldsymbol{\mu}_2 \in \mathbb{R}^d$ and two positive semidefinite matrices $\Sigma_1, \Sigma_2 \in \mathbb{R}^{d \times d}$, assume $\boldsymbol{\mu}_1 \preceq \boldsymbol{\mu}_2$, $\Sigma_1 \preceq \Sigma_2$ [2] and $\mathbf{w}^\top \boldsymbol{\mu}_1 \geqslant c$ for $\mathbf{w} \in \mathbb{W}$. If both the mean and covariance matrix are included in the box uncertainty set, that is,

$$\mathcal{S}_1(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \Sigma_1, \Sigma_2) := \{F \in \mathcal{P}_2(\mathbb{R}^d) : \boldsymbol{\mu}_1 \preceq \mathbb{E}^F[\mathbf{X}] \preceq \boldsymbol{\mu}_2, \Sigma_1 \preceq \mathrm{Var}^F(\mathbf{X}) \preceq \Sigma_2\},$$

then we can obtain similar results by applying the result in Theorem 3.1. We consider the distributionally robust portfolio optimization problem under the new mean-variance uncertainty set:

$$\sup_{\mathbf{w} \in \mathbb{W}} \inf_{F \in \mathcal{S}_1(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \Sigma_1, \Sigma_2)} \Omega^F_{\mathbf{w}^\top \mathbf{X}}(c), \tag{11}$$

where $\Omega^F_{\mathbf{w}^\top \mathbf{X}}(c)$ represents the Omega ratio of $\mathbf{w}^\top \mathbf{X}$ that is calculated under the constraint that the distribution of $\mathbf{X}$ is F. One can verify that

$$\inf_{F \in \mathcal{S}_1(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \Sigma_1, \Sigma_2)} \Omega^F_{\mathbf{w}^\top \mathbf{X}}(c) = \inf_{\substack{\boldsymbol{\mu}_1 \preceq \boldsymbol{\mu} \preceq \boldsymbol{\mu}_2 \\ \Sigma_1 \preceq \Sigma \preceq \Sigma_2}} \inf_{F \in \mathcal{S}(\boldsymbol{\mu}, \Sigma)} \Omega^F_{\mathbf{w}^\top \mathbf{X}}(c) = \inf_{\substack{\boldsymbol{\mu}_1 \preceq \boldsymbol{\mu} \preceq \boldsymbol{\mu}_2 \\ \Sigma_1 \preceq \Sigma \preceq \Sigma_2}} \frac{\sqrt{1 + S_{\mathbf{w}}^2} + S_{\mathbf{w}}}{\sqrt{1 + S_{\mathbf{w}}^2} - S_{\mathbf{w}}},$$

where $S_{\mathbf{w}} = \frac{\mathbf{w}^\top \boldsymbol{\mu} - c}{\sqrt{\mathbf{w}^\top \Sigma \mathbf{w}}}$. Note that $S_{\mathbf{w}}$ is increasing with respect to $\boldsymbol{\mu}$ and decreasing with respect to Σ. Furthermore, $\frac{\sqrt{1+S_{\mathbf{w}}^2}+S_{\mathbf{w}}}{\sqrt{1+S_{\mathbf{w}}^2}-S_{\mathbf{w}}}$ is increasing with respect to $S_{\mathbf{w}}$. Thus, problem (11) is equivalent to the

[2] For two vectors $\boldsymbol{\mu}_1, \boldsymbol{\mu}_2 \in \mathbb{R}^d$, the notation $\boldsymbol{\mu}_1 \preceq \boldsymbol{\mu}_2$ means that each component of $\boldsymbol{\mu}_1$ is no more than the corresponding component of $\boldsymbol{\mu}_2$. For two positive semidefinite matrices $\Sigma_1, \Sigma_2 \in \mathbb{R}^{d \times d}$, the notation $\Sigma_1 \preceq \Sigma_2$ means that $\Sigma_2 - \Sigma_1$ is a positive semidefinite matrix.

https://doi.org/10.1017/S0269964823000141 Published online by Cambridge University Press

following optimization problem:

$$\sup_{\mathbf{w}\in\mathbb{W}} \frac{\mathbf{w}^\top \boldsymbol{\mu}_1 - c}{\sqrt{\mathbf{w}^\top \Sigma_2 \mathbf{w}}}, \tag{12}$$

in the sense that two problems have the same optimal solution.

3.2. Mean-Wasserstein uncertainty set

We consider a distributionally robust portfolio optimization problem based on Omega ratio in the portfolio under the mean-Wasserstein uncertainty set. Let $F \in \mathcal{P}_p(\mathbb{R}^d)$ be a prespecified distribution used as a benchmark. Assume the distribution of $\mathbf{X}$, defined at the beginning of Section 3, is G, which satisfies $W_p(F, G) \leqslant \varepsilon$. The distribution of $\mathbf{w}^\top \mathbf{X}$ is denoted by $G_{\mathbf{w}}$ when the distribution of $\mathbf{X}$ is G. Then, we propose the distributionally robust portfolio optimization problem under mean-Wasserstein uncertainty set:

$$\sup_{\mathbf{w}\in\mathbb{W}} \inf_{G\in\mathcal{S}^d_{p,\varepsilon}(\boldsymbol{\mu},F)} \Omega^G_{\mathbf{w}^\top\mathbf{X}}(c) \tag{13}$$

where $\varepsilon > 0$, $F \in \mathcal{P}_p(\mathbb{R}^d)$,

$$\mathcal{S}^d_{p,\varepsilon}(\boldsymbol{\mu}, F) := \{G \in \mathscr{B}^p_\varepsilon(F) : \mathbb{E}^G[\mathbf{X}] = \boldsymbol{\mu}\},$$

and $\Omega^G_{\mathbf{w}^\top\mathbf{X}}(c)$ represents the Omega ratio of $\mathbf{w}^\top\mathbf{X}$ that is calculated under the constraint that the distribution of X is G.

In the following theorem, we give the explicit solution to the inner problem of Eq. (13) for $p = 1$ and simplify the inner problem of Eq. (13) for $p > 1$.

Theorem 3.5. *For $p \geqslant 1$, let $F \in \mathcal{P}_p(\mathbb{R}^d)$ be a benchmark distribution, $\mathbb{E}^F[\mathbf{X}] = \boldsymbol{\mu}$ and c be a sure payoff. Suppose $\mathbf{w}^\top\boldsymbol{\mu} \geqslant c$ for $\mathbf{w} \in \mathbb{W}$, and denote*

$$\sup_{\mathbf{w}\in\mathbb{W}} \inf_{G\in\mathcal{S}^d_{p,\varepsilon}(\boldsymbol{\mu},F)} \Omega^G_{\mathbf{w}^\top\mathbf{X}}(c) := \sup_{\mathbf{w}\in\mathbb{W}} \underline{\Omega}^F_{\mathbf{X}}(\mathbf{w}, c, \boldsymbol{\mu}, \varepsilon). \tag{14}$$

(i) If $p = 1$, then we have

$$\underline{\Omega}^F_{\mathbf{X}}(\mathbf{w}, c, \boldsymbol{\mu}, \varepsilon) = \begin{cases} 1, & \text{if } \mathbf{w}^\top\boldsymbol{\mu} \in \mathcal{A}^*_1 \cup \mathcal{B}^*_1, \\ \left[1 - \dfrac{2(\mathbf{w}^\top\boldsymbol{\mu} - c)}{\mathbf{w}^\top\boldsymbol{\mu} - \mu_{F_{\mathbf{w}}} + \varepsilon}\right]^{-1}, & \text{if } \mathbf{w}^\top\boldsymbol{\mu} \in \mathcal{A}^*_2 \cup \mathcal{B}^*_2, \end{cases}$$

where $\mathcal{A}^*_1 = \{\mu : \mu > \mu_{F_{\mathbf{w}}} - \mu^{c_+}_{F_{\mathbf{w}}}, \mu > \mu_{F_{\mathbf{w}}} - 2\mu^{c_+}_{F_{\mathbf{w}}} + \varepsilon\}$, $\mathcal{A}^*_2 = \{\mu : \mu > \mu_{F_{\mathbf{w}}} - \mu^{c_+}_{F_{\mathbf{w}}}, \mu \leqslant \mu_{F_{\mathbf{w}}} - 2\mu^{c_+}_{F_{\mathbf{w}}} + \varepsilon\}$, $\mathcal{B}^*_1 = \{\mu : \mu \leqslant \mu_{F_{\mathbf{w}}} - \mu^{c_+}_{F_{\mathbf{w}}}, \mu < \mu_{F_{\mathbf{w}}} - \varepsilon\}$ and $\mathcal{B}^*_2 = \{\mu : \mu \leqslant \mu_{F_{\mathbf{w}}} - \mu^{c_+}_{F_{\mathbf{w}}}, \mu \geqslant \mu_{F_{\mathbf{w}}} - \varepsilon\}$.

(ii) If $p > 1$, then we have

$$\underline{\Omega}^F_{\mathbf{X}}(\mathbf{w}, c, \boldsymbol{\mu}, \varepsilon) = \left[1 - \frac{\mathbf{w}^\top\boldsymbol{\mu} - c}{T(\mathbf{w}^\top\boldsymbol{\mu})}\right]^{-1},$$

https://doi.org/10.1017/S0269964823000141 Published online by Cambridge University Press

where

$$T(\mathbf{w}^\top \boldsymbol{\mu}) = \inf_{\lambda \geqslant 0, \gamma \in \mathbb{R}} \Big\{ \lambda(\|\mathbf{w}\|_* \varepsilon)^p + \gamma(\mathbf{w}^\top \boldsymbol{\mu} - \mu_{F_\mathbf{w}}) + |\gamma| \left(1 - \frac{1}{p}\right) \left(\frac{|\gamma|}{\lambda p}\right)^{\frac{1}{p-1}} + \mathbb{E}^F[(\mathbf{w}^\top \mathbf{X} - c_1(\lambda, \gamma))_+] \Big\},$$

and $c_1(\lambda, \gamma)$ *is defined in Theorem* 2.6.

For the benchmark distribution F, define

$$\mathcal{F}_{\mathbf{w},p,\varepsilon}(F) = \{G_{\mathbf{w}^\top \mathbf{X}} \in \mathcal{P}_p(\mathbb{R}) : G \in \mathscr{B}^p_\varepsilon(F), \; G_\mathbf{w} \text{ is the distribution of } \mathbf{w}^\top \mathbf{X} \text{ when the distribution of } \mathbf{X} \text{ is } G\}.$$

The following lemma can be found in [14], which will be used in the proof of Theorem 3.5.

Lemma 3.6. *For* $\varepsilon \geqslant 0$, $p \geqslant 1$ *and* $\mathbf{w} \neq \mathbf{0}$, *assume a prespecified distribution* $F \in \mathcal{P}_p(\mathbb{R}^d)$, *it holds that*

$$\mathcal{F}_{\mathbf{w},p,\varepsilon}(F) = \mathscr{B}^p_{\|\mathbf{w}\|_* \varepsilon}(F_\mathbf{w}),$$

where $\|\cdot\|_*$ *is the dual norm of* $\|\cdot\|$.

Now we use the projection result in Lemma 3.6 to prove Theorem 3.5.

Proof of Theorem 3.5 We focus on a portfolio selection problem where the unknown distribution of $\mathbf{X}$ is in the uncertainty set $\mathcal{S}_4$. For $F \in \mathcal{P}_p(\mathbb{R}^d)$, let

$$\mathcal{S}_{p,\varepsilon}(\mathbf{w}, \boldsymbol{\mu}, F) = \{G_\mathbf{w} \in \mathcal{P}_p(\mathbb{R}) : G \in \mathcal{S}^d_{p,\varepsilon}(\boldsymbol{\mu}, F), \; G_\mathbf{w} \text{ is the distribution of } \mathbf{w}^\top \mathbf{X} \text{ when the distribution of } \mathbf{X} \text{ is } G\}.$$

By the Lemma 3.6, we have $\mathcal{F}_{\mathbf{w},p,\varepsilon}(F) = \mathscr{B}^p_{\|\mathbf{w}\|_* \varepsilon}(F_\mathbf{w})$. It directly applies that $\mathcal{S}_{p,\varepsilon}(\mathbf{w}, \boldsymbol{\mu}, F) = \mathcal{S}^1_{p,\|\mathbf{w}\|_* \varepsilon}(\mathbf{w}^\top \boldsymbol{\mu}, F_\mathbf{w})$. As a result,

$$\inf_{G \in \mathcal{S}^d_{p,\varepsilon}(\boldsymbol{\mu}, F)} \Omega^G_{\mathbf{w}^\top \mathbf{X}}(c) = \inf_{G_\mathbf{w} \in \mathcal{S}_{p,\varepsilon}(\mathbf{w}, \boldsymbol{\mu}, F)} \Omega^{G_\mathbf{w}}_{\mathbf{w}^\top \mathbf{X}}(c) = \inf_{H \in \mathcal{S}^1_{p,\|\mathbf{w}\|_* \varepsilon}(\mathbf{w}^\top \boldsymbol{\mu}, F_\mathbf{w})} \Omega^H_Y(c).$$

Hence, for $\mathbf{w}^\top \boldsymbol{\mu} \geqslant c$, applying Theorem 2.6 to $\mathbf{w}^\top \mathbf{X}$, the result follows immediately. □

In Theorem 3.5, we apply problem (7) in portfolio selection. In Section 3.3, we will discuss finite sample guarantee based on Wasserstein ball.

3.3. Finite sample guarantee based on Wasserstein ball

In portfolio simulation, the distribution of portfolio assets is unknown, and we often use an empirical distribution function $\hat{F}_N$ to approximate the true distribution of assets. This uncertainty set provides attractive performance guarantees under the common light tail assumption of generating a distribution F over unknown data.

https://doi.org/10.1017/S0269964823000141 Published online by Cambridge University Press

Assumption 3.7. (Light-tailed distribution). *For $F \in \mathcal{P}_p(\mathbb{R}^d)$, there exists an exponent $a > p$ such that*

$$A := \mathbb{E}^F[\exp(\|X\|^a)] = \int_{\mathbb{R}^d} \exp(\|\xi\|^a) F(d\xi) < +\infty.$$

Assumption 3.7 essentially requires that the tails of the distribution F decay at an exponential rate. Let $\mathbb{W}_1 = \{\mathbf{w} \in \mathbb{R}^d : w_i \geqslant 0,\ i = 1, \ldots, d,\ \sum_{i=1}^d w_i = 1\}$ and $\mathbf{e}_{1/d} = (1/d, \ldots, 1/d)^\top$. We have $U_{\mathbb{W}_1} := \sup_{\mathbf{w} \in \mathbb{W}_1} \|\mathbf{w}\|_* = 1 < \infty$ and $L_{\mathbb{W}_1} := \inf_{\mathbf{w} \in \mathbb{W}_1} \|\mathbf{w}\|_* = \left\|\mathbf{e}_{1/d}\right\|_* > 0$. For $\mathbf{w} \in \mathbb{W}_1$, $\hat{F}_{\mathbf{w},N}$ is constructed with i.i.d samples in the population whose true distribution is $F_{\mathbf{w}}$. We have the following lemma from Theorem 2 of Fournier and Guillin [7].

Lemma 3.8. *If Assumption 3.7 holds, then for any $\mathbf{w} \in \mathbb{W}_1$, $N \geqslant 1$ and $\varepsilon > 0$, we have*

$$\mathbb{P}^N\left(W_p\left(F_{\mathbf{w}}, \hat{F}_{\mathbf{w},N}\right) \geqslant \varepsilon\right) \leqslant \begin{cases} c_1 \exp\left(-c_2 N \varepsilon^2\right), & \text{if } \varepsilon \leqslant 1, \\ c_1 \exp\left(-c_2 N \varepsilon^{a/p}\right), & \text{if } \varepsilon > 1, \end{cases} \tag{15}$$

where c_1, c_2 are positive constants that only depend on a, A and p.

For fixed $\mathbf{w} \in \mathbb{W}_1$, Lemma 3.8 provides an a priori estimate of the probability that the unknown data generating distribution $F_{\mathbf{w}}$ lies outside the Wasserstein ball $\mathscr{B}_\varepsilon^p(\hat{F}_{\mathbf{w},N})$. Equating the RHS of Eq. (15) to β and solving for ε yields

$$\varepsilon_{p,N}(\beta) := \begin{cases} \left(\frac{\log(c_1 \beta^{-1})}{c_2 N}\right)^{1/2}, & \text{if } N \geqslant \frac{\log(c_1 \beta^{-1})}{c_2}, \\ \left(\frac{\log(c_1 \beta^{-1})}{c_2 N}\right)^{p/a}, & \text{if } N < \frac{\log(c_1 \beta^{-1})}{c_2}. \end{cases} \tag{16}$$

Thus, we can use Lemma 3.8 to estimate the radius of the smallest Wasserstein ball containing distribution $F_{\mathbf{w}}$ with $1-\beta$ confidence for $\beta \in (0,1)$. Specifically, if Assumption 3.7 holds, then for any $\mathbf{w} \in \mathbb{W}_1$, $N \geqslant 1$ and $\beta \in (0,1)$, we have

$$\mathbb{P}^N\left(W_p\left(F_{\mathbf{w}}, \hat{F}_{\mathbf{w},N}\right) \leqslant \varepsilon_{p,N}(\beta)\right) \geqslant 1 - \beta, \tag{17}$$

where $\varepsilon_{p,N}(\beta)$ is defined in Eq. (16). Then the concentration inequality gives rise to the following finite sample guarantee.

Theorem 3.9. (Finite sample guarantee). *Suppose Assumption 3.7 holds. For $\beta \in (0,1)$, let $\varepsilon_{p,N}(\beta)$ be defined in (16). Then we have*

$$\mathbb{P}^N\left(\Omega^F_{\mathbf{w}^\top \mathbf{X}}(c) \geqslant \inf_{G \in \mathscr{B}^p_{\varepsilon_{p,N}(\beta)/L_{\mathbb{W}_1}}(\hat{F}_N)} \Omega^G_{\mathbf{w}^\top \mathbf{X}}(c)\right) \geqslant 1 - \beta,\ \forall \mathbf{w} \in \mathbb{W}_1.$$

Proof. Denoted by $\varepsilon^* := \varepsilon_{p,N}(\beta)/L_{\mathbb{W}_1}$. One can verify that $F_{\mathbf{w}} \in \mathcal{F}_{\mathbf{w},p,\varepsilon^*}(\hat{F}_N)$ implies that

$$\Omega^F_{\mathbf{w}^\top \mathbf{X}}(c) \geqslant \inf_{G \in \mathscr{B}^p_{\varepsilon^*}(\hat{F}_N)} \Omega^G_{\mathbf{w}^\top \mathbf{X}}(c).$$

https://doi.org/10.1017/S0269964823000141 Published online by Cambridge University Press

Then, we have

$$\mathbb{P}^N\left(\Omega^F_{\mathbf{w}^\top\mathbf{X}}(c) \geqslant \inf_{G\in\mathscr{B}^p_{\varepsilon^*}(\hat{F}_N)} \Omega^G_{\mathbf{w}^\top\mathbf{X}}(c)\right) \geqslant \mathbb{P}^N\left(F_{\mathbf{w}} \in \mathcal{F}_{\mathbf{w},p,\varepsilon^*}(\hat{F}_N)\right).$$

By Lemma 3.6, we obtain

$$\mathbb{P}^N\left(F_{\mathbf{w}} \in \mathcal{F}_{\mathbf{w},p,\varepsilon^*}(\hat{F}_N)\right) = \mathbb{P}^N\left(F_{\mathbf{w}} \in \mathscr{B}^p_{\|\mathbf{w}\|_*\varepsilon^*}(\hat{F}_{\mathbf{w},N})\right) = \mathbb{P}^N\left(W_p\left(F_{\mathbf{w}}, \hat{F}_{\mathbf{w},N}\right) \leqslant \|\mathbf{w}\|_*\varepsilon^*\right).$$

Since $\varepsilon^* = \varepsilon_{p,N}(\beta)/L_{\mathbb{W}_1}$, we have

$$\mathbb{P}^N\left(W_p\left(F_{\mathbf{w}}, \hat{F}_{\mathbf{w},N}\right) \leqslant \|\mathbf{w}\|_*\varepsilon^*\right) \geqslant \mathbb{P}^N\left(W_p\left(F_{\mathbf{w}}, \hat{F}_{\mathbf{w},N}\right) \leqslant \varepsilon_{p,N}(\beta)\right).$$

Thus, one can verify that

$$\mathbb{P}^N\left(\Omega^F_{\mathbf{w}^\top\mathbf{X}}(c) \geqslant \inf_{G\in\mathscr{B}^p_{\varepsilon_{p,N}(\beta)/L_{\mathbb{W}_1}}(\hat{F}_N)} \Omega^G_{\mathbf{w}^\top\mathbf{X}}(c)\right) \geqslant 1-\beta$$

directly from inequality (17). □

In this section, we first proposed a concentration inequality for one-dimensional Wasserstein ball. The concentration inequality provided the finite sample guarantee based on one-dimensional Wasserstein ball.

4. Numerical experiments

In this section, we examine the performance of our models and compare them with the classical (non-robust) Omega ratio (OR) model as well as two alternative distributionally robust portfolio optimization models proposed by [3] and [16], respectively. We first give some notations: μ_i denotes the expected return on asset i, r_0 denotes the expected rate of return on the portfolio, w_i denotes the proportion of asset X_i in the portfolio ($0 \leqslant w_i \leqslant 1$) and $\mathrm{Cov}(X_i, X_j)$ denotes the covariance between the return on asset X_i and the return on asset X_j for $i, j = 1, \ldots, d$. Let $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_d)^\top$, $\Sigma = (\mathrm{Cov}(X_i, X_j))_{d\times d}$ and $\mathbb{W}_1 = \{\mathbf{w} \in \mathbb{R}^d : w_i \geqslant 0,\ i = 1, \ldots, d,\ \sum_{i=1}^d w_i = 1\}$. Below we provide the specific expressions for the models mentioned above to obtain optimal solutions.

Classical (non-robust) OR optimization model:

$$\max_{\mathbf{w}\in\mathbb{W}_1} \Omega_{\mathbf{w}^\top\mathbf{X}}(c) = \frac{\sum_{i=1}^d (w_i\mu_i - c)_+}{\sum_{i=1}^d (w_i\mu_i - c)_-} \tag{18}$$

$$\text{s.t.} \sum_{i=1}^d w_i\mu_i \geqslant r_0,\ \sum_{i=1}^d w_i = 1,\ w_i \geqslant 0,\ i = 1, \ldots, d. \tag{19}$$

Distributionally robust Omega ratio optimization based on mean-variance information (DROR-MV) model (Theorem 3.1):

$$\max_{\mathbf{w}\in\mathbb{W}_1} \frac{\sum_{i=1}^d w_i\mu_i - c}{\sqrt{\sum_{i=1}^d \sum_{j=1}^d w_i w_j \mathrm{Cov}(X_i, X_j)}} \text{ s.t. } (19) \tag{20}$$

https://doi.org/10.1017/S0269964823000141 Published online by Cambridge University Press

Distributionally robust Omega ratio optimization based on mean-Wasserstein ball (DROR-MW) model (Theorem 3.5): Set $p=2$, $a=2$ and $\varepsilon=(\log N/N)^{1/d}$, where d is the dimension of $\mathbf{X}$ and N is the sample size of the data.

$$\begin{aligned}
&\max_{\mathbf{w}\in\mathbb{W}_1}\underline{\Omega}_{\mathbf{X}}^{\hat{F}_N}(\mathbf{w},c,\boldsymbol{\mu},\varepsilon)=\left\{1-\frac{\sum_{i=1}^{d}(w_i\mu_i-c)_+}{T^{\hat{F}_N}(\mathbf{w}^\top\boldsymbol{\mu})}\right\}^{-1}\\
&\text{s.t. } T^{\hat{F}_N}(\mathbf{w}^\top\boldsymbol{\mu})\\
&\quad=\min_{\lambda\geqslant 0,\gamma\in\mathbb{R}}\left\{\lambda(\|\mathbf{w}\|_2\,\varepsilon)^2+\gamma\left(\sum_{i=1}^{d}\mu_i-\mu_{\hat{F}_{\mathbf{w},N}}\right)+\frac{\gamma^2}{4\lambda}+\frac{1}{N}\sum_{j=1}^{N}\left[\sum_{i=1}^{d}w_iX_{ij}-\left(c+\frac{2\gamma-1}{4\lambda}\right)\right]_+\right\},\\
&\quad\sum_{i=1}^{d}w_i\mu_i\geqslant\max(r_0,c),\quad\sum_{i=1}^{d}w_i=1,\quad w_i\geqslant 0,\ i=1,\ldots,d.
\end{aligned}\tag{21}$$

In general, the objective function $\underline{\Omega}_{\mathbf{X}}^{\hat{F}_N}(\mathbf{w},c,\boldsymbol{\mu},\varepsilon)$ is not convex with respect to $\mathbf{w}$. Therefore, DROR-MW model is not a convex optimization problem and may have only locally optimal solutions. We conduct simulations to obtain the optimal portfolio by selecting different initial values that match the actual situation and calculating the final combined weight after optimization.

Blanchet et al. [3] proposed the distributionally robust mean-variance portfolio selection problem with Wasserstein distances. Specifically, the distributionally robust mean-variance (DRMV) model is given by

$$\min_{\mathbf{w}\in\mathbb{W}_1}\max_{F\in\mathscr{B}_\varepsilon^p(\hat{F}_N)}\mathbf{w}^\top\mathrm{Var}^F(\mathbf{X})\mathbf{w}\tag{22}$$

$$\text{s.t. }\min_{F\in\mathscr{B}_\varepsilon^p(\hat{F}_N)}\mathbb{E}^F[\mathbf{w}^\top\mathbf{X}]\geqslant r_0,\quad\sum_{i=1}^{d}w_i=1,\quad w_i\geqslant 0,\ i=1,\ldots,d.\tag{23}$$

Neufeld et al. [16] proposed a fully data-driven markov decision problem under Wasserstein uncertainty set. Specifically, the distributionally robust expected return optimization based on Wasserstein ball (DRER-W) model is given by

$$\sup_{\mathbf{w}\in\mathbb{W}_1}\inf_{F\in\mathscr{B}_\varepsilon^p(\hat{F}_N)}\mathbb{E}^F\left[\sum_{t=1}^{\infty}\sum_{i=1}^{d}\alpha^t w_iX_{t+1}^i\right]\tag{24}$$

$$\text{s.t. }\sum_{i=1}^{d}w_i=1,\quad w_i\geqslant 0,\ i=1,\ldots,d,\tag{25}$$

where X_t^i represents the return of the ith asset at time t and the empirical distribution $\hat{F}_N$ is based on the training data; see [16] for more details.

4.1. Simulation

The forthcoming simulations aim to evaluate the performance of the DROR-MV and DROR-MW models, comparing them to the DRMV and DRER-W models as well as the classical non-robust OR model.

We conduct experiments in the following setting. In each simulation, we generate n realizations of the normal random vector $\mathbf{X}_i\in\mathbb{R}^d$ $(i=1,\ldots,n)$, which satisfies $\mathbf{X}_i\sim N(\boldsymbol{\mu},\Sigma)$. We set $d=10$,

https://doi.org/10.1017/S0269964823000141 Published online by Cambridge University Press

Table 1. *Return performance indices based on different models—Mean, Sharpe ratio, Omega ratio and their respective variances ($\times 10^{-2}$).*

Method	Mean	Sharpe ratio	Omega ratio
DROR-MW	0.033 (2.38)	1.22 (0.19)	1.053 (0.082)
DROR-MV	0.030 (1.81)	1.51 (0.14)	1.032 (0.063)
DRER-W	0.039 (3.17)	1.12 (0.62)	1.013 (0.110)
DRMV	0.026 (1.45)	1.57 (0.13)	0.954 (0.032)
OR	0.031 (2.79)	0.98 (0.48)	1.056 (0.096)

Table 2. *Return performance indices based on different models—Mean, Sharpe ratio, Omega ratio and their respective variances ($\times 10^{-2}$).*

Method	Mean	Sharpe ratio	Omega ratio
DROR-MW	0.054 (2.06)	2.00 (0.35)	1.089 (0.157)
DROR-MV	0.052 (1.72)	2.13 (0.28)	1.074 (0.150)
DRER-W	0.061 (2.96)	1.64 (0.79)	1.027 (0.377)
DRMV	0.038 (1.42)	2.42 (0.16)	1.012 (0.108)
OR	0.047 (2.65)	1.35 (0.73)	1.091 (0.191)

$n = 10^4$, $\boldsymbol{\mu} = (0.0297, 0.039, 0.038, 0.026, 0.023, 0.025, 0.026, 0.036, 0.022, 0.028)^\top$, $\Sigma = \mathrm{diag}(1, \ldots, 1)$, $c = 0.03$. We compute the optimal portfolio of robust models and perform 10 simulation runs to ensure the stability of the simulation results. In each simulation run, we generate 10^4 realizations of standard normal random vectors. All the results are summarized in Table 1. The first column of the table lists the aforementioned models. The second, third and fourth columns show the mean, Sharpe ratio and Omega ratio corresponding to each model, respectively, along with their respective variances denoted within parentheses.

Based on the simulation results, we can draw the following observations: (i) The expected return of the DROR-MV and DROR-MW models is higher than that of the DRMV model but lower than that of the DRER-W model. (ii) Based on the variances corresponding to the mean in the table, it can be inferred that the DROR-MV and DROR-MW models have lower volatility than the DRER-W model but higher volatility than the DRMV model. (iii) The Sharpe ratio of the DROR-MV and DROR-MW models is higher than that of the DRER-W model. Moreover, the DROR-MV and DROR-MW models have higher Omega ratios than other robust models. The results can be attributed to the fact that the DRMV model aims to minimize variance without considering returns, the DRER-W model aims to maximize returns, while the DROR-MV and DROR-MW models aim to maximize worst-case Omega ratio, considering both returns and losses.

Since the previous data were randomly generated from multivariate normal distributions, but in practice the data are usually correlated, we generated the data that follows the $ARMA(p, q)$ model, which is expressed as

$$\mathbf{X}_t = \phi_0 + \phi_1 \mathbf{X}_{t-1} + \cdots + \phi_p \mathbf{X}_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q},$$

where $\{\varepsilon_i\}_{i=1}^t$ is an independent identically distributed zero-mean white noise series. In the second experiment, for each d and n, we generate random variables $\mathbf{X}_i \in \mathbb{R}^d$, $i = 1, \ldots, n$ such that $\mathbf{X}_t = \rho \mathbf{X}_{t-1} + \boldsymbol{\varepsilon}_t + \theta \boldsymbol{\varepsilon}_{t-1}$, where $\mathbf{X}_0 \sim N(\boldsymbol{\mu}, \Sigma)$, $\boldsymbol{\varepsilon}_i \sim N(0, 1)$. We set $\rho = 0.7$, $\theta = 0.01$ and the other parameters were taken to be consistent with the previous simulation.

The results of the simulated data generated based on the ARMA time series are shown in Table 2. The same column name representation as in Table 2 is used. When the simulated data is generated based on the ARMA model and the initial data follows a normal distribution, the portfolio of our robust

https://doi.org/10.1017/S0269964823000141 Published online by Cambridge University Press

Table 3. *The assets from the US stock market.*

Stock code	Return	Stock code	Return	Stock code	Return	Stock code	Return	Stock code	Return
AAPL	X_1	BAC	X_2	CVX	X_3	GOOG	X_4	JNJ	X_5
MSFT	X_6	PEP	X_7	ADBE	X_8	TSM	X_9	WMT	X_{10}

Table 4. *Means, variances and correlation coefficients of stock returns.*

Stock	Mean	Variance	Correlation coefficient									
X_1	0.00162	0.00080	1	0.42	0.49	0.64	0.38	0.53	0.37	0.53	0.53	0.36
X_2	0.00001	0.00341	0.42	1	0.46	0.44	0.39	0.44	0.32	0.5	0.44	0.34
X_3	0.00040	0.00059	0.49	0.46	1	0.54	0.65	0.63	0.59	0.6	0.56	0.5
X_4	0.00067	0.00061	0.64	0.44	0.54	1	0.47	0.58	0.42	0.59	0.53	0.42
X_5	0.00005	0.00017	0.38	0.39	0.65	0.47	1	0.56	0.6	0.52	0.45	0.57
X_6	0.00031	0.00057	0.53	0.44	0.63	0.58	0.56	1	0.53	0.64	0.5	0.45
X_7	0.00007	0.00023	0.37	0.32	0.59	0.42	0.6	0.53	1	0.52	0.48	0.46
X_8	0.00028	0.00078	0.53	0.5	0.6	0.59	0.52	0.64	0.52	1	0.56	0.47
X_9	0.00049	0.00083	0.53	0.44	0.56	0.53	0.45	0.5	0.48	0.56	1	0.48
X_{10}	0.00029	0.00027	0.36	0.34	0.5	0.42	0.57	0.45	0.46	0.47	0.48	1

model outperforms other robust models in terms of both return and robustness. This observation seems to confirm the common denominator.

4.2. Stock market

The DRO framework has been criticized for being conservative, with a pessimistic bias in its forecasting. For instance, Hu et al. [9] trained models that minimize adversarial reweighting and account for worst-case distribution changes. However, the model's predictions are based on the training data performed pessimistically on the test set. To determine whether the proposed framework is overly conservative in the up-trending market and more effective in controlling losses during the down-trending market, we run the aforementioned models separately in bear and bull markets.

We focus on the stock market between 2007 and 2020. The global financial crisis, triggered by the subprime crisis in 2008, led to an overall decline in the stock market, characterized by a bear market. For our analysis, we selected data from 2007 to 2009. In the remaining period (2010 to 2020), the stock market as a whole rose and behaved as a bull market.

4.2.1. Bear market

We consider a selection of well-known stocks from various industry sectors of the US stock market for the empirical analysis, with the selected stock codes shown in Table 3, where $\mathbf{X} = (X_1, \ldots, X_{10})^\top$ is a 10-dimensional vector, where X_i represents the ith stock return for $i = 1, \ldots, 10$.

The daily returns of ten stocks were recorded as the dataset for the numerical experiment. The sample period for these historical return values is from 1 January 2007 to 31 December 2009, with 755 valid observations, that is, $N = 755$. Table 4 gives the means, variances and the correlation coefficients for the ten stocks.

Due to the limited sample size, it is impossible to determine the accurate distribution of stock returns. In this paper, the ambiguous distribution of stock returns is estimated from several observations in the dataset. More specifically, the central distribution of the Wasserstein uncertainty set is a discrete empirical distribution, while the radius of the Wasserstein uncertainty set represents the imprecision

https://doi.org/10.1017/S0269964823000141 Published online by Cambridge University Press

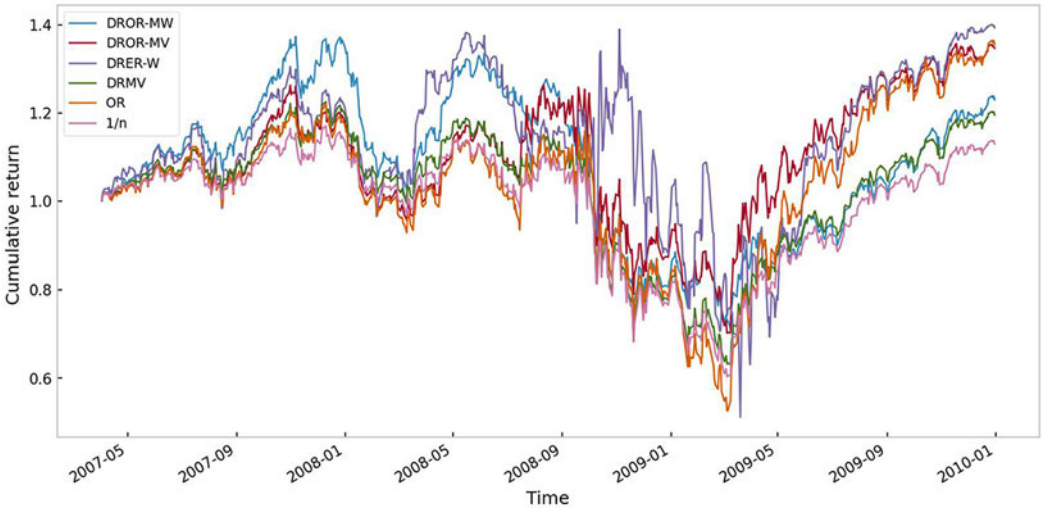

Figure 1. *Cumulative returns of optimal portfolio strategies under different models over the period 2007 to 2009.*

of the distribution. We derive some theoretical results under the proposed mean-variance uncertainty set and the mean-Wasserstein uncertainty set in Sections 2 and 3, respectively. Theorems 3.1 and 3.5, respectively, give the investment weights for maximizing the worst-case Omega ratio under the mean-variance uncertainty set and the mean-Wasserstein uncertainty set, reducing the worst-case impact in the portfolio. Furthermore, the robust optimization problem in Theorems 3.1 and 3.5 is solved to obtain the optimal portfolio strategy.

Based on data from the US bear market, we employed a 30-day sliding time window with an initial wealth of 1. Using the data from the first 30 days, we selected the best-weighted portfolio for day 31 and calculated the returns obtained under the DROR-MV and DROR-MW models. We also considered the DRER-W and DRMV models, the non-robust OR model and the classical $1/n$ portfolio model for comparison. The $1/n$ portfolio model, also known as the equal weighting model, is a strategy that allocates equal capital to each asset in the portfolio and is a simple yet effective approach. The results are shown in Figure 1.

Figure 1 shows the cumulative portfolio returns under the six models. The figure demonstrates that, in terms of model robustness, the DROR-MV and DROR-MW models are more effective at controlling losses during economic decline, followed by the DRMV and $1/n$ models, and lastly, the DRER-W and OR models. This is because the DROR-MV and DROR-MW models aim to maximize the worst-case Omega ratio, which considers both returns and losses. While the DRER-W model performed the best in cumulative returns, it does not account for losses and is less stable in volatile market. The DRMV model minimizes variance, which can result in lower overall returns for portfolios with constant returns and zero variance.

4.2.2. Bull market

We have selected the same 10 stocks with a total of 2,515 valid observations from 2010 to 2020. Table 5 gives important characteristics of the returns of the 10 stocks, including the means, variances and correlation coefficients.

We used the data of the first 30 days for portfolio selection of the best weight for day 31 and calculated the change in returns under different models, and the results are shown in Figure 2.

As shown in Figure 2, in terms of the final cumulative returns of the six models, the DRER-W and DROR-MW models lead by a substantial margin. The second-tier league includes the DROR-MV, OR

https://doi.org/10.1017/S0269964823000141 Published online by Cambridge University Press

Table 5. *Means, variances and correlation coefficients of stock returns.*

Stock	Mean	Variance	Correlation coefficient									
X_1	0.00103	0.00026	1	0.36	0.35	0.46	0.29	0.45	0.26	0.39	0.42	0.23
X_2	0.00053	0.00042	0.36	1	0.48	0.39	0.35	0.43	0.27	0.41	0.38	0.25
X_3	0.00026	0.00018	0.35	0.48	1	0.37	0.43	0.44	0.35	0.37	0.41	0.28
X_4	0.00069	0.00024	0.46	0.39	0.37	1	0.35	0.54	0.3	0.48	0.43	0.24
X_5	0.00037	0.00009	0.29	0.35	0.43	0.35	1	0.4	0.49	0.38	0.33	0.35
X_6	0.00075	0.00021	0.45	0.43	0.44	0.54	0.4	1	0.38	0.55	0.46	0.31
X_7	0.00036	0.00008	0.26	0.27	0.35	0.3	0.49	0.38	1	0.31	0.3	0.39
X_8	0.00103	0.00031	0.39	0.41	0.37	0.48	0.38	0.55	0.31	1	0.43	0.26
X_9	0.00077	0.00025	0.42	0.38	0.41	0.43	0.33	0.46	0.3	0.43	1	0.24
X_{10}	0.00037	0.00012	0.23	0.25	0.28	0.24	0.35	0.31	0.39	0.26	0.24	1

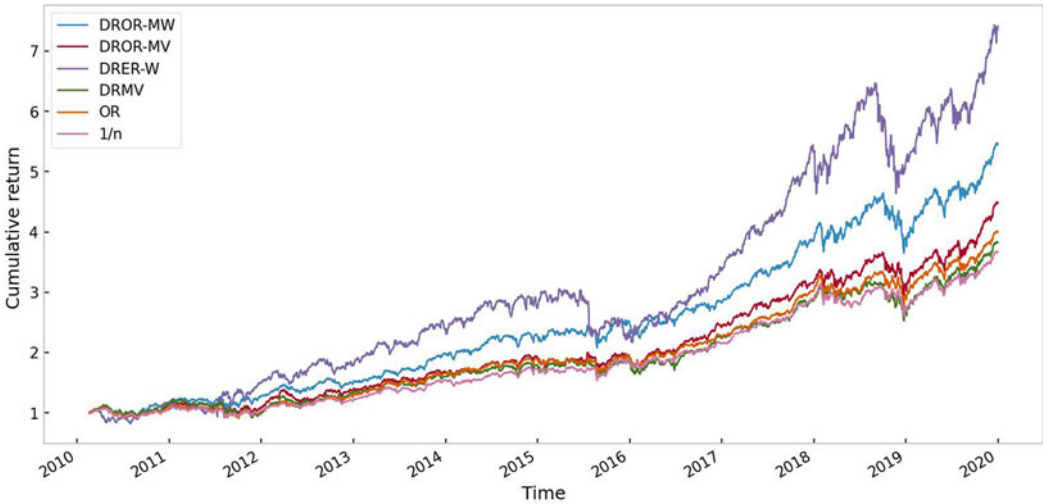

Figure 2. *Cumulative returns of optimal portfolio strategies under different models over the period 2010 to 2020.*

and DRMV models. The classical equal-weighting model lags behind. This is due to the DRER-W model, which aims to maximize portfolio returns and obviously outperforms the other models. In contrast, the DROR-MV and DROR-MW models maximize the worst-case Omega ratio, while the DRMV model aims to minimize variance, which is not monotonic with respect to returns.

In summary, the DROR-MW and DROR-MV models perform well in a bear market as they consider loss and robustness, controlling losses more than classical non-robust models during large price swings. At the same time, the DROR-MV and DROR-MW models have higher returns than the classical non-robust models in a robust long-term bull market and perform less conservatively.

4.2.3. The Influence of the Wasserstein Radius ε

The parameter ε represents the radius of the Wasserstein uncertainty set, and its value determines the size of a Wasserstein ball. In other words, as ε increases, the Wasserstein ball becomes larger.

To obtain a relatively suitable Wasserstein radius, we refer to relevant literature to choose the value of the radius for numerical experiments. Typically, we take $\varepsilon_0 = (\log N/N)^{1/d}$ as the radius. In the numerical experiments, we calculate the portfolio return results for the cases $\varepsilon = 0$, $\varepsilon_0/2$, $3\varepsilon_0/4$ and ε_0.

https://doi.org/10.1017/S0269964823000141 Published online by Cambridge University Press

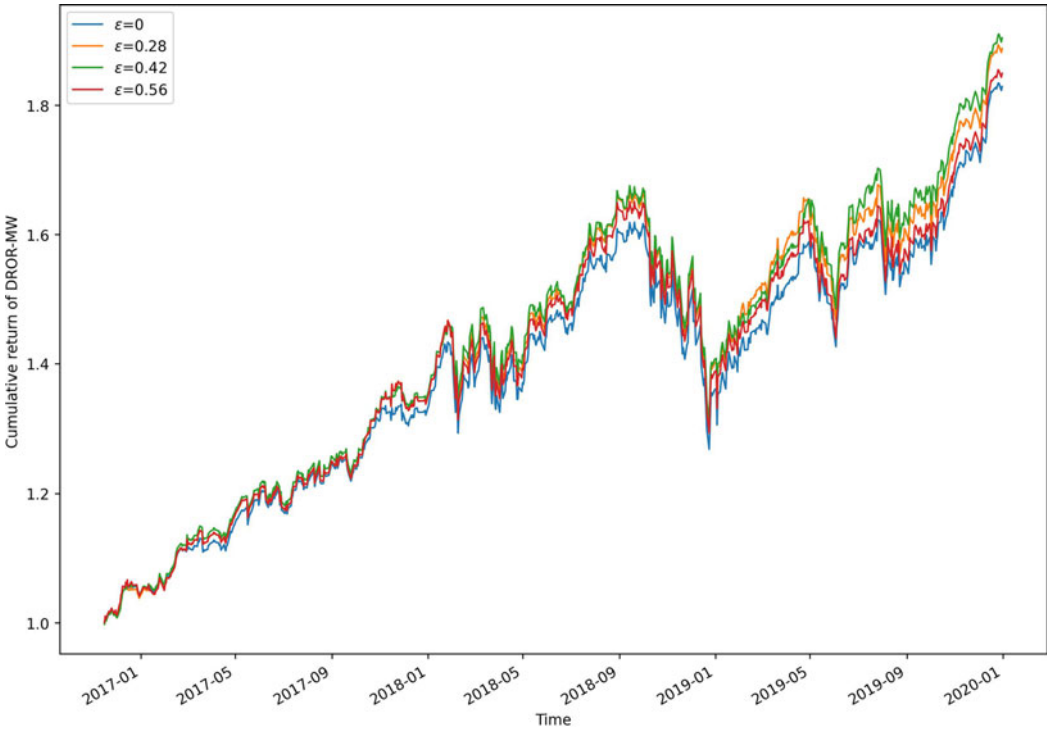

Figure 3. *Cumulative returns of optimal portfolio strategies under different ε.*

The calculation results show that small changes in ε can have a large impact on the solution of the model. Thus, the optimal solution for the investment weights is highly sensitive to the parameter ε. The calculation results of the portfolio strategy's returns for different values of ε are presented in Figure 3. The results show that the optimal portfolio allocation varies based on the radius ε of the Wasserstein uncertainty set, and we obtain a relatively diversified portfolio strategy.

Acknowledgments. The authors express their gratitude to the Editor, an anonymous referee, Tiantian Mao, and Qinyu Wu for providing insightful comments on an early version of the paper. Xinqiao Xie is the corresponding author.

Competing interests. The authors declare no conflict of interest.

References

[1] Bertrand, P. & Prigent, J.L. (2012). Omega performance measure and portfolio insurance. *The Journal of Banking and Finance* **35**(7): 1811–1823.

[2] Bertsimas, D., Doan, X.V., Natarajan, K., & Teo, C.P. (2010). Models for minimax stochastic linear optimization problems with risk aversion. *Mathematics of Operations Research* **35**(3): 580–602.

[3] Blanchet, J., Chen, L., & Zhou, X.Y. (2022). Distributionally robust mean-variance portfolio selection with Wasserstein distances. *Management Science* **68**(9): 6382–6410.

[4] Calafiore, G.C. & Ghaoui, L.E. (2006). On distributionally robust chance-constrained linear programs. *Journal of Optimization Theory and Applications* **130**(1): 1–22.

[5] Chen, L., He, S., & Zhang, S. (2011). Tight bounds for some risk measures with application to robust portfolio selection. *Operational Research* **59**(4): 847–865.

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

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

[8] Gao, R. & Kleywegt, A.J. (2016). Distributionally robust stochastic optimization with Wasserstein distance. Preprint arXiv:1604.02199.

https://doi.org/10.1017/S0269964823000141 Published online by Cambridge University Press

[9] Hu, W., Niu, G., Sato, I., & Sugiyama, M.. (2018). Does distributionally robust supervised learning give robust classifiers?. In *Proceedings of the 35th International Conference on Machine Learning*: Stockholm, Sweden. pp. 2029–2037.
[10] Jagannathan, R. (1977). Minimax procedure for a class of linear programs under uncertainty. *Operations Research* **25**(1): 173–177.
[11] Kang, Z., Li, X., Li, Z., & Zhu, S. (2019). Data-driven robust mean-CVaR portfolio selection under distribution ambiguity. *Quantitative Finance* **19**(1): 105–121.
[12] Kang, Z., Zhao, L., & Sun, J. (2019). The optimal portfolio of α-maxmin mean-VaR problem for investors. *Physica A: Statistical Mechanics and its Applications* 526: 120778.
[13] Markowitz, H. (1952). Portfolio selection. *The Journal of Finance* **7**(1): 77–91.
[14] Mao, T., Wang, R., & Wu, Q. (2022). Model aggregation for risk evaluation and robust optimization. Preprint arXiv:2201.06370.
[15] Natarajan, K., Sim, M., & Uichanco, J. (2010). Tractable robust expected utility and risk models for portfolio optimization. *Mathematical Finance* **20**(4): 695–731.
[16] Neufeld, A., Sester, J., & Šikić, M. (2022). Markov decision processes under model uncertainty. Preprint arXiv:2206.06109.
[17] Popescu, I. (2007). Robust mean-covariance solutions for stochastic optimization. *Operations Research* **55**(1): 98–112.
[18] Shadwick, W.F. & Keating, C. (2002). A universal performance measure. *Journal of Performance Measurement* **6**(3): 59–84.
[19] Sharpe, W.F. (1994). Sharpe ratio. *The Journal of Portfolio Management* **21**(1): 49–58.
[20] Villani, C. (2008). *Optimal Transport: Old and New*, Vol. 338. Berlin: Springer Science and Media.
[21] Wiesemann, W., Kuhn, D., & Sim, M. (2014). Distributionally robust convex optimization. *Operations Research* **62**(6): 1203–1466.
[22] Zhu, Z., Zhang, J., & Ye, Y. (2013). Newsvendor optimization with limited distribution information. *Optimization Methods and Software* **28**(3): 640–667.

Cite this article: Li Q. and Xie X. (2024). Worst-case Omega ratio under distribution uncertainty with its application in robust portfolio selection. *Probability in the Engineering and Informational Sciences* 38(2): 318–340. https://doi.org/10.1017/S0269964823000141