Distributionally robust mean-absolute deviation portfolio optimization using wasserstein metric

Dali Chen · Yuwei Wu · Jingquan Li · Xiaohui Ding · Caihua Chen

Received: 29 August 2021 / Accepted: 19 April 2022 / Published online: 16 May 2022
© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2022

Abstract
Data uncertainty has a great impact on portfolio selection. Based on the popular mean-absolute deviation (MAD) model, we investigate how to make robust portfolio decisions. In this paper, a novel Wasserstein metric-based data-driven distributionally robust mean-absolute deviation (DR-MAD) model is proposed. However, the proposed model is non-convex with an infinite-dimensional inner problem. To solve this model, we prove that it can be transformed into two simple finite-dimensional linear programs. Consequently, the problem can be solved as easily as solving the classic MAD model. Furthermore, the proposed DR-MAD model is compared with the 1/N, classic MAD and mean-variance model on S&P 500 constituent stocks in six different settings. The experimental results show that the portfolios constructed by DR-MAD model are superior to the benchmarks in terms of profitability and stability in most fluctuating markets. This result suggests that Wasserstein distributionally robust optimization framework is an effective approach to address data uncertainty in portfolio optimization.

Keywords Uncertainty modelling · Wasserstein distributionally robust optimization · MAD portfolio model · Nonconvex optimization

Mathematics Subject Classification 90C90 · 90C15

Yuwei Wu
wyw@nus.edu.sg

Jingquan Li
ljq@nju.edu.cn

Dali Chen
dalichen@smail.nju.edu.cn

Xiaohui Ding
dingxh@smail.nju.edu.cn

Caihua Chen
chchen@nju.edu.cn

1 School of Management and Engineering, Nanjing University, Nanjing 210093, China
2 Department of Mathematics, National University of Singapore, Singapore 117576, Singapore
1 Introduction

Modern portfolio theory originated from Markowitz’s seminal work on mathematical analysis of portfolio selection [35]. A major contribution of Markowitz’s work is the introduction of a trade-off between risk and expected return, which results in the mean-variance model seeking minimum variance portfolio with a given level of return. Since then, the mean-variance model attracts significant attentions from multiple fields, such as stock investment [13, 17, 40], supply chain management [8–10], and energy economics [43, 44].

Despite significant theoretical contributions of Markowitz’s theory, the analysis is built on strong market conditions. One fundamental assumption is that investor’s decision is determined by mean and variance of the portfolio return. This assumption holds under the frequently used condition that the portfolio return follows a multivariate Gaussian distribution [19]. However, [11] finds that normality for investment returns appears contradict with empirical evidence. Besides, even with the normal distribution assumption, the real distribution parameters are unknown. A tactical way is to resort these parameters from empirical data, but it might be highly deviated, especially under poor quality datasets [4]. Moreover, the solution to mean-variance model is believed to be sensitive to the mean and covariance matrix [6, 23], which offers a poor out-of-sample performance due to the estimation error of parameters. Last but not least, for single-period investment, the Markowitz model indicates the in-sample and out-of-sample data follow the same distribution, which is rare in real situation. Generally, these issues can be attributed to the uncertainty in the true distribution, which makes the application of Markowitz model problematic in the real-world financial market.

To compensate the drawbacks of Markowitz model under uncertain environments, recent studies motivate us to consider distributionally robust optimization (DRO) method. DRO is a popular technique used to deal with problems with uncertain distributions. Compared to the Markowitz model considering a single distribution, DRO constructs a distribution uncertainty set and optimizes the model with respect to all inner distributions. Therefore, DRO is well accepted as a method that resists against distribution uncertainty since it usually allows both the distribution and their parameters’ perturbation in certain scale. As a result, DRO can achieve a reasonable performance in practical even if the amount of data is limited [36]. These properties inspire us to apply the idea of DRO to deal with data uncertainties.

In this paper, we consider the combination of DRO with mean-absolute deviation (MAD) portfolio model. MAD model is first proposed in portfolio selection by [29] and uses mean-absolute deviation as risk measure. Compared with Markowitz model, it is more consistent with Von Neumann’s principle of “maximization of expected utility (MEU)” [28, 39] and computationally easier. In this regard, this paper considers MAD model under DRO framework for portfolio selection.

To better understand the proposed distributionally robust MAD (DR-MAD) model, we review the related literature below.

1.1 Related literature

In this subsection, we compare MAD model with mean-variance model in portfolio selection, and discuss major studies on DRO methods.

1.1.1 Comparison of MAD and mean-variance model

As an alternative to mean-variance criteria, MAD can also be used as the risk measure. For the early application since 1990s in portfolio selection, MAD model attracts much attention.
as it preserves many positive properties of mean-variance model and sometimes provides advantages. To illustrate, both models produce the same results under the assumption of joint normality for asset returns [29, 52]. Moreover, MAD model is demonstrated to be consistent with the second degree stochastic dominance (SSD) when considering a bounded set of mean-risk trade-off [38, 52]. By contrast, the mean-variance model is not SSD consistent in most cases [33]. More importantly, MAD model has a simpler linear program (LP) formulation and does not require an estimation of covariance matrix for asset returns. The simplicity makes it more attractive to deal with dataset without good quality, which motivates a lot of research to consider absolute deviation as the risk measure [5, 34, 41].

As MAD model simply replaces the risk measure with absolute deviation, it also inherits some drawbacks of Markowitz model mentioned above. For this reason, some recent studies try to address the limitations in MAD framework. In this effort, [30] considers the third moment of asset returns, in addition to mean and absolute deviation, to cope with skewed distributions that explain empirical evidence better. Besides, [53] and [52] develop multi-period models to characterize the dynamic properties of the financial market. Moreover, [37] applies robust optimization (RO) techniques to control the impact of uncertain parameters on model performance. Overall, each study mentioned above only considers parts of the limitations on applying MAD model for portfolio selection. To further improve its practical performance, we intend to investigate a more general approach to model data uncertainty.

1.1.2 Distributionally robust optimization

As a systematic way to model data uncertainty, DRO is designed to make robust decisions by optimizing the worst-case performance over the ambiguity set constructed from distributional information of uncertain data. The idea of DRO is first introduced by [45] in the seminal work on risk-averse newsvendor problem. The modeling technique is increasingly popular because of the great advances in mathematical understanding of DRO problems in the last decade [36, 50]. Compared with classic RO method, e.g., [37], DRO is often thought to be less conservative and it can incorporate distributional information to utilize all available data better [42]. Indeed, the ambiguity set is designed to include unknown data-generating distribution as much as possible, whereas the true distribution can be highly skewed or fat-tail. Meanwhile, DRO allows uncertain distribution to be varied within the ambiguity set, which also reflects the dynamic features of real-world problem. In this sense, we consider DRO as a more general approach to handle data uncertainty in MAD model.

To combine DRO with MAD model, there are two major approaches to build ambiguity sets in the literature. In the first case, the ambiguity set contains distributions satisfying some moment constraints, thus this kind of method is called moment-based DRO [14, 21, 50]. In the other case of metric-based DRO, the ambiguity set is defined as a ball in the space of probability distributions centered at a nominal distribution with the radius measured by distance functions such as the Kullback–Leibler divergence [24], the Wasserstein metric [51], or the Prohorov metric [18]. Compared with moment-based DRO, metric-based method builds ambiguity set considering distance to the nominal distribution, which includes all moment information implicitly instead of just a few moments. Therefore, it is intuitive and more flexible to control model’s conservativeness [36, 55]. Among different distances in metric-based DRO, Wasserstein metric obeys axioms of distance better and often enjoys tractable reformulation under mild conditions. Moreover, Wasserstein metric-based DRO (WDRO) can usually offer better out-of-sample performance guarantees [31]. For these reasons, we consider to combine MAD model with Wasserstein-based DRO. This work is close in spirit to the earlier paper of Mohajerin Esfahani and Kuhn [36], which provides detailed discussions.
on the tractability of DRO models in many different settings. We note our problem is mostly related to the DRO problem with piecewise affine loss functions considered in [36]. However, in this paper we consider expected return as the central tendency. As the expected return also depends on the uncertain parameters, the objective function is not simply a piecewise affine loss function. As a result, we cannot apply the Corollary 5.1 in [36] to simplify the problem. This motivates us to develop new techniques to solve the problem.

1.2 Contributions and overview

In this paper, we study portfolio selection by combining MAD model with Wasserstein-based DRO. To the best of our knowledge, there is no prior literature in this combination. As we explained before, this problem is difficult because the expected return also depends on the unknown data distribution. To simplify the proposed model, several transformations can be applied to it as shown in Fig. 1. First, to linearize the inner problem with respect to data distribution, we treat the expected portfolio return as a decision variable to make the problem a min-max-max formula. As the new formula is three-stage with infinite-dimensional inner problem, we apply dual theory twice to equate it to a min-max finite-dimensional problem. Although the problem is greatly simplified, it remains non-convex and is difficult to solve. For this issue, we observe an optimal closed-form solution exists for the inner problem when considering infinity norm in Wasserstein metric. As a result, DR-MAD model is transformed into two LPs. It is noteworthy that the resulting model is no more difficult to solve, but more robust to defend financial risks compared with the classic MAD model.

To check the competence of the proposed method, we compare the performance of DR-MAD model with 1/N (equal amounts of every stock), classic MAD and mean-variance model on S&P 500 component stock data. In our experiments, the performance of all methods is evaluated using a rolling horizon procedure similar to [15]. Firstly, we consider weekly return of S&P 500 stocks and tested the difference between Sharpe ratios of DR-MAD model and benchmarks using inference method introduced by [32]. The test result shows that our method is superior to benchmarks in most cases, and the differences are statistically significant. Considering a weekly rebalancing strategy, the maximum drawdown rate (MDR) of DR-MAD model is consistently lower than benchmarks, which shed some light on the robustness of our model. To further validate the models’ robustness, we select five daily datasets in periods of fluctuating markets due to the financial risk, trade frictions, or pandemic, etc. In this experiment, DR-MAD model also performs better in terms of final value of portfolio and MDR in most cases. To conclude, DR-MAD model achieves an attractive balance between risk and return.

Our work makes three main contributions: (i) To cope with the real-world data uncertainty, we presented the first distributionally robust MAD portfolio model using Wasserstein metric. (ii) Although the DR-MAD model is a non-convex infinite-dimensional problem, we introduced a tractable reformulation to make it equivalent to two simple finite LPs under mild assumptions. As a result, the problem is easily solvable as the classic MAD model. (iii)
Compared with 1/N, classic MAD, and mean-variance model, DR-MAD model achieved superior performance on S&P 500 component stock data in terms of portfolio return and stability in most instances.

The rest of the paper is organized as follows. In Sect. 2, we review the classic MAD model and extend it to DR-MAD model. In Sect. 3, we show the tractable reformulation of DR-MAD model. Section 4 is devoted to numerical experiments. The conclusion remarks are drawn in Sect. 5.

Notation 1 For \( x, y \in \mathbb{R}^m \), the inner product is denoted by \( \langle x, y \rangle := x^T y \). \( \| \cdot \| \) denotes an arbitrary vector norm on \( \mathbb{R}^m \). The dual norm is defined as \( \| y \|_* := \sup_{\| x \| \leq 1} \langle x, y \rangle \). \( \mathcal{M}(\Xi) \) denotes the set of all probability distributions supported on \( \Xi \), and \( \mathcal{M}(\Xi_1) \times \mathcal{M}(\Xi_2) \) denotes the set of all probability distributions supported on \( \Xi_1 \times \Xi_2 \). The \( e \) is a vector of all ones and has a matching dimension. In addition, all vectors in this paper are columns for the sake of unity.

2 Model formulation

In this section, we review the MAD model proposed by [29], and establish the mathematical formulation of the DR-MAD model.

2.1 Mean-absolute deviation model

Suppose there are \( m \) assets in the financial market, the random vector \( \xi = (\xi_1, \ldots, \xi_m)^T \in \Xi \) denotes the assets return in a certain period. In this paper, we consider \( \Xi = \mathbb{R}^m \) as the domain of \( \xi \). Given the portfolio weights \( x = (x_1, \ldots, x_m)^T \) which satisfy \( \sum_{i=1}^m x_i = 1 \), the portfolio return can be calculated as \( \xi^T x \). The MAD model aims to minimize the average deviation of portfolio return and requires the expected return to be above a target value. An optimal trade-off between risk and return can be obtained by solving the following problem:

\[
\text{(MAD)}: \min_x \mathbb{E}_Q|\xi^T x - \mathbb{E}_Q(\xi^T x)|,
\]

s.t. \( \mathbb{E}_Q(\xi^T x) \geq \rho, \ x \in \mathcal{X} \),

where \( \rho \) denotes the target portfolio return, and \( Q \) denotes the true distribution of the assets return \( \xi \). Moreover, \( \mathcal{X} \subseteq \mathbb{R}^m \) denotes the extra constraints on portfolio weights. In this paper, we focus on portfolio selection without short sales and hence \( \mathcal{X} = \{x : x \geq 0, e^T x = 1\} \). It is noteworthy that \( \mathcal{X} \) can also include other constraints to meet the investors’ requirements, such as cardinality bound, diversification and centralization constraints.

As the real distribution of portfolio return \( Q \) is usually unknown in real-world markets, solving problem (1) is impossible in practice. To deal with this difficulty, empirical distribution built from financial data is used to approximate the true distribution. Suppose \( \hat{\xi} = \{\hat{\xi}_1, \ldots, \hat{\xi}_N\} \) is a set of independent observations of \( \xi \), then the empirical distribution of assets return can be written as:

\[
\mathbb{P}_N = \frac{1}{N} \sum_{i=1}^N \delta_{\hat{\xi}_i},
\]
where $\delta_{\hat{\xi}_i}$ denotes the Dirac point measure at $\hat{\xi}_i$. Substituting $Q$ by $P_N$ in problem (1), the MAD model can be reformulated as

$$(\text{classic MAD}) : \min_{x} \mathbb{E}_{P_N} |\xi^T x - \mathbb{E}_{P_N} (\xi^T x)|,$$

s.t. $\mathbb{E}_{P_N} (\xi^T x) \geq \rho, \ x \in \mathcal{X},$ \hspace{1cm} (2)

which is a sample average approximation to model (1). Moreover, it can be easily transformed into an LP if $\mathcal{X}$ contains only linear constraints.

### 2.2 DR-MAD model

Sample average approximation may cause several problems in practice. Firstly, if the data size is small, the approximation might be biased [36]. This often happens when using low-frequency financial data [15]. Secondly, the aforementioned data uncertainties in mean-variance model still remain in MAD model [47]. These drawbacks make using classic MAD model in real-word financial market data problematic.

To address these problems, we consider the data-driven WDRO method proposed by [36]. Although the true distribution of the investment return $Q$ in (1) can not be identified exactly, it is reasonable to believe that $Q$ may not be “far” from the empirical distribution $P_N$. Based on this assumption, we define the ambiguity set as a distribution-based ball centered at $P_N$ using Wasserstein distance as the radius measure. In particular, the definition of Wasserstein distance is given below:

**Definition 1** Wasserstein metric [27]: $d_W : \mathcal{M}(\Xi) \times \mathcal{M}(\Xi) \to \mathbb{R}_+$ is defined as

$$d_W(P_1, P_2) = \inf_{\Pi \in \mathcal{M}(\Xi^2)} \left\{ \int_{\Xi^2} \|\xi_1 - \xi_2\| \Pi(d\xi_1, d\xi_2) : \Pi(\Xi, d\xi_2) = P_1(d\xi_1), \right\},$$

where $\Pi \in \mathcal{M}(\Xi^2)$ is the joint distribution of $P_1 \in \mathcal{M}(\Xi)$ and $P_2 \in \mathcal{M}(\Xi)$.

In this definition, $d_W$ does not guarantee a real distance. The following assumption is made to prevent $d_W$ from being infinite on set $\mathcal{M}(\Xi)$ [2].

**Assumption 1** For any distribution $P \in \mathcal{M}(\Xi)$, the following constraint holds

$$\int_{\Xi} \|\xi\| P(d\xi) < \infty.$$

Assumption 1 only sacrifices little modeling power as has been stated in [48]. In this respect, we define the ambiguity set as the Wasserstein ball:

$$\mathbb{B}_\varepsilon(P_N) = \{P \in \mathcal{M}(\Xi) : d_W(P, P_N) \leq \varepsilon\}.$$

This ambiguity set contains all distributions within the $\varepsilon$-Wasserstein distance from the empirical distribution $P_N$. Indeed, the Wasserstein ball contains the true distribution $Q$ with a high probability under mild conditions [36]. Therefore, we introduce the Wasserstein distributionally robust MAD model as follows

$$(\text{DR-MAD}) : \min_{x} \max_{P} \mathbb{E}_{P} |\xi^T x - \mathbb{E}_{P} (\xi^T x)|,$$

s.t. $\mathbb{E}_{P} (\xi^T x) \geq \rho, \ P \in \mathbb{B}_\varepsilon(P_N), \ x \in \mathcal{X},$ \hspace{1cm} (3)
As demonstrated in [36], WDRO offers powerful out-of-sample performance guarantees. Therefore, DR-MAD model might be more capable at coping with data uncertainty in portfolio selection. However, the solution to problem (3) is non-trivial. To see this, we note the inner-most expectation with respect to $\mathbb{P}$ renders the model an infinite-dimensional non-convex optimization problem. For this reason, our problem is distinct from the piecewise affine problem considered in [36]. Close in spirit to [36], we show that the problem can be transformed into a finite convex program problem, which allows the DR-MAD model to be solved efficiently. We discuss several mathematical transformations in the next section.

3 Tractable reformulation of DR-MAD model

In this section, we provide a tractable reformulation for the DR-MAD model. Since problem (3) is non-convex, we fix the expected portfolio value $E_{\mathbb{P}}(\xi^T x) = \alpha$. Then, problem (3) is equivalent to the following min-max-max formulation

$$\min_{x} \max_{\alpha \geq \rho} \max_{\mathbb{P}} E_{\mathbb{P}}|\xi^T x - \alpha|, \quad \text{s.t. } E_{\mathbb{P}}(\xi^T x) = \alpha, \quad \mathbb{P} \in B_\epsilon(\mathbb{P}_N), \quad x \in \mathcal{X}. \tag{4}$$

Clearly, (4) is a three-stage program and the inner problem is now linear with respect to distribution $\mathbb{P}$, i.e.,

$$\max_{\mathbb{P}} E_{\mathbb{P}}|\xi^T x - \alpha|, \quad \text{s.t. } E_{\mathbb{P}}(\xi^T x) = \alpha, \quad \mathbb{P} \in B_\epsilon(\mathbb{P}_N). \tag{5}$$

Note that $\alpha$ and $x$ are regarded as two constants in (5), and the decision variable $\mathbb{P}$ is infinite-dimensional. To simplify this problem, (5) can be transformed into a conic linear program. Applying the dual theory as in [36] results in the following proposition.

**Proposition 1** Considering the uncertain set $\Xi = \mathbb{R}^m$, for any $\varepsilon \geq 0$, if the classic MAD model (2) is feasible, then the optimal value of (5) equals the optimal value of the following program

$$\min_{\gamma_1, \gamma_2} \frac{1}{N} \sum_{i=1}^{N} |(\hat{\xi}_i)^T x - \alpha| + \gamma_1(\hat{\mu}^T x - \alpha) + \gamma_2 \varepsilon, \tag{6}$$

$$\text{s.t. } \|((1 - \gamma_1)x)_*\| \leq \gamma_2, \quad \|((1 + \gamma_1)x)_*\| \leq \gamma_2,$$

where $\hat{\mu} = \frac{1}{N} \sum_{i=1}^{N} \hat{\xi}_i$.

According to Proposition 1, the inner problem (5) is simplified as a two-dimension convex program. Plugging the expression (6) into (4) still yields a three-stage formulation. Next, we take the dual of (6) again so that the DR-MAD model can be reduced to a two-stage program in the following proposition:
Proposition 2 For any $\varepsilon \geq 0$, the optimal value of (3) equals to the optimal value of the following problem

$$\begin{align*}
\min_{x} \max_{\alpha, \lambda_1, \lambda_2, \varepsilon, v_1, v_2} & \quad \frac{1}{N} \sum_{i=1}^{N} |(\hat{\varepsilon}_i)^T x - \alpha| + (v_1 + v_2)^T x, \\
\text{s.t.} & \quad \hat{\mu}^T x - v_1^T x + v_2^T x = \alpha, \\
& \lambda_1 + \lambda_2 \leq \varepsilon, \\
& \|v_1\| \leq \lambda_1, \\
& \|v_2\| \leq \lambda_2, \\
& \alpha \geq \rho, \ x \notin X.
\end{align*}$$

Note that (7) is a two-stage finite-dimensional problem, but still non-trivial since the inner maximization problem is non-concave with respect to $\alpha$. Fortunately, we are able to simplify (7) by considering infinity norm $\| \cdot \|_{\infty}$ in the Wasserstein metric, as a closed-form solution can be identified for the inner problem with fixed $\alpha$. This result is stated in the next proposition.

Proposition 3 Suppose $\| \cdot \| = \| \cdot \|_{\infty}$ and $x \in X$ hold. If $\hat{\mu}^T x - \alpha \in [-\varepsilon, \varepsilon]$, the following problem is feasible:

$$\begin{align*}
\max_{\lambda_1, \lambda_2, \varepsilon, v_1, v_2} & \quad (v_1 + v_2)^T x, \\
\text{s.t.} & \quad \hat{\mu}^T x - v_1^T x + v_2^T x = \alpha, \\
& \lambda_1 + \lambda_2 \leq \varepsilon, \\
& \|v_1\|_{\infty} \leq \lambda_1, \\
& \|v_2\|_{\infty} \leq \lambda_2.
\end{align*}$$

Moreover, $\lambda_1^* = \frac{\varepsilon + (\hat{\mu}^T x - \alpha)}{2}, \lambda_2^* = \frac{\varepsilon - (\hat{\mu}^T x - \alpha)}{2}, v_1^* = \lambda_1^* \varepsilon, v_2^* = \lambda_2^* \varepsilon$ is an optimal solution to (8), while the optimal value is $\varepsilon$. If $\hat{\mu}^T x - \alpha \notin [-\varepsilon, \varepsilon]$, the problem is infeasible.

Indeed, infinity norm is a common choice of distance measure in DRO and portfolio selection [20, 25, 52]. Moreover, infinity norm generally induces a larger feasible region. Therefore, problem (8) provides a point-wise upper bound to the inner problem in (7) with $p$ norm for $p \in (0, +\infty)$. When restricting to infinity norm, Proposition 3 gives us the feasible region, optimal value and an optimal solution to problem (8). Using these results, the inner problem of (7) reduces to a one-dimensional convex program with respect to $\alpha$. Now, we are able to reformulate the DR-MAD model as two finite-dimensional LPs. The following theorem describes the resulting LPs.

Theorem 1 Suppose the uncertain set is $\Xi = \mathbb{R}^m$, and infinity norm $\| \cdot \|_{\infty}$ is considered in Wasserstein metric. For any $\varepsilon \geq 0$, problem (2) is equivalent to the following program

$$\begin{align*}
\min_{x \in \mathcal{X}} \left\{ \max \left( \frac{1}{N} \sum_{i=1}^{N} |\hat{\mu}^T x - (\hat{\varepsilon}_i)^T x - \varepsilon| + \varepsilon, \frac{1}{N} \sum_{i=1}^{N} |\hat{\mu}^T x - (\hat{\varepsilon}_i)^T x + \varepsilon| + \varepsilon \right), \right. \\
\left. \text{if } \hat{\mu}^T x - \varepsilon \geq \rho, \right. \\
\left. \min \left( \frac{1}{N} \sum_{i=1}^{N} |\rho - (\hat{\varepsilon}_i)^T x| + \varepsilon, \frac{1}{N} \sum_{i=1}^{N} |\hat{\mu}^T x - (\hat{\varepsilon}_i)^T x + \varepsilon| + \varepsilon \right), \right. \\
\left. \text{if } \hat{\mu}^T x - \varepsilon \leq \rho \leq \hat{\mu}^T x + \varepsilon, \right. \\
\left. +\infty, \text{ if } \hat{\mu}^T x + \varepsilon \leq \rho. \right. \tag{9}
\end{align*}$$

\textcopyright Springer
Moreover, the optimal value of problem (9) equals to the minimum of optimal values of the following two LPs. The first LP is

\[
\begin{align*}
    \min_{t, x, y_1, y_2} & \quad t, \\
    \text{s.t.} & \quad t \geq \frac{1}{N} \sum_{i=1}^{N} y_{1,i} + \varepsilon, \\
    & \quad t \geq \frac{1}{N} \sum_{i=1}^{N} y_{2,i} + \varepsilon, \\
    & \quad y_{1,i} \geq \mu^T x - (\hat{\xi}_i)^T x - \varepsilon, \quad i = 1, \ldots, N, \\
    & \quad y_{1,i} \geq -\mu^T x + (\hat{\xi}_i)^T x + \varepsilon, \quad i = 1, \ldots, N, \\
    & \quad y_{2,i} \geq \mu^T x - (\hat{\xi}_i)^T x + \varepsilon, \quad i = 1, \ldots, N, \\
    & \quad y_{2,i} \geq -\mu^T x + (\hat{\xi}_i)^T x - \varepsilon, \quad i = 1, \ldots, N, \\
    & \quad \hat{\mu}^T x - \varepsilon \geq \rho, \\
    & \quad \hat{\mu}^T x + \varepsilon \geq \rho, \\
    & \quad x \in \mathcal{X},
\end{align*}
\]

and the second LP is

\[
\begin{align*}
    \min_{t, x, y_1, y_2} & \quad t, \\
    \text{s.t.} & \quad t \geq \frac{1}{N} \sum_{i=1}^{N} y_{1,i} + \varepsilon, \\
    & \quad t \geq \frac{1}{N} \sum_{i=1}^{N} y_{2,i} + \varepsilon, \\
    & \quad y_{1,i} \geq (\hat{\xi}_i)^T x - \rho, \quad i = 1, \ldots, N, \\
    & \quad y_{1,i} \geq -(\hat{\xi}_i)^T x + \rho, \quad i = 1, \ldots, N, \\
    & \quad y_{2,i} \geq \mu^T x - (\hat{\xi}_i)^T x + \varepsilon, \quad i = 1, \ldots, N, \\
    & \quad y_{2,i} \geq -\mu^T x + (\hat{\xi}_i)^T x - \varepsilon, \quad i = 1, \ldots, N, \\
    & \quad \hat{\mu}^T x - \varepsilon \leq \rho, \\
    & \quad \hat{\mu}^T x + \varepsilon \geq \rho, \\
    & \quad x \in \mathcal{X}.
\end{align*}
\]

If we take the radius of Wasserstein ball \( \varepsilon \) to be 0, it is not difficult to verify that the first LP will achieve a smaller value, which is equivalent to the classic MAD model. Noticeably, the DR-MAD model absorbs distributional robustness as compared with the classic MAD model, but is also LP solvable according to Theorem 1.

**Remark 1** One popular way to deal with portfolio selection model is to penalize the expected return constraint to the objective. By mimicking our method, the combination of this kind of model with WDRO can be reduced to a single LP. The main reason that it has one less LP than (10) and (11) is the lack of expected return constraint.

### 4 Numerical experiments

In this section, we conduct several empirical experiments to compare the performance of the proposed method with classic portfolio selection models on real market datasets. These experiments focus on the profitability and robustness of different models in the out-of-sample period.
Table 1  List of datasets considered

| Datasets      | # Stocks | Time period            | Sample size | E. Window | P. Window |
|---------------|----------|------------------------|-------------|-----------|-----------|
| Week2012      | 465      | 12/28/2012–10/21/2016  | 200 weeks   | 150 weeks | 1 week    |
| Day2007       | 419      | 07/27/2007–07/23/2009  | 502 days    | 250 days  | 21 days   |
| Day2010       | 447      | 02/06/2010–05/25/2012  | 502 days    | 250 days  | 21 days   |
| Day2014       | 476      | 05/28/2014–05/23/2016  | 502 days    | 250 days  | 21 days   |
| Day2017       | 488      | 08/04/2017–05/03/2019  | 439 days    | 250 days  | 21 days   |
| Day2019       | 497      | 01/15/2019–07/13/2020  | 376 days    | 250 days  | 21 days   |

‘E. Window’ is the estimation window and ‘P. Window’ is the prediction window. ‘# Stocks’ is the number of valid stocks in each dataset for which the stocks with missing data are all removed.

4.1 Experimental setup

4.1.1 Description of models and parameters

Models compared with the proposed DR-MAD are 1/N, classic MAD and mean-variance model. 1/N strategy refers to distributing investors’ wealth uniformly to all available assets and the mean-variance model refers to the minimum variance portfolio strategy proposed by [35]. The experiments have been repeated on several datasets in order to achieve a convincing result. For each dataset, a testing period is specified to evaluate model performances. Similar to [16], a rolling window procedure is applied in which estimation windows and prediction windows are identified for choosing models’ parameters and making predictions, respectively. Before the description of dataset and experiments, we introduce the method to determine parameters in these models.

For DR-MAD model, there are two parameters need to be estimated: target return $\rho$ and radius of the WDRO-based uncertain ball $\varepsilon$. The target return is set to the average stock returns in the estimation window as stated in [26] and denoted as $\bar{\rho}$. We also consider $\rho = 0.5 \times \bar{\rho}$ and $\rho = 1.5 \times \bar{\rho}$ since real-world investors may have different targets. Moreover, we note the choice of $\varepsilon$ is important since a small $\varepsilon$ may result in a small ambiguity ball that doesn’t contain all possible distributions, while a large $\varepsilon$ may lead to over-conservative decisions. To estimate the radius of Wasserstein ball $\varepsilon$, we employ 5-fold cross-validation method as in [36]. Particularly, the $\varepsilon$ is chosen as the value that results in the highest average portfolio return. In our experiment, we first observe that when searching $\varepsilon$ in the interval [0.05, 0.20] with step length of 0.01 for daily data, the model produces an estimation with superior performance. Based on this observation, we decide that the logarithmic return of weekly data should be about 5 times that of the daily data, so the interval for weekly data is set to be [0.25, 1] with the step length of 0.05.

Classic MAD and mean-variance model use sample average method to estimate mean and covariance of assets, and the parameter $\rho$ in these two models is determined in the same way as in the DR-MAD model for consistency.

4.1.2 Description of datasets

Datasets used in this paper are stock price from S&P 500 market components which are downloaded from the Wind-Economic Database.1 The logarithmic return was calculated and

---

1 A commercial financial database, see https://www.wind.com.cn/.
used in the experiments. The description of the datasets is given in Table 1 and abbreviations of all datasets are listed in the second column.

To validate the robustness of DR-MAD model, datasets considered in this paper are specified in certain time periods, in which the financial market fluctuates significantly. The selected periods cover some major unforeseen adverse events, such as the financial crisis in 2008, trade frictions, and the COVID-19 pandemic. The stock number varies because those stocks with missing data are omitted during the time period considered. Note that the first dataset denoted as Week2012 consists of weekly stock returns, and other datasets consist of daily returns.

To apply the rolling window procedure in [16], we separate the stock price data with total size $T$ into an estimation window with size $E$ and testing period with size $T - E$. More specifically, Fig. 2 gives an example of the S&P 500 index from August 04th 2017 to May 03rd 2019, which is the same as the time period considered in dataset Day2017. The black line and grey line show the index data in the 1st estimation window and data in the testing period respectively. In the 1st step, the optimal portfolio strategy is obtained using data from the 1st estimation window. Then the strategy is applied on the prediction window with length $P$. In the 2nd step, the earliest $P$ observations are dropped, and newest $P$ observations are added to form the 2nd estimation window. Portfolio can be obtained similarly as in the 1st step. These processes are repeated for $H = \frac{T - E}{P}$ times till the prediction results cover the entire testing period.

However, real-world financial market data may fluctuate greatly among different time periods. Stocks that are relatively stable in the estimation window may be volatile in the testing period, making it difficult to decide a proper portfolio strategy. For example in Fig. 2, the black line is flatter than the grey one, indicating data from testing period has a higher variance. This happens frequently when investors are in panic moods. To better understand this phenomenon, we check the popular financial index VIX\footnote{VIX is the Chicago Board Options Exchange (CBOE) Market Volatility Index, a measure of the stock market’s expectation of volatility based on S&P 500 index options. The data is also downloaded from the Wind-Economic Database.} which is usually regarded as the barometer of investor fear [49]. Table 2 describes statistics of VIX index and the return of S&P 500 index for each dataset. As can be seen from the table, the VIX and the variance of return in the testing period are often higher than that in the estimating period, indicating the higher uncertainty in financial markets during testing periods. It is interesting to investigate
Table 2 Statistics of VIX Index and Return of S&P 500 Index in datasets

| Datasets | Estimating period | Testing period |
|----------|-------------------|----------------|
|          | Mean(10^{-2}) | Variance(10^{-2}) | VIX | Mean(10^{-2}) | Variance(10^{-2}) | VIX |
| Week2012 | 12.793 | 1.241 | 14.937 | 1.979 | 1.692 | 16.440 |
| Day2007  | -14.643 | 4.244 | 23.160 | -27.257 | 20.702 | 40.431 |
| Day2010  | 21.532 | 2.303 | 20.941 | -0.595 | 5.380 | 24.468 |
| Day2014  | 10.702 | 1.367 | 14.591 | -3.739 | 2.771 | 17.789 |
| Day2017  | 13.032 | 1.596 | 13.718 | 6.126 | 2.677 | 16.728 |
| Day2019  | 23.644 | 1.390 | 15.141 | 21.348 | 33.539 |

The column ‘Estimating Period’ and ‘Testing Period’ are the periods of first estimation window and test set for each dataset correspondingly. The column ‘Mean’ and ‘Variance’ are the average and variance of the annual return of S&P 500 index, and ‘VIX’ is the average VIX index for the corresponding dataset.

whether the tested methods perform well in this situation, as it is hard to believe the stock return follows the same distributions in the estimation and test periods.

On the other hand, the weekly dataset are often more stable than the daily dataset. Particularly, volatilities are prominent for datasets Day2007 and Day2019, which indicates a more unstable financial environment as compared to others. This helps us to check the stability of different portfolio strategies. In the following experiments, we use weekly rebalancing policy for Week2012 dataset and monthly rebalancing policy for datasets Day2007, Day2010, Day2014, Day2017, Day2019.

4.1.3 Evaluation criteria

To compare the performance of different models, evaluation metrics are defined in this subsection. Consider a dataset with m candidate stocks and the testing period consisted of H prediction windows, each prediction window has P samples. Let \( V_{t,k}^p \), \( t = 1, \ldots, H \), \( k = 1, \ldots, P \) denotes the cumulative portfolio return in prediction window \( t \) at time \( k \). Let \( \xi_{t,k} = (\xi_{t,k}^1, \ldots, \xi_{t,k}^m) \) denote the logarithmic returns in prediction window \( t \) at time \( k \) for all stocks, and \( x_t = (x_t^1, \ldots, x_t^m) \) denote the portfolio weights in prediction window \( t \). At the beginning of the first prediction window, we set investor’s initial wealth to 1 and the cumulative return for asset \( i \) at time \( k \) equals \( V_{1,k}^i = x_1^i e^{\sum_{j=1}^k \xi_{1,j}^i} \). For \( t = 2, \ldots, H \), the cumulative return for asset \( i \) time \( k \) accumulates \( V_{t,k}^i = V_{t-1,k}^p x_t^i e^{\sum_{j=1}^k \xi_{t,j}^i} \). Therefore, the cumulative portfolio return equals

\[
V_{t,k}^p = \sum_{i=1}^m V_{t,k}^i.
\]

The last portfolio value \( V_{H,P}^p \) describes the final result of investment, and it is used to measure the portfolio return. The maximum drawdown rate (MDR) representing the maximum peak-to-trough decline ratio of the portfolio in out-of-sample period is defined as

\[
MDR = \max_{1 \leq i \leq m} \max_{1 \leq j \leq n} \frac{U_i - U_j}{U_i},
\]

where \( U_{P(t-1)+k}^p = V_{t,k}^p \) and \( n = H \cdot P \). The black downward arrow in Fig. 2 gives a visual display of MDR for S&P 500 index on dataset Day2017. Considering investors often quit the
investment after a large drawdown as stated in [54], we use the MDR to measure portfolio risk.

In addition, investors also care about the investment return and risk in each period. Using the definition of the portfolio return $V^P_t = V^P_{t-1}$ in prediction window $t$ as mentioned in above, the logarithmic return in such period equals $R^p_1 = \log V^P_1$ and $R^p_t = \log V^P_t - \log V^P_{t-1}$, $t \geq 2$. The mean $\hat{\mu}^p$ and standard deviation $\hat{\sigma}^p$ of the portfolio return is defined by

$$
\hat{\mu}^p = \frac{1}{H} \sum_{t=1}^{H} R^p_t,
$$

$$
\hat{\sigma}^p = \sqrt{\frac{1}{H-1} \sum_{t=1}^{H} (R^p_t - \hat{\mu}^p)^2}.
$$

Let $R^e_t = R^p_t - R^f_t$ denotes the excessive portfolio return, in which $R^f_t$ denotes the risk-free ratio. Then, Sharpe ratio $\hat{SR}$ can be defined as

$$
\hat{SR} = \frac{\hat{\mu}^e}{\hat{\sigma}^e},
$$

where $\hat{\mu}^e = \frac{1}{H} \sum_{t=1}^{H} R^e_t$ and $\hat{\sigma}^e = \sqrt{\frac{1}{H-1} \sum_{t=1}^{H} (R^e_t - \hat{\mu}^e)^2}$, respectively. Note that the Sharpe ratio is used to measure the excessive return per unit risk for all models.

### 4.2 Experiments on weekly data

The performance of DR-MAD model, 1/N strategy, classic MAD model and mean-variance model are compared using the Week2012 dataset in this subsection. As displayed in Table 1, there are 200 weekly observations for 465 stocks in the dataset. Investors are assumed to rebalance their portfolio on a weekly basis with the initial investment wealth equals 1. Rolling window procedure is applied with the estimation window of 150 weeks ($E = 150$, about 3 years) and prediction window of 1 week ($P = 1$). The remaining 50 weeks data ($H = 50$, about a year) from November 13, 2015 to October 21, 2016 is chosen as the testing period.

Table 3 reports the comparison of four models under different target returns. The out-of-sample performance is reported for the cumulative return in the final state (FinVal), MDR, mean return (Mean), standard deviation (Std) and Sharpe ratios ($\hat{SR}$) as defined in the previous subsection. The best performance is highlighted for each setting. Considering model robustness, MDR and Std suggest that 1/N strategy has the worst stability compared to other models. On the contrary, DR-MAD seems to be more robust than mean-variance and classic MAD. Additionally, FinVal is used to compare the cumulative return and DR-MAD also achieves the best performance. Note that classic MAD model also attains good results with target return $0.5 \times \hat{\rho}$ and $\bar{\rho}$, but it becomes much worse when target return reaches $1.5 \times \hat{\rho}$.

Moreover, we test the statistical differences between the Sharpe ratios of DR-MAD model and benchmarks using bootstrapping methods proposed by [32]. The one-side null hypothesis is defined as $H_0 : \hat{SR}_{dr} - \hat{SR}_{ben} \leq 0$, where $\hat{SR}_{dr}$ and $\hat{SR}_{ben}$ represent the Sharpe ratio of DR-MAD model and one of the benchmarks. The studentized circular block bootstrap is used for each inference, and the p-value is computed with 1000 resamples at the block size

---

3 The Treasure bills (three-month maturity) is assumed to be a proxy for risk-free ratio like [22], and the data is downloaded from the official website of the US.
Table 3  Statistical results of Portfolio values and returns during out-of-sample period

| \( \rho \) | Model            | FinVal | MDR(10\(^{-2}\)) | Mean(10\(^{-3}\)) | Std(10\(^{-2}\)) | \( \bar{SR} \) |
|---------|-----------------|--------|-----------------|-----------------|-----------------|-------------|
| 0.5 \( \times \bar{\rho} \) | DR-MAD          | 1.120  | 4.38            | 2.257           | 1.476           | 1.056       |
|         | 1/N             | 1.055  | 12.06           | 1.077           | 1.993           | 0.364 (0.0879)* |
|         | Mean-Variance   | 1.060  | 6.32            | 1.168           | 1.510           | 0.522 (0.0889)* |
|         | Classic MAD     | 1.075  | 5.31            | 1.439           | 1.525           | 0.643 (0.0929)* |
| 1.0 \( \times \bar{\rho} \) | DR-MAD          | 1.118  | 4.76            | 2.239           | 1.470           | 1.051       |
|         | 1/N             | 1.055  | 12.06           | 1.077           | 1.993           | 0.364 (0.0909)* |
|         | Mean-Variance   | 1.062  | 6.32            | 1.197           | 1.488           | 0.544 (0.0799)* |
|         | Classic MAD     | 1.070  | 5.44            | 1.346           | 1.508           | 0.607 (0.0800)* |
| 1.5 \( \times \bar{\rho} \) | DR-MAD          | 1.113  | 4.68            | 2.142           | 1.464           | 1.009       |
|         | 1/N             | 1.055  | 12.06           | 1.077           | 1.993           | 0.364 (0.0959)* |
|         | Mean-Variance   | 1.060  | 6.62            | 1.162           | 1.460           | 0.537 (0.0889)* |
|         | Classic MAD     | 1.045  | 5.82            | 0.873           | 1.517           | 0.382 (0.0360)** |

The \( p \)-value of inference for Sharpe Ratio is in brackets in the last column of the table; one asterisk near the \( p \)-value indicates the significance level of 10%, and two asterisks indicate 5%. The bold number refers to the best result.

Fig. 3  Evolution of cumulative wealth for different models during out-of-sample period: \( \rho = \bar{\rho} \)

As can be seen from Table 3, the Sharpe ratio of the DR-MAD model is higher than benchmarks in all cases, and the differences are statistically significant.

The cumulative return for target returns \( \bar{\rho} \) is plotted in Fig. 3 to check the performance of different models over the test periods. As can be seen from the figure, the cumulative returns of DR-MAD, classic MAD and mean-variance model coincide with each other in the first 25 weeks (about half an year). By contrast, 1/N strategy performs much more poorly. However, in the last 25 weeks, the cumulative wealth of DR-MAD model and 1/N strategy achieve more prominent growth than the other approaches. Finally, DR-MAD model performs the best on almost all weeks. Similar pattern can be found for different target returns in Figs. 4 and 5.
To conclude, DR-MAD model shows the best profitability and robustness in the experiment. Since the ambiguity set in the distributionally robust model considers more risky scenarios, the DR-MAD model is likely to give better performance especially in financial market with great uncertainties.

4.3 Experiments on daily data

In this section, we consider the performance of the DR-MAD model on the daily data. Note that both weekly data and daily data are commonly used in portfolio optimization, e.g., [1, 7, 12]. Compared to the weekly data, daily data keeps all the daily fluctuations, which may be more informative but also noisy. For this reason, we test the robustness of the proposed model using daily data from the 5 most fluctuating periods. The data collection periods can be found...
Table 4  Results of Portfolio values and MDR during out-of-sample period

| ρ   | Model      | Day2007 | Day2010 | Day2014 | Day2017 | Day2019 |
|-----|------------|---------|---------|---------|---------|---------|
|     |            | FinVal  | MDR(%)  | FinVal  | MDR(%)  | FinVal  | MDR(%)  | FinVal  | MDR(%)  |
| 0.5 | DR-MAD     | 0.9168  | 27.699  | 1.0610  | 9.988   | 1.0491  | 8.690   | 1.0391  | 13.143  | 0.9654  | 27.881  |
|     | 1/N        | 0.8699  | 48.321  | 1.0359  | 21.451  | 1.0099  | 14.332  | 1.0509  | 19.867  | 0.8916  | 39.032  |
|     | Mean-variance | 0.8708  | 32.449  | 1.0575  | 10.299  | 1.0056  | 9.065   | 1.0489  | 10.782  | 0.9323  | 29.819  |
|     | Classic MAD | 0.9051  | 28.943  | 1.0616  | 10.549  | 1.0538  | 8.746   | 1.0253  | 12.862  | 0.9518  | 28.050  |
| 1.0 | DR-MAD     | 0.9153  | 28.243  | 1.0629  | 10.392  | 1.0475  | 8.745   | 1.0380  | 13.179  | 0.9572  | 27.550  |
|     | 1/N        | 0.8699  | 48.321  | 1.0359  | 21.451  | 1.0099  | 14.332  | 1.0509  | 19.867  | 0.8916  | 39.032  |
|     | Mean-variance | 0.8908  | 32.449  | 1.0535  | 11.321  | 1.0520  | 9.065   | 1.0488  | 10.789  | 0.9335  | 29.810  |
|     | Classic MAD | 0.9051  | 28.943  | 1.0563  | 11.559  | 1.0505  | 8.752   | 1.0233  | 13.863  | 0.9528  | 28.050  |
| 1.5 | DR-MAD     | 0.9183  | 27.943  | 1.0581  | 10.747  | 1.0521  | 8.839   | 1.0349  | 13.035  | 0.9553  | 27.881  |
|     | 1/N        | 0.8699  | 48.321  | 1.0359  | 21.451  | 1.0099  | 14.332  | 1.0509  | 19.867  | 0.8916  | 39.032  |
|     | Mean-variance | 0.8908  | 32.449  | 1.0478  | 13.649  | 1.0557  | 9.118   | 1.0467  | 11.139  | 0.9469  | 29.045  |
|     | Classic MAD | 0.9051  | 28.943  | 1.0494  | 13.532  | 1.0617  | 8.744   | 1.0332  | 14.190  | 0.9552  | 28.428  |

The bold number refers to the best result.

in Table 1. From the statistics of S&P 500 index in Table 2, it can be observed that the daily datasets are more unstable than the weekly dataset. Moreover, the VIX index in datasets Day2007, Day2010 and Day2019 exceeds 20, which is approximately the median value according to historical data [49]. This demonstrates that during these periods the market has high uncertainty and investors are relatively more nervous. Experiments on the daily datasets were conducted considering monthly rebalance. In this sense, the estimation window and the prediction window are 250 and 21 days (one month), respectively. Similar to the experiment on weekly data, the same approach for cross-validation was used in the parameter estimation. The only difference is the searching window and step length.

In this subsection, we consider FinVal for profitability and MDR for robustness. Table 4 presents the performance of all four models with different target returns. As can be seen from the table, DR-MAD model outperforms benchmarks in both FinVal and MDR in Day2007, Day2010 and Day2019, which are more volatile and have higher VIX than other datasets. However, DR-MAD model does not result in significant superiority in the rest of two datasets, especially in Day2017. The naive 1/N strategy outperforms all the other methods in cumulative return regardless of different target returns. The reason might be that, although 1/N strategy has an poor ability to defend uncertainty, it acquires a high investment returns due to the strong bull market in the second half of the testing period, and the robust strategy might be too conservative in contrast. Nevertheless, compared with classic MAD model, DR-MAD model still achieves a better MDR under most circumstances in this situation, which demonstrates that the DR-MAD model is more suitable when data distribution is with perturbations. Together with the experiments on weekly data, we conclude that DR-MAD model is highly competent for portfolio management in uncertain markets.

5 Conclusion

In this work, a novel Wasserstein metric-based data-driven distributionally robust MAD model is proposed for portfolio selection. Although MAD model is widely used for risk control, its combination with the WDRO framework is brand new. To solve the proposed
model, we show that it can be reformulated as two simple finite-dimensional LPs, which can be solved as easily as solving the classic MAD model. Furthermore, the proposed model is compared with the 1/N, classic MAD and mean-variance model on S&P 500 constituent stocks in different settings. The experimental results show that the portfolios constructed by DR-MAD model are superior to the benchmarks in terms of profitability and stability in most cases. The experiments suggest the model is highly competent for portfolio management in uncertain markets.

Nevertheless, the proposed DR-MAD model is a single-period framework and does not consider transaction costs and short selling. Future works can be conducted on extending the model to dynamic settings with more realistic constraints and designing effective algorithms.

Acknowledgements This work was supported by the National Natural Science Foundation of China [Grant Numbers 11871269 and 71801124]; and the Natural Science Foundation of Jiangsu Province [Grant Number BK20181259].

Data Availability All data generated or analyzed during this study are included in this published article. The S&P 500 data is deposited in the Wind-Economic Database (https://www.wind.com.cn/).

6 Appendix

6.1 Appendix I: Proof of Proposition 1.

For any feasible distribution \( P \in \mathcal{M}(\Xi) \), and support set \( \Xi = \mathbb{R}^m \), we use the similar idea as in [36] to transform problem (5) into the following conic linear program

\[
\begin{align*}
\max_{\Pi(\xi, \hat{\xi}_i) \geq 0} & \quad \int_{\Xi} \frac{1}{N} \sum_{i=1}^{N} |\xi^T x - \alpha| \Pi(\xi, \hat{\xi}_i), \\
\text{s.t.} & \quad \int_{\Xi} \sum_{i=1}^{N} \xi^T x \Pi(\xi, \hat{\xi}_i) = \alpha, \\
& \quad \int_{\Xi} \Pi(\xi, \hat{\xi}_i) = \frac{1}{N}, \quad i = 1, \ldots, N, \\
& \quad \int_{\Xi} N \sum_{i=1}^{N} \| \xi - \hat{\xi}_i \| \Pi(\xi, \hat{\xi}_i) \leq \varepsilon. 
\end{align*}
\]

The Lagrangian function for (12) is

\[
L(\xi, s, \gamma_1, \gamma_2) = \int_{\Xi} \sum_{i=1}^{N} (|\xi^T x - \alpha| - s_i - \gamma_1 \xi^T x - \gamma_2 \| \xi - \hat{\xi}_i \|) \Pi(\xi, \hat{\xi}_i) + \frac{1}{N} \sum_{i=1}^{N} s_i + \gamma_1 \alpha + \gamma_2 \varepsilon.
\]

where \( s \in \mathbb{R}^{N}, \gamma_1 \in \mathbb{R}, \gamma_2 \geq 0 \). Then, the Lagrange dual function is

\[
\max_{\Pi(\xi, \hat{\xi}_i) \geq 0} L(\xi, s, \gamma_1, \gamma_2),
\]

and the dual problem of (12) can be represented as

\[
\begin{align*}
\min_{s, \gamma_1, \gamma_2} & \quad \frac{1}{N} \sum_{i=1}^{N} s_i + \gamma_1 \alpha + \gamma_2 \varepsilon, \\
\text{s.t.} & \quad s_i + \gamma_1 \xi^T x + \gamma_2 \| \xi - \hat{\xi}_i \| \geq |\xi^T x - \alpha|, \quad \forall \xi \in \Xi, \ i = 1, \ldots, N, \\
& \quad \gamma_2 \geq 0.
\end{align*}
\]
For any $\varepsilon > 0$, we can easily verify that if the classic MAD (2) is feasible, $\Pi_0 = \mathbb{P}_N \times \mathbb{P}_N$ is a strictly feasible solution for the primal problem (12). Then the Slater condition for strong duality holds according to [46], and the optimal value of (13) is equivalent to (12). Furthermore, problem (13) can be transformed into

$$\begin{align*}
\min_{s, y_1, y_2} & \quad \frac{1}{N} \sum_{i=1}^{N} s_i + \gamma_1 y_1 + \gamma_2 \varepsilon, \\
\text{s.t.} & \quad s_i + \gamma_1 \xi^T x + \gamma_2 \| \xi - \hat{\xi}_i \| \geq \xi^T x - \alpha, \quad \forall \xi \in \Xi, \ i = 1, \ldots, N, \\
& \quad s_i + \gamma_1 \xi^T x + \gamma_2 \| \xi - \hat{\xi}_i \| \geq -\xi^T x + \alpha, \quad \forall \xi \in \Xi, \ i = 1, \ldots, N, \\
& \quad \gamma_2 \geq 0.
\end{align*}$$

As the definition of the dual norm, problem (15) can be reformulated as

$$\begin{align*}
\min_{s, y_1, y_2} & \quad \frac{1}{N} \sum_{i=1}^{N} s_i + \gamma_1 y_1 + \gamma_2 \varepsilon, \\
\text{s.t.} & \quad s_i \geq \max_{\xi \in \Xi} \min_{\|z_{i,1}\| \leq \gamma_2} \{ \xi^T x - \alpha - \gamma_1 \xi^T x - z_{i,1}^T (\xi - \hat{\xi}_i) \}, \ i = 1, \ldots, N, \\
& \quad s_i \geq \max_{\xi \in \Xi} \min_{\|z_{i,2}\| \leq \gamma_2} \{ -\xi^T x + \alpha - \gamma_1 \xi^T x - z_{i,2}^T (\xi - \hat{\xi}_i) \}, \ i = 1, \ldots, N, \\
& \quad \gamma_2 \geq 0.
\end{align*}$$

In accordance with the min-max theorem in [3][Proposition 5.5.4], we exchange the max-min order in (16) constraints. Moreover, by adding $z_{i,1}$ and $z_{i,2}$, $i = 1, \ldots, N$ to decision variables, the resulting problem can be further reduced to

$$\begin{align*}
\min_{s, y_1, y_2} & \quad \frac{1}{N} \sum_{i=1}^{N} s_i + \gamma_1 y_1 + \gamma_2 \varepsilon, \\
\text{s.t.} & \quad s_i \geq \max_{\xi \in \Xi} \min_{\|z_{i,1}\| \leq \gamma_2} \{ \xi^T x - \alpha - \gamma_1 \xi^T x - z_{i,1}^T (\xi - \hat{\xi}_i) \}, \ i = 1, \ldots, N, \\
& \quad s_i \geq \max_{\xi \in \Xi} \min_{\|z_{i,2}\| \leq \gamma_2} \{ -\xi^T x + \alpha - \gamma_1 \xi^T x - z_{i,2}^T (\xi - \hat{\xi}_i) \}, \ i = 1, \ldots, N, \\
& \quad \|z_{i,1}\|_* \leq \gamma_2, \ i = 1, \ldots, N, \\
& \quad \|z_{i,2}\|_* \leq \gamma_2, \ i = 1, \ldots, N.
\end{align*}$$

$$\begin{align*}
\min_{s, y_1, y_2} & \quad \frac{1}{N} \sum_{i=1}^{N} s_i + \gamma_1 y_1 + \gamma_2 \varepsilon, \\
\text{s.t.} & \quad s_i \geq \max_{\xi \in \Xi} \min_{\|z_{i,1}\| \leq \gamma_2} \{ \xi^T (x - y_1 x - z_{i,1}) - \alpha + (\hat{\xi}_i)^T z_{i,1} \}, \ i = 1, \ldots, N, \\
& \quad s_i \geq \max_{\xi \in \Xi} \min_{\|z_{i,2}\| \leq \gamma_2} \{ \xi^T (x - y_1 x - z_{i,2}) + \alpha + (\hat{\xi}_i)^T z_{i,2} \}, \ i = 1, \ldots, N, \\
& \quad \|z_{i,1}\|_* \leq \gamma_2, \ i = 1, \ldots, N, \\
& \quad \|z_{i,2}\|_* \leq \gamma_2, \ i = 1, \ldots, N.
\end{align*}$$
Consider $\mathcal{E} = \mathbb{R}^m$, $z_{i,1} = (x - \gamma_1 x)$ and $z_{i,2} = (-x - \gamma_1 x)$ hold to ensure that (18) make sense. By further eliminating $z_i$, 1 and $z_i$, 2, $i = 1, \ldots, N$, we obtain

$$
\begin{align*}
\min_{s, \gamma_1, \gamma_2} & \quad \frac{1}{N} \sum_{i=1}^{N} s_i + \gamma_1 \alpha + \gamma_2 \varepsilon, \\
\text{s.t.} & \quad s_i \geq -\alpha + (\hat{\xi}_i)^T (x - \gamma_1 x), \quad i = 1, \ldots, N, \\
& \quad s_i \geq \alpha - (\hat{\xi}_i)^T (x + \gamma_1 x), \quad i = 1, \ldots, N, \\
& \quad \|x - \gamma_1 x\|_* \leq \gamma_2, \\
& \quad \|x + \gamma_1 x\|_* \leq \gamma_2, \\
\end{align*}
$$

(19)

$$
\begin{align*}
\min_{s, \gamma_1, \gamma_2} & \quad \frac{1}{N} \sum_{i=1}^{N} s_i + \gamma_1 \alpha + \gamma_2 \varepsilon, \\
\text{s.t.} & \quad s_i \geq |(\hat{\xi}_i)^T x - \alpha| - \gamma_1 (\hat{\xi}_i)^T x, \quad i = 1, \ldots, N, \\
& \quad \|x - \gamma_1 x\|_* \leq \gamma_2, \\
& \quad \|x + \gamma_1 x\|_* \leq \gamma_2. \\
\end{align*}
$$

(20)

Indeed, the optimal solution exists only if the first $N$ inequality constraints in (20) satisfies $s_i = |(\hat{\xi}_i)^T x - \alpha| - \gamma_1 (\hat{\xi}_i)^T x$, $i = 1, \ldots, N$. Based on this, we put $s_i$, $i = 1, \ldots, N$, into the objective function. Then, problem (20) is equivalent to

$$
\begin{align*}
\min_{\gamma_1, \gamma_2} & \quad \frac{1}{N} \sum_{i=1}^{N} |(\hat{\xi}_i)^T x - \alpha| + \gamma_1 (\hat{\mu}^T x - \alpha) + \gamma_2 \varepsilon, \\
\text{s.t.} & \quad \|(1 - \gamma_1) x\|_* \leq \gamma_2, \\
& \quad \|(1 + \gamma_1) x\|_* \leq \gamma_2, \\
\end{align*}
$$

(21)

where $\hat{\mu} = \frac{1}{N} \sum_{i=1}^{N} \hat{\xi}_i$, and the proof completes.

### 6.2 Appendix II: Proof of Proposition 2.

Model (6) is equivalent to

$$
\begin{align*}
\min_{\gamma_1, \gamma_2, \omega_1, \omega_2} & \quad \frac{1}{N} \sum_{i=1}^{N} |(\hat{\xi}_i)^T x - \alpha| + \gamma_1 (\hat{\mu}^T x - \alpha) + \gamma_2 \varepsilon, \\
\text{s.t.} & \quad \|\omega_1\|_* \leq \gamma_2, \\
& \quad \|\omega_2\|_* \leq \gamma_2, \\
& \quad \omega_1 = (1 - \gamma_1) x, \\
& \quad \omega_2 = (1 + \gamma_1) x, \\
\end{align*}
$$

(22)

Dual theory is applied to the above model and the Lagrangian function for (22) is

$$
L(\gamma, \omega, \lambda, \nu) = \frac{1}{N} \sum_{i=1}^{N} |(\hat{\xi}_i)^T x - \alpha| + \gamma_1 (\hat{\mu}^T x - \alpha) + \gamma_2 \varepsilon + \lambda_1 (\|\omega_1\|_* - \gamma_2) + \lambda_2 (\|\omega_2\|_* - \gamma_2) + v_1^T ((1 - \gamma_1) x - \omega_1) + v_2^T ((1 + \gamma_1) x - \omega_2),
$$

$$
= \sum_{i=1}^{N} |(\hat{\xi}_i)^T x - \alpha| + (v_1 + v_2)^T x + \gamma_1 (\hat{\mu}^T x - \alpha - v_1^T x + v_2^T x) + \gamma_2 (\varepsilon - \lambda_1 - \lambda_2) + (\lambda_1 \|\omega_1\|_* - v_1^T \omega_1) + (\lambda_2 \|\omega_2\|_* - v_2^T \omega_2),
$$

where $v_1, v_2 \in \mathbb{R}^m$ and $\lambda_1, \lambda_2 \in \mathbb{R}_+$. Then, the Lagrange dual function is

$$
\min_{\gamma_2 \geq 0, \gamma_1, \omega} L(\gamma, \omega, \lambda, \nu).
$$
According to the conjugate function of norm, the dual problem of (22) can be represented as

\[
\begin{align*}
\max_{\lambda, \gamma, \nu, \mu} \quad & (v_1 + v_2)^T x, \\
\text{s.t.} \quad & \gamma_1 = 0, \quad \gamma_2 = 2, \quad \omega_1 = \omega_2 = x, \\
\quad & -\lambda_1 x \leq v_1 \leq \lambda_1 x, \\
\quad & -\lambda_2 x \leq v_2 \leq \lambda_2 x.
\end{align*}
\] (23)

Considering the equivalence of norms, we set \( \| \cdot \| = \| \cdot \|_1 \) in (22) without loss of generality.

For any \( x \in X \), It can be verified that \( \gamma_1 = 0, \gamma_2 = 2, \omega_1 = \omega_2 = x \) is a strictly feasible solution for (22), then the strong duality holds and the proof completes.

**Appendix III: Proof of Proposition 3.**

Suppose \( \| \cdot \| = \| \cdot \|_\infty \), problem (8) can be expressed as

\[
\begin{align*}
\max_{\lambda, \gamma, \nu, \mu} \quad & (v_1 + v_2)^T x, \\
\text{s.t.} \quad & \hat{\mu}^T x - v_1^T x + v_2^T x = \alpha, \\
\quad & \lambda_1 + \lambda_2 \leq \epsilon, \\
\quad & -\lambda_1 e \leq v_1 \leq \lambda_1 e, \\
\quad & -\lambda_2 e \leq v_2 \leq \lambda_2 e.
\end{align*}
\] (24)

As the portfolio weights \( x \geq 0 \), the following inequality holds

\[
- (\lambda_1 + \lambda_2) e^T x \leq (v_1 + v_2)^T x \leq (\lambda_1 + \lambda_2) e^T x.
\] (25)

Given that \( e^T x = 1 \), we obtain \( (v_1 + v_2)^T x \leq (\lambda_1 + \lambda_2) \leq \epsilon \), which means the maximum value of (24) will not be greater than \( \epsilon \). Later, we prove that if the problem (24) is feasible, there is always an solution under which (24) reaches the optimal value \( \epsilon \).

Similarly to (25), the following inequality holds

\[
- (\lambda_1 + \lambda_2) e^T x \leq (v_1 - v_2)^T x \leq (\lambda_1 + \lambda_2) e^T x,
\] (26)

and \( (v_1 - v_2)^T x \in [-\epsilon, \epsilon] \). Note that \( \alpha \) and \( x \) are two given constants in (24), and the first constraint is \( \hat{\mu}^T x - \alpha = (v_1 - v_2)^T x \). Then, it can be observed that if \( \hat{\mu}^T x - \alpha \notin [-\epsilon, \epsilon] \), problem (24) will be infeasible. If \( \hat{\mu}^T x - \alpha \in [-\epsilon, \epsilon] \), it can be verified that \( \lambda_1^* = \frac{\epsilon - (\hat{\mu}^T x - \alpha)}{2}, \lambda_2^* = \frac{\epsilon - (\hat{\mu}^T x - \alpha)}{2}, v_1^* = \lambda_1^* e, v_2^* = \lambda_2^* e \) is an optimal solution.

**Appendix IV: Proof of Theorem 1**

Suppose \( \Xi = \mathbb{R}^m \), and using the infinity norm \( \| \cdot \|_\infty \) in Wasserstein metric, Propositions 1–3 hold. Then, put the result of Proposition 3 into (7), DR-MAD model can be expressed as

\[
\min_{x \in X} \max_{\alpha \geq \rho} \left\{ \frac{1}{N} \sum_{i=1}^{N} |(\hat{\xi}_i)^T x - \alpha| + \epsilon, \quad \text{if} \quad \mu^T x - \alpha \in [-\epsilon, \epsilon], \right. \quad \text{otherwise},
\] (27)
\[
\begin{align*}
\min_{x \in \mathcal{X}} & \quad \alpha \sum_{i=1}^{N} |(\hat{\xi}_i)^T x - \alpha| + \varepsilon \\
\text{s.t.} & \quad \alpha \geq \rho \\
& \quad \alpha \geq \hat{\mu}^T x - \varepsilon \\
& \quad \alpha \leq \hat{\mu}^T x + \varepsilon 
\end{align*}
\] (28)

Note that the objective function of the inner problem in (28) is one-dimensional convex to \(\alpha\). Thus, if the inner problem is feasible, the maximum value must be obtained at one of the endpoints. Based on this, (28) can be reduced to

\[
\min_{x \in \mathcal{X}} \begin{cases} 
\max \left\{ \frac{1}{N} \sum_{i=1}^{N} |\hat{\mu}^T x - (\hat{\xi}_i)^T x - \varepsilon| + \varepsilon, \frac{1}{N} \sum_{i=1}^{N} |\hat{\mu}^T x - (\hat{\xi}_i)^T x + \varepsilon| + \varepsilon \right\}, & \text{if } \hat{\mu}^T x - \varepsilon \geq \rho, \\
\max \left\{ \frac{1}{N} \sum_{i=1}^{N} |\rho - (\hat{\xi}_i)^T x| + \varepsilon, \frac{1}{N} \sum_{i=1}^{N} |\hat{\mu}^T x - (\hat{\xi}_i)^T x + \varepsilon| + \varepsilon \right\}, & \text{if } \hat{\mu}^T x - \varepsilon \leq \rho \leq \hat{\mu}^T x + \varepsilon, \\
+\infty, & \text{if } \hat{\mu}^T x + \varepsilon \leq \rho. 
\end{cases}
\] (29)

Finally, problem (29) can be equivalent to two linear programs using some basic optimization techniques, which completes the proof.

**References**

1. Al Janabi, M.A.M., Hernandez, J.A., Berger, T., Khuong, N.D.: Multivariate dependence and portfolio optimization algorithms under illiquid market scenarios. Eur. J. Oper. Res. 259(3), 1121–1131 (2017)
2. Ambrosio, L., Gigli, N.: A user’s guide to optimal transport. In: Modelling and optimisation of flows on networks. Springer, 1–155 (2013)
3. Bertsekas, D.P.: Convex Optimization Theory. Athena Scientific Belmont, Belmont (2009)
4. Blanchet, J., Chen, L., Zhou, X.Y.: Distributionally robust mean-variance portfolio selection with wasserstein distances (2018)
5. Calafiore, G.C.: Ambiguous risk measures and optimal robust portfolios. SIAM J. Optim. 18(3), 853–877 (2007)
6. Chan, L.K.C., Karceski, J.: On portfolio optimization: forecasting covariances and choosing the risk model. Rev. Financ. Stud. 12(5), 937–974 (1999)
7. Caihua, C., Xindan, L., Caleb, T., Suyang, W., Yinyu, Y.: Sparse portfolio selection via quasi-norm regularization. arXiv:1312.6350 (2013)
8. Chiu, C.-H., Choi, T.-M.: Supply chain risk analysis with mean-variance models: a technical review. Ann. Oper. Res. 240(2), 489–507 (2016)
9. Choi, T.-M., Li, D., Yan, H.: Mean-variance analysis of a single supplier and retailer supply chain under a returns policy. Eur. J. Oper. Res. 184(1), 356–376 (2008)
10. Choi, T.-M., Wen, X., Sun, X., Chung, S.-H.: The mean-variance approach for global supply chain risk analysis with air logistics in the blockchain technology era. Transp. Res. Part E Logist. Transp. Rev. 127, 178–191 (2019)
11. Cootner, P.H.: The Random Character of Stock Market Prices. MIT Press, Cambridge (1967)
12. Cui, X., Sun, X., Zhu, S., Jiang, R., Li, D.: Portfolio optimization with nonparametric value at risk: a block coordinate descent method. INFORMS J. Comput. 30(3), 454–471 (2018)
13. De Roon, F.A., Nijman, T.E.: Currency hedging for international stock portfolios: the usefulness of mean-variance analysis. J. Bank. Financ. 27(2), 327–349 (2003)
14. Delage, E., Ye, Y.: Distributionally robust optimization under moment uncertainty with application to data-driven problems. Oper. Res. 58(3), 595–612 (2010)
15. De Miguel, V., Nogales, F.J.: Portfolio selection with robust estimation. Oper. Res. 57(3), 560–577 (2009)
16. DeMiguel, V., Garlappi, L., Nogales, F.J., Uppal, R.: A generalized approach to portfolio optimization: improving performance by constraining portfolio norms. Manag. Sci. 55(5), 798–812 (2009)
17. Engel, C., Frankel, J.A., Froot, K.A., Rodrigues, A.P.: Tests of conditional mean-variance efficiency of the US stock market. J. Empir. Financ. 2(1), 3–18 (1995)
18. Erdoğan, E., Iyengar, G.: Ambiguous chance constrained problems and robust optimization. Math. Program. 107(1–2), 37–61 (2006)
19. Feldstein, M.S.: Mean-variance analysis in the theory of liquidity preference and portfolio selection. Rev. Econ. Stud. 36(1), 5–12 (1969)
20. Gao, R., Chen, X., Kleywegt, A.J.: Wasserstein distributionally robust optimization and variation regularization. arXiv:1712.06050 (2017)
21. Goh, J., Sim, M.: Distributionally robust optimization and its tractable approximations. Oper. Res. 58(4–part-1), 902–917 (2010)
22. Hodges, C.W., Taylor, W.R.L., Yoder, J.A.: Stocks, bonds, the sharpe ratio, and the investment horizon. Financ. Anal. J. 53(6), 74–80 (1997)
23. Jagannathan, R., Ma, T.: Risk reduction in large portfolios: why imposing the wrong constraints helps. J. Financ. 58(4), 1651–1683 (2003)
24. Jiang, R., Guan, Y.: Data-driven chance constrained stochastic program. Math. Program. 158(1–2), 291–327 (2016)
25. Jiang, R., Guan, Y.: Risk-averse two-stage stochastic program with distributional ambiguity. Oper. Res. 66(5), 1390–1405 (2018)
26. Kang, Z., Li, X., Li, Z., Zhu, S.: Data-driven robust mean-CVaR portfolio selection under distribution ambiguity. Quant. Financ. 19(1), 105–121 (2019)
27. Kantorovich, L., Rubinshtein, G.: On a space of totally additive functions, vestn. Vestn Lening. Univ. 13, 52–59 (1958)
28. Konno, H., Koshizuka, T.: Mean-absolute deviation model. Iie Trans. 37(10), 893–900 (2005)
29. Konno, H., Yamazaki, H.: Mean-absolute deviation portfolio optimization model and its applications to tokyo stock market. Manag. Sci. 37(5), 519–531 (1991)
30. Konno, H., Shirakawa, H., Yamazaki, H.: A mean-absolute deviation-skewness portfolio optimization model. Ann. Oper. Res. 45(1), 205–220 (1993)
31. Daniel, K., Esfahani, P.M., Nguyen, V.A., Shafieezadeh-Abadeh, S.: Wasserstein distributionally robust optimization: theory and applications in machine learning. In: Operations research & management science in the age of analytics, pp. 130–166. INFORMS (2019)
32. Ledoit, O., Wolf, M.: Robust performance hypothesis testing with the sharpe ratio. J. Empir. Financ. 15(5), 850–859 (2008)
33. Levy, H.: Stochastic dominance and expected utility: survey and analysis. Manag. Sci. 38(4), 555–593 (1992)
34. Li, J., Wang, L.: A minimax regret approach for robust multi-objective portfolio selection problems with ellipsoidal uncertainty sets. Comput. Ind. Eng. 147, 106646 (2020)
35. Markowitz, H.: Portfolio selection. J. Financ, 77–91 (1952)
36. Peyman, M.E., Daniel, K.: Data-driven distributionally robust optimization using the wasserstein metric: performance guarantees and tractable reformulations. Math. Program. 171(1–2), 115–166 (2018). https://doi.org/10.1007/s10107-017-1172-1
37. Moon, Y., Yao, T.: A robust mean absolute deviation model for portfolio optimization. Comput. Oper. Res. 38(9), 1251–1258 (2011)
38. Ogryczak, W., Ruszczyński, A.: From stochastic dominance to mean-risk models: semideviations as risk measures. Eur. J. Oper. Res. 116(1), 33–50 (1999)
39. Ogryczak, W., Ruszczyński, A.: On consistency of stochastic dominance and mean-semideviation models. Math. Program. 89(2), 217–232 (2001)
40. Emmanouil, P., Charles, S., Xiaoxia, Y.: Horses for courses: mean-variance for asset allocation and 1/n for stock selection. Eur. J. Oper. Res. 288(1), 302–317 (2021)
41. Qin, Z.: Random fuzzy mean-absolute deviation models for portfolio optimization problem with hybrid uncertainty. Appl. Soft Comput. 56, 597–603 (2017)
42. Rahimian, H., Mehotra, S.: Distributionally robust optimization: a review. arXiv:1908.05659, (2019)
43. Roques, F., Hiroux, C., Saguan, M.: Optimal wind power deployment in europe-a portfolio approach. Energy Policy 38(7), 3245–3256 (2010)
44. Roques, F.A., Newbery, D.M., Nuttall, W.J.: Fuel mix diversification incentives in liberalized electricity markets: a mean-variance portfolio theory approach. Energy Econ. 30(4), 1831–1849 (2008)
45. Scarf, H.: A min-max solution of an inventory problem. Studies in the mathematical theory of inventory and production (1958)
46. Shapiro, A.: On duality theory of conic linear problems. In: Semi-infinite programming, pp. 135–165. Springer, (2001)
47. Simaan, Y.: Estimation risk in portfolio selection: the mean variance model versus the mean absolute deviation model. Manag. Sci. 43(10), 1437–1446 (1997)
48. Wang, Z., You, K., Shiji Song, S., Zhang, Y.: Wasserstein distributionally robust shortest path problem. Eur. J. Oper. Res. 284(1), 31–43 (2020)
49. Whaley, R.E.: Understanding the vix. J. Portfolio Manag. 35(3), 98–105 (2009)
50. Wiesemann, W., Kuhn, D., Sim, M.: Distributionally robust convex optimization. Oper. Res. 62(6), 1358–1376 (2014)
51. Wozabal, D.: A framework for optimization under ambiguity. Ann. Oper. Res. 193(1), 21–47 (2012)
52. Mei, Yu., Wang, S.: Dynamic optimal portfolio with maximum absolute deviation model. J. Global Optim. 53(2), 363–380 (2012)
53. Mei, Yu., Takahashi, S., Inoue, H., Wang, S.: Dynamic portfolio optimization with risk control for absolute deviation model. Eur. J. Oper. Res. 201(2), 349–364 (2010)
54. Zabarankin, M., Pavlikov, K., Uryasev, S.: Capital asset pricing model (capm) with drawdown measure. Eur. J. Oper. Res. 234(2), 508–517 (2014)
55. Zhu, R., Wei, H., Bai, X.: Wasserstein metric based distributionally robust approximate framework for unit commitment. IEEE Trans. Power Syst. 34(4), 2991–3001 (2019)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.