Applied Mathematical Modelling 117 (2023) 513–528

Contents lists available at ScienceDirect

Applied Mathematical Modelling

journal homepage: www.elsevier.com/locate/apm

Distributionally robust optimization with Wasserstein metric for multi-period portfolio selection under uncertainty

Zhongming Wu*, Kexin Sun

Research Center of Risk Management and Emergency Decision Making, School of Management Science and Engineering, Nanjing University of Information Science and Technology, Nanjing 210044, PR China

ARTICLE INFO

Article history:
Received 21 August 2022
Revised 15 December 2022
Accepted 20 December 2022
Available online 24 December 2022

Keywords:
Portfolio selection
Multi-period
Mean-variance
Distributionally robust optimization
Wasserstein metric

ABSTRACT

The mean-variance model formulated by Markowitz for a single period serves as a fundamental method of modern portfolio selection. In this study, we consider a multi-period case with uncertainty that better matches the reality of the financial market. Using the Wasserstein metric to characterize the uncertainty of returns in each period, a new distributionally robust mean-variance model is proposed to solve multi-period portfolio selection problem. We further transform the developed model into a tractable convex problem using duality theory. We also apply a nonparametric bootstrap method and provide a specific algorithm to estimate the radius of the Wasserstein ball. The effects of the parameters on the corresponding strategy and evaluation criteria of portfolios are analyzed using in-sample data. The analysis indicate that the return and risk of our portfolio selections are relatively immune to parameter values. Finally, a series of out-of-sample experiments demonstrate that the proposed model is superior to some other models in terms of final wealth, standard deviation, and Sharpe ratio.

1. Introduction

The core of modern portfolio theory is focused on balancing returns and risks in the investment of available assets, with the aim of determining an effective portfolio strategy that allocates wealth in a manner that minimizes the variance of returns of the portfolio while maximizing its mean. However, the return rate of each asset is usually not explicitly given. Thus, in this study, we consider multi-period portfolio selection problem under uncertainty based on the classical mean-variance (MV) framework.

Markowitz [1] first suggested measuring the mean of return rates as the investment gain and treating the variance of returns as the risk. Since then, many studies have extended the MV model under different settings, such as minimum investment, maximum investment, and cardinality constraints, to improve its realism (e.g., [2–5]), and Cui et al. [6] reviewed the MV portfolio selection model and its variants. However, the above-mentioned studies were limited to single-period situations. While in the actual decision-making process, investors would like to adjust their asset positions multiple times depending on market conditions, that is, to have a selection model for a multi-period process [6–8]. Wealth is reallocated at the beginning of each period with the ultimate goal of maximizing returns upon exit from the market. Actually, the MV model was first extended to a multi-period situation by Li and Ng [7]. Cui et al. [8] enforced a no-shorting constraint on

* Corresponding author.
E-mail address: wuzm@seu.edu.cn (Z. Wu).

https://doi.org/10.1016/j.apm.2022.12.037

a dynamic MV model and derived a semi-analytical expression of the value function. Cui et al. [9] used the mean-field formulation to address the issue of non-separability for multi-period MV portfolio selection problems. More recently, Xiao et al. [10] developed a time-consistent strategy for multi-period MV portfolio selection. Corsaro et al. [11] introduced a multi-period mean-variance model based on the fused LASSO approach to reduce holding and transaction costs. Li et al. [12] studied the multi-period portfolio optimization problem using mean-variance and risk parity asset allocation frameworks. We refer to recent research involving a multi-period mean-variance portfolio selection model [6]. Note that the above works were all based on the given return rate of each invested asset, while characterizing uncertainty is another challenging problem.

Broadly speaking, the underlying distribution of random variables, such as the rate of return of investment assets, is not known in practice [13,14]. In this case, robust optimization (RO) and distributionally robust optimization (DRO) are efficient tools for coping with uncertainty. DRO assumes that the uncertain distribution is contained in an ambiguity set, where the ambiguity set is constructed by the incomplete distribution information of the random variables. More details on DRO can be found in [15–17]. Now we introduce two kinds of ambiguity set in DRO framework. The first type is the ambiguity set with generalized moment information developed by Delage and Ye [18]. The authors introduced two parameters as tools for controlling the size of the moment-based ambiguity set. Kang et al. [19] and Liu et al. [20] proposed new distributionally robust portfolio models with similar moment-based ambiguity sets. Inspired by the moment-based DRO approach, Li et al. [21] proposed two distributionally robust model predictive control algorithms for a class of discrete linear systems with chance constraints. Besides, the moment-based DRO was applied to solve many other practical problems in many fields such as dynamical system [22], multiple linear regression [23], MIMO system [24] and beamformer design [25]. The second type of ambiguity set is constructed based on a metric. Compared with the moment-based ambiguity set, the metric-based one contains considerable moment information by empirical distribution, which is more flexible in controlling conservativeness [26]. Metric ambiguity is mainly based on relative entropy, for example, the Kullback-Leibler (KL) divergence in Hu and Hong [27] and the Wasserstein distance [28].

DRO approaches have been extended to study the uncertainty in field of portfolio selection, which investigate the uncertainty of return of investment assets instead of assuming that the return or its distribution is known. Lobo and Boyd [29] selected box and ellipsoidal constraints to estimate the mean and covariance of return rates, respectively, and applied semidefinite programming and projection methods to compute the worst-case solution of single-stage portfolio selection. Gülpınar and Rustem [30] proposed a robust multi-period MV optimization and used a scenario tree to estimate the uncertainty for future stages. In terms of the single-period DRO, Pflflug and Wozabal [31] used Wasserstein distance to describe a portfolio ambiguity sets. Subsequently, Wozabal [32] considered a distributionally robust stochastic problem based on Wasserstain balls and extended a study conducted by Pflug and Wozabal [31] to a more general case. Kang et al. [19] used a zero net adjustment framework by adding a linear constraint to a moment-based ambiguity set to reduce the conservativeness of the DRO. Sun et al. [33] proposed a new mean-CVaR model whose ambiguity set is constructed by moment information for portfolio optimization. Du et al. [34] proposed a new method for the DRO portfolio problem, in which they developed a mean-CVaR model with a Wasserstein ball as its ambiguity set. They also derived a tractable problem under the uncertainty set of the cone and obtained some theoretical conclusions about the box, budget, and ellipsoid uncertainty sets. However, when it comes to multi-period DRO, even fewer papers have been studied. Liu and Chen [35] introduced new multi-period mean-CVaR models for markets with regime-switching properties, which made the model more robust in reflecting a changing market environment. Jiang et al. [36] considered a multi-period portfolio model based on moment DRO consisting of riskless and risky assets. Nevertheless, compared with moment-based DRO, metric-based DRO generally includes all moment information because it considers the nominal distribution. All the above studies examined only single-period situations or moment-based DRO, which motivate us to address a new multi-period portfolio selection problem using metric-based methods.

In practice, most investors prefer to make medium- or long-term investment decisions, and the investment strategy should be adjusted dynamically by changing market information. This requirement also inspire us to consider a combination of DRO and the mean-variance model in a multi-period case. Among all the metrics-based ambiguity sets, the KL ambiguity set is the most popular one which can provide more modeling flexibility while keep the computational tractability [37]. However, KL ambiguity sets are not capable of expressing confidence sets for unknown distributions. In comparison, as noted in [38], the Wasserstein ball can provide natural confidence sets for unknown distributions. Additionally, Wasserstein distance can be used to measure the difference between distributions, even if their support sets are completely different [28], compared with KL divergence. By modifying the radius of the ambiguity set, we can flexibly adjust the degree of conservativeness in optimization problem. If the radius equals to zero, the DRO model will be reduced to an ambiguity-free stochastic program. Hence, from the point of view of numerical experiments, Wasserstein-based DRO model can achieve better out-of-sample performance. These advantages inspire us to choose an ambiguity set using the Wasserstein metric.

The contributions of this study can be summarized as follows: First, we develop a multi-period distributionally robust MV model using the Wasserstein metric. However, the objective function and constraint are nonconvex. To alleviate this difficulty, we apply the dual theory twice to equate the problem to a convex problem. Moreover, we use the multiplier method to transform the single-objective problem into a multi-objective problem containing a transaction cost term, which captures the risk-return trade-off. Furthermore, we adopt a bootstrap method to estimate the radius of the Wasserstein balls under time-varying market information and examine whether the investment strategy, return, and risk of portfolios can be remarkably influenced by the different radius of the Wasserstein balls. Finally, we compare the performance of the new method with the following three ones: the multi-period mean-variance model with known distribution, and Olivares-Nadal-DeMiguel model (ODM) in DeMiguel and Olivares-Nadal [39] and two-stage model (Two-Stage) in Li et al. [40]. In our experiments, the S&P 500 is selected as our stock data, and we then apply the rolling window procedure in DeMiguel et al. [41] for the out-of-sample test and show the better performance of our model on four evaluation values. Computational experiments reveal that, in most cases, our approach outperforms other models in terms of the final wealth, the mean and standard deviation, and Sharpe ratio of optimal portfolios.

The remainder of this paper is organized as follows: In Section 2, we describe the formulation of the multi-period distributionally robust MV model using the Wasserstein metric and definition of the Wasserstein ball. In Section 3, we describe the transformation of the model into a tractable form. Section 4 presents the empirical data used to demonstrate the robustness and practicality of our model. Concluding remarks are presented in Section 5.

2. Multi-period distributionally robust MV portfolio selection model

In this section, we develop the multi-period distributionally robust MV model using the Wasserstein metric, where the objective is to minimize the sum of the variance of returns of the multi-period in the worst case.

We consider a financial market with n risky assets available to be invested in a time horizon with T time periods. The random return rates of n assets are denoted as $\xi_t = \left[\xi_t^1, \xi_t^2, \dots, \xi_t^n\right]^\top$ which is defined on a probability space $(\Omega, \mathcal{F}_t, \mathbb{P}_t)$. We require that the filtration process $\mathcal{F}_t$, for $t = 1, 2, \dots, T$, satisfies $\mathcal{F}_0 = \{0, \Omega\}$ and $\mathcal{F}_t \subseteq \mathcal{F}_{t+1}$. $\mathbb{P}_t$ denotes the true distribution of ξ_t, $t = 1, 2, \dots, T$. We assume that an investor allocates his wealth among the n risky assets with initial wealth w_0 at $t = 0$ and reallocates the wealth at the beginning of the following period. Let $u_t = \left[u_t^1, u_t^2, \dots, u_t^n\right]^\top$ be the allocation of the investor on n assets in the market. At the beginning of each period, the investor can reallocate the current wealth w_t, $t = 0, 1, \dots, T-1$. The wealth w_t is equal to $u_{t+1}^\top e$, where $e = [1, 1, \dots, 1]^\top$ is the column vector of all ones with n elements. Correspondingly, the investment loss can be denoted as $-w_t = -u_t^\top \xi_t$, $t = 1, 2, \dots, T$.

In multi-period setting, the investment risk at time t in the random loss process on the time horizon $T = [t, m]$ can be measured as a conditional risk mapping $\rho_{t,m}(\boldsymbol{.})$. Chen et al. [42] presented a time consistent multi-period MV model based on separable expected conditional mapping, meaning that it is built by summing single-period risks. Additionally, in order to ensure that the preferences of investors remain consistent over time, the conception of time consistent also should be defined. In the following, we introduce the definition of separable expected conditional mapping by Chen et al. [42], which has been proved to be time consistent.

Definition 1. A conditional risk mapping $\rho_{t,m}(\boldsymbol{.})$ between t and m is separable if it can be expressed as

$$\rho_{t,m}(-w_{t,m}) = \sum_{i=t+1}^{m} \mathbb{E}[\rho_i(-w_i|\mathcal{F}_i)|\mathcal{F}_t)],$$

where $-w_{t,m}$ is the investment loss estimated from t to m, and $-w_t$ is the random loss adapted to the filtration $\mathcal{F}_t$.

Definition 1 states that multi-period risk can be decomposed into the sum of the single-period terms. In this study, we use the variance of returns as the risk measurement in a multi-period setting. More precisely, we consider the minimum variance in each period and minimize the sum of the variance of all periods:

$$\min_{w} \sum_{t=1}^{T} \mathrm{Var}(w_t).$$

The above model is efficient if the return rate of each asset in each period is exactly known. However, in practice, the return rate is usually not pre-determined and its distribution is not easy to characterize. This motivates us to utilize the DRO method to hedge against the distributional uncertainty.

Generally speaking, most investors care about not only minimizing investment risk, but also determining how to maximize investment returns. In the classic single-period MV framework, the targeted expected return of the portfolio is constrained by $u^\top \mathbb{E}_{\mathbb{P}}(\xi) = \sigma$, where $\mathbb{P}$ is the underlying distribution of the random return vector ξ. Considering the uncertainty of asset returns, we set the lowest acceptable target return as the lower bound of the single-period minimum investment returns estimated using an ambiguity set. Moreover, the dynamic processes of wealth should also be considered as

$$w_t = \sum_{i=1}^{n} u_t^i \xi_t^i, \quad t = 1, 2, \dots, T.$$

Based on the above, we develop the following multi-period distributionally robust mean-variance portfolio selection model with a no-shorting constraint (DRMV):

$$
\text{(DRMV)} \qquad \begin{aligned}
&\min_{w} \sum_{t=1}^{T} \max_{\mathbb{P}_t \in \mathcal{Q}_{\theta_t}\left(\mathbb{P}_{t,N_t}\right)} \mathrm{Var}_{\mathbb{P}_t}\left(w_t\right) & \\
&\text{s.t. } w_{t-1} = u_t^{\top} e, & t = 1, 2, \ldots, T, \\
&\min_{\mathbb{P}_t \in \mathcal{Q}_{\theta_t}\left(\mathbb{P}_{t,N_t}\right)} \mathbb{E}_{\mathbb{P}_t}(w_t) \geq \sigma_t, & t = 1, 2, \ldots, T, \\
&w_t = u_t^{\top} \xi_t, & t = 1, 2, \ldots, T, \\
&u_t \geq 0, & t = 1, 2, \ldots, T,
\end{aligned} \tag{1}
$$

where $w = [w_1, w_2, \ldots, w_T]^{\top}$, and σ_t denotes the lowest acceptable target return at t. The definition of $\mathcal{Q}_{\theta_t}\left(\mathbb{P}_{t,N_t}\right)$ will be detailed later (see (2)), and $\mathbb{P}_{t,N_t}$ is the empirical distribution of ξ_t with sample size $N_t,\ t = 1, 2, \ldots, T$, which is denoted by

$$
\mathbb{P}_{t,N_t} = \frac{1}{N_t} \sum_{i=1}^{N_t} \delta_{\hat{\xi}_{t,i}},
$$

where $\delta_{\hat{\xi}_{t,i}}$ is Dirac point measure at $\hat{\xi}_{t,i}$. Notably, the first and third constraints imply that the wealth at $t-1$ is the source that constitutes the allocation at t, and the allocation of n assets multiplied by the investment return is equivalent to the wealth at time t. The second constraint sets a lower bound on expectation of wealth at period t.

Unlike the study of portfolios by Xiao et al. [10] and Li and Ng [7] who minimize the terminal variance like

$$
\min \mathrm{Var}(w_T),
$$

we let the sum of single-period variances at each time t as our objective function inspired by Corsaro et al. [43]

$$
\min \sum_{t=1}^{T} u_t^{\top} C_t u_t,
$$

where C_t is the covariance matrix estimated at time t, and it is more in line with the multi-period situation. Nevertheless, Corsaro et al. [[43] did not take the uncertainty of random returns into account. We next give a following definition of the ambiguity set constructed by Wasserstein metric. The Wasserstein metric is defined on the space $\mathcal{M}(\Xi)$ of all probability distributions $\mathbb{Q}$ supported on Ξ with $\mathbb{E}^{\mathbb{Q}}[\|\xi\|] = \int_{\Xi} \|\xi\| \mathbb{Q}(\mathrm{d}\xi) < \infty$.

Definition 2. (Wasserstein metric [44]), the Wasserstein metric $D_W : \mathcal{M}(\Xi) \times \mathcal{M}(\Xi) \to \mathbb{R}_+$ is defined by

$$
D_W(\mathbb{Q}_1, \mathbb{Q}_2) := \inf \left\{ \int_{\Xi^2} \left\| \xi - \xi' \right\|^p \Pi\left(\mathrm{d}\xi, \mathrm{d}\xi'\right) : \begin{array}{l} \Pi \text{ is of joint distribution of } \xi, \xi' \\ \text{with marginal distribution } \mathbb{Q}_1, \mathbb{Q}_2 \end{array} \right\},
$$

where $p = 2$ leads to a Wasserstein distance of order two.

In Definition 2, D_W does not guarantee a real distance, it can be interpreted as the minimum transportation cost of moving the mass from $\mathbb{Q}_1$ to $\mathbb{Q}_2$. Thus, the ambiguity set at $t,\ t = 1, 2, \ldots, T$, can be defined as

$$
\mathcal{Q}_{\theta_t}(\mathbb{P}_{t,N_t}) = \{\mathbb{P}_t : D_W(\mathbb{P}_t, \mathbb{P}_{t,N_t}) \leq \theta_t\}, \tag{2}
$$

where $\mathcal{Q}_{\theta_t}\left(\mathbb{P}_{t,N_t}\right)$ containing all distributions can be viewed as the Wasserstein ball of radius θ_t centered at empirical distribution $\mathbb{P}_{t,N_t}$. Indeed, $\mathcal{Q}_{\theta_t}\left(\mathbb{P}_{t,N_t}\right)$ contains true distribution $\mathbb{P}_t$ with a high probability with mild situation [26].

For comparison, we further introduce the ambiguity set based on moment information from the study of Liu et al. [20]:

$$
\mathcal{P}_t = \left\{ P_t \in \mathcal{M}_t \;\middle|\; \begin{array}{l} \left(\mathbb{E}_{P_t}[\xi_t] - \mu_t^0\right)^{\top} \left(\Gamma_t^0\right)^{-1} \left(\mathbb{E}_{P_t}[\xi_t] - \mu_t^0\right) \leq k_t^1 \\ \mathrm{Cov}_{P_t}[\xi_t] \preceq k_t^2 \Gamma_t^0 \end{array} \right\}, t = 1, 2, \ldots, T,
$$

where $\mathcal{M}_t$ is the set of all probability distributions on the measure space $(\mathbb{R}^n, \mathcal{F}_t)$. μ_t^0 and Γ_t^0 are estimation values of the mean and the covariance matrix of ξ_t. $k_t^1, k_t^2 \in \mathbb{R}_+,\ t = 1, 2, \ldots, T$, control the size of the ambiguity set at time t. Obviously, the first constraint defines a ellipsoidal constraint centered on μ_t^0, that is, it constrains the true expectation of returns within a given range and the second, the size of true covariance of returns is also related to given Γ_t^0. Thus it can be seen, the Wasserstein ball which measures the distance between true distribution and empirical distribution of random returns, contains more informations than the ambiguity set built by Liu et al. [20].

However, the differences from previous models present some difficulties for our study. On the one hand, although Blanchet et al. [[44] came up with a new combination of single-period portfolio selection and Wasserstein distance, our multi-period extension still face some challenges in modeling. Especially, the objective and the connections between different periods should be reconsidered for multi-period case. On the other hand, minimizing the sum of variances in each period is much more difficult than minimizing the portfolio terminal variance. Technically, the DRO method leads to a more sophisticated problem because the min-max problem is non-convex and infinite dimensional. This prompted us to transform the model into a convex problem of finite dimensions. Moreover, the ambiguity set defined by the Wasserstein metric includes mass of moment information, which complicates the processing of model transformation.

3. Transformation of the multi-period DRMV model

In this section, we establish a reformulation of DRMV model since our initial problem is non-convex. Precisely, the objection problem of the model and the second constraint is related to the ambiguity set defined by Section 2. Thus, we should transform them into tractable reformulations. We first employ the dual theory to rewrite the second constraint of problem (1) in the following Lemma.

Lemma 1. *For $t = 1, 2, \ldots, T$, the second constraint in problem (1) is equivalent to*

$$u_t^\top \mathbb{E}_{\mathbb{P}_{t,N_t}}(\xi_t) - \sqrt{\theta_t}\|u_t\| \geq \sigma_t, \quad t = 1, 2, \ldots, T. \tag{3}$$

Proof. For $t = 1, 2, \ldots, T$, we consider the following problem

$$\min_{\mathbb{P}_t \in \mathcal{Q}_{\theta_t}(\mathbb{P}_{t,N_t})} \mathbb{E}_{\mathbb{P}_t}(w_t), \tag{4}$$

where $\mathcal{Q}_{\theta_t}\left(\mathbb{P}_{t,N_t}\right)$ is defined in (2). Note that (4) is equivalent to

$$-\max_{\mathbb{P}_t \in \mathcal{Q}_{\theta_t}(\mathbb{P}_{t,N_t})} \mathbb{E}_{\mathbb{P}_t}(-w_t).$$

According to $w_t = u_t^\top \xi_t$ in (1), for feasible $\mathbb{P}_t \in \mathcal{M}(\Xi_t)$ and support set $\Xi_t \in \mathbb{R}^n, \quad t = 1, \ldots, T$, we have

$$\begin{aligned}
&\max_{\mathbb{P}_t \in \mathcal{Q}_{\theta_t}(\mathbb{P}_{t,N_t})} \mathbb{E}_{\mathbb{P}_t}(-w_t) = \max_{\mathbb{P}_t \in \mathcal{Q}_{\theta_t}(\mathbb{P}_{t,N_t})} \mathbb{E}_{\mathbb{P}_t}\left[(-u_t)^\top \xi_t\right] \\
\Leftrightarrow &\begin{cases} \max\limits_{\Pi_t, \mathbb{P}_t} & \int_{\Xi_t} (-u_t)^\top \xi_t \mathbb{P}_t(\mathrm{d}\xi_t) \\ \text{s.t.} & \dfrac{1}{N_t}\sum\limits_{i=1}^{N_t} \int_{\Xi_t} \left\|\xi_t - \hat{\xi}_{t,i}\right\|^2 \Pi_t(\mathrm{d}\xi_t, \mathrm{d}\xi_t') \leq \theta_t \end{cases} \\
\Leftrightarrow &\begin{cases} \max\limits_{\mathbb{P}_t^i \in \mathcal{M}(\Xi_t)} & \dfrac{1}{N_t}\sum\limits_{i=1}^{N_t} \int_{\Xi_t} (-u_t)^\top \xi_t \mathbb{P}_t^i(\mathrm{d}\xi_t) \\ \text{s.t.} & \dfrac{1}{N_t}\sum\limits_{i=1}^{N_t} \int_{\Xi_t} \left\|\xi_t - \hat{\xi}_{t,i}\right\|^2 \mathbb{P}_t^i(\mathrm{d}\xi_t) \leq \theta_t, \end{cases}
\end{aligned} \tag{5}$$

where Π_t is a joint distribution of ξ_t and ξ_t', $t = 1, 2, \ldots, T$, with marginals $\mathbb{P}_t$ and $\mathbb{P}_{t,N_t}$, respectively, and the second equivalence follows from the law of total probability, which means that any joint probability distribution Π_t of ξ_t and ξ_t' can be constructed from the marginal distribution $\mathbb{P}_{t,N_t}$ of ξ_t' and the conditional distribution $\mathbb{P}_t^i$ of ξ_t given $\xi_t' = \hat{\xi}_{t,i}$, $t = 1, 2, \ldots, T,\ i \leq N_t$, which means that we may write $\Pi_t = \frac{1}{N_t}\sum_{i=1}^{N_t} \delta_{\hat{\xi}_{t,i}} \otimes \mathbb{P}_t^i$. The corresponding Lagrangian function for (5) is

$$L_t(\xi_t, \alpha_t) = \alpha_t\theta_t + \frac{1}{N_t}\sum_{i=1}^{N_t} \int_{\Xi_t} (-u_t)^\top \xi_t - \alpha_t \left\|\xi_t - \hat{\xi}_{t,i}\right\|^2 \mathbb{P}_t^i(\mathrm{d}\xi_t),$$

where α_t is the Lagrangian multiplier satisfying $\alpha_t \geq 0$.

In light of the proof of Theorem 4.2 in Mohajerin Esfahani and Kuhn [26], we let $\Delta_t = \xi_t - \hat{\xi}_{t,i}$ and the dual problem of (5) is

$$\begin{aligned}
&\begin{cases} \min\limits_{\alpha_t, R_t} \left\{\alpha_t\theta_t + \frac{1}{N_t}\sum\limits_{i=1}^{N_t} R_{t,i}\right\} \\ \text{s.t. } R_{t,i} \geq \max\limits_{\xi_t \in \Xi_t} \left\{(-u_t)^\top \xi_t - \alpha_t \left\|\xi_t - \hat{\xi}_{t,i}\right\|^2\right\} \quad i = 1, 2, \ldots, N_t, \\ \qquad \alpha_t \geq 0. \end{cases} \\
\Leftrightarrow \quad &\begin{cases} \min\limits_{\alpha_t, R_t} \left\{\alpha_t\theta_t + \frac{1}{N_t}\sum\limits_{i=1}^{N_t} R_{t,i}\right\} \\ \text{s.t. } R_{t,i} \geq \max\limits_{\xi_t \in \Xi_t} \left\{\|u_t\|\|\Delta_t\| - \alpha_t\|\Delta_t\|^2 - u_t^\top \hat{\xi}_{t,i}\right\} \quad i = 1, 2, \ldots, N_t, \\ \qquad \alpha_t \geq 0. \end{cases} \\
\Leftrightarrow \quad &\min\nolimits_{\alpha_t \geq 0} \left\{\alpha_t\theta_t + \frac{\|u_t\|^2}{4\alpha_t} - u_t^\top \mathbb{E}_{\mathbb{P}_{t,N_t}}(\xi_t)\right\} \\
&= \sqrt{\theta_t}\|u_t\| - u_t^\top \mathbb{E}_{\mathbb{P}_{t,N_t}}(\xi_t).
\end{aligned}$$

Therefore, we obtain (3). The proof is completed. □

In view of $w_t = u_t^\top \xi_t,\ t = 1, 2, \ldots, T$, in (1), we have

$$\mathrm{Var}_{\mathbb{P}_t}(w_t) = u_t^\top \mathbb{E}_{\mathbb{P}_t}\left(\xi_t \xi_t^\top\right) u_t - \left[u_t^\top \mathbb{E}_{\mathbb{P}_t}(\xi_t)\right]^2,\ t = 1, 2, \ldots, T.$$

We denote $\hat{\sigma}_t = u_t^\top \mathbb{E}_{\mathbb{P}_t}(\xi_t)$. Hence, we have $\hat{\sigma}_t \geq \sigma_t$ according to

$$\min_{\mathbb{P}_t \in \mathcal{Q}_{\theta_t}(\mathbb{P}_{t,N_t})} u_t^T \mathbb{E}_{\mathbb{P}_t}(\xi_t) \geq \sigma_t, \ \ t = 1, 2, \dots, T.$$

Then, substituting them into the objective function of problem (1), we obtain

$$\min_u \sum_{t=1}^{T} \left\{ \max_{\hat{\sigma}_t \geq \sigma_t} \left[\max_{\substack{\mathbb{P}_t \in \mathcal{Q}_{\theta_t}(\mathbb{P}_{t,N_t}) \\ \hat{\sigma}_t = \mathbb{E}_{\mathbb{P}_t}(u_t^\top \xi_t)}} \left\{ u_t^\top \mathbb{E}_{\mathbb{P}_t}\left(\xi \xi^\top\right) u_t \right\} - \hat{\sigma}_t^2 \right] \right\}. \tag{6}$$

By introducing $\hat{\sigma}_t = \mathbb{E}_{\mathbb{P}_t}\left(u_t^\top \xi_t\right)$, we just need to focus on how to solve the innermost maximum problem. We then provide a theorem to reformulate the primal problem (1) combining Lemma 1, whose proof is given in Appendix.

Theorem 1. *The primal* DRMV *problem equivalent to the following dual problem:*

$$\begin{array}{lll} \min_u & \sum_{t=1}^{T} \left\{ \sqrt{u_t^\top \operatorname{Var}_{\mathbb{P}_{t,N_t}}(\xi_t) u_t} + \sqrt{\theta_t} \|u_t\| \right\} & \\ \text{s.t.} & w_{t-1} = u_t^\top e & t = 1, 2, \dots, T, \\ & u_t^\top \mathbb{E}_{\mathbb{P}_{t,N_t}}(\xi_t) - \sqrt{\theta_t} \|u_t\| \geq \sigma_t & t = 1, 2, \dots, T, \\ & u_t \geq 0 & t = 1, 2, \dots, T, \end{array} \tag{7}$$

in the sense that the two problems have the same optimal solutions and optimal value, where $u = [u_1, u_2, \dots, u_T]^\top$.

We can easily see that the objective function $\sqrt{u_t^\top \operatorname{Var}_{\mathbb{P}_{t,N_t}}(\xi_t) u_t} + \sqrt{\theta_t}\|u_t\|$ and inequality constraint $u_t^\top \mathbb{E}_{\mathbb{P}_{t,N_t}}(\xi_t) - \sqrt{\theta_t}\|u_t\| \geq \sigma_t$ are convex. Moreover, the additional term of problem (7), $\sqrt{\theta_t}\|u_t\|$, is a reasonable approach to enhance the sparsity of portfolio weights, which is a way to reduce the variety of stocks in the portfolio and to solve the overfitting problem. Now, we add the second inequality constraint into the objective function for a more intuitive view of our proposed model. so we need to transform (7) to the following problem in Theorem 2.

Theorem 2. *For some constants* $\gamma_t > 0$, $\kappa_t > 0$, *we can obtain the following equivalent problem from (7):*

$$\begin{array}{lll} \min\limits_{u, \gamma, \kappa} & \sum_{t=1}^{T} \left\{ \gamma_t \sqrt{u_t^\top \operatorname{Var}_{\mathbb{P}_{t,N_t}}(\xi_t) u_t} - u_t^\top \mathbb{E}_{\mathbb{P}_{t,N_t}}(\xi_t) + \kappa_t \|u_t\| \right\} & \\ \text{s.t.} & w_{t-1} = u_t^\top e & t = 1, 2, \dots, T, \\ & u_t \geq 0 & t = 1, 2, \dots, T, \end{array} \tag{8}$$

where $\gamma = [\gamma_1, \gamma_2, \dots, \gamma_T]^\top$ *and* $\kappa = [\kappa_1, \kappa_2, \dots, \kappa_T]^\top$.

Proof. We just consider the problem including the objective function and the second constraint of (7). At moment t, for $t = 1, 2, \dots, T$, the corresponding Lagrangian function is

$$\begin{aligned} L_t(u_t, \lambda_t) &= \sqrt{u_t^\top \operatorname{Var}_{\mathbb{P}_{t,N_t}}(\xi_t) u_t} + \sqrt{\theta_t}\|u_t\| + \lambda_t\left(\sigma_t - u_t^\top \mathbb{E}_{\mathbb{P}_{t,N_t}}(\xi_t) + \sqrt{\theta_t}\|u_t\|\right) \\ &= \sqrt{u_t^\top \operatorname{Var}_{\mathbb{P}_{t,N_t}}(\xi_t) u_t} - \lambda_t u_t^\top \mathbb{E}_{\mathbb{P}_{t,N_t}}(\xi_t) + (1 + \lambda_t)\sqrt{\theta_t}\|u_t\| + \lambda_t \sigma_t, \end{aligned} \tag{9}$$

where $\lambda_t > 0$ is the Lagrangian multipliers. Since the problem (7) is convex, the optimal Lagrangian multipliers exist. Let λ_t^* be the optimal Lagrangian multipliers, and define $\gamma_t = \frac{1}{\lambda_t^*} > 0$, $\kappa_t = (1 + \gamma_t)\sqrt{\theta_t} > 0$. According to the KKT condition, the problem (7) is equivalent to the following Lagrangian problem

$$\begin{array}{lll} \min\limits_{u} & \sum_{t=1}^{T} \left\{ \gamma_t \sqrt{u_t^\top \operatorname{Var}_{\mathbb{P}_{t,N_t}}(\xi_t) u_t} - u_t^\top \mathbb{E}_{\mathbb{P}_{t,N_t}}(\xi_t) + \kappa_t \|u_t\| + \sigma_t \right\} & \\ \text{s.t.} & w_{t-1} = u_t^\top e & t = 1, 2, \dots, T, \\ & u_t \geq 0 & t = 1, 2, \dots, T. \end{array} \tag{10}$$

Ignore the constant term in the objective function, we obtain the conclusion and thus complete the proof. □

From Theorem 2, we can solve problem (7) by way of solving (8). Noticeable, the objective function (8) is equal to the sum of the portfolio risk as well as a holding cost term (see, e.g., [39,42]). The problem (8) is a second-order conic program and thus tractable, which can be efficiently solved by CPLEX and MOSEK solver.

4. Numerical results

We randomly choose 15 stocks ($n = 15$) from S&P 500 indices, **AMZN, BSX, CME, CVX, DIS, EL, EW, GE, IBM, NDAQ, PEP, PFE, SCHW, WMT** and **XRAY.** We used the adjusted weekly returns of these stocks for approximately 883 weeks from May 6, 2005, to April 1, 2022, for the numerical experiments. The historical data were separated into two parts: the in-sample

Table 1
Statistics of return rates with in-sample data.

	mean(10^{-2})	var(10^{-2})	Kurtosis	Skewness	$K-S$ Statistic*	P-value
AMZN	0.6559	0.2554	9.1045	0.8673	0.4519	0.0000
BSX	0.1555	0.197	9.5197	-0.2293	0.4451	0.0000
CME	0.4688	2.0052	519.452	20.1265	0.1438	0.0000
CVX	0.2228	0.1059	10.593	-0.7495	0.4693	0.0000
DIS	0.302	0.1065	7.3841	-0.0897	0.4659	0.0000
EL	0.2689	0.3819	85.0891	1.4097	0.3747	0.0000
EW	0.3476	0.4113	87.1916	1.8203	0.386	0.0000
GE	-0.0027	0.1801	12.9404	0.7613	0.4269	0.0000
IBM	0.1615	0.0896	6.3253	-0.2058	0.4521	0.0000
NDAQ	0.399	0.2452	7.5696	0.205	0.4265	0.0000
PEP	0.1942	0.0445	8.6919	-0.7272	0.468	0.0000
PFE	0.1718	0.0918	8.2336	-0.4472	0.4546	0.0000
SCHW	0.3268	0.1978	5.3478	0.1713	0.4564	0.0000
WMT	0.1994	0.067	5.9391	-0.4851	0.4845	0.0000
XRAY	0.0969	0.1439	44.5387	-3.1948	0.4482	0.0000

* : The critical value is 0.0489.

period data of 782 weeks ($N_{IN} = 782$), from May 6, 2005, to April 24, 2020, and the out-of-sample period data of 101 weeks ($N_{OUT} = 101$), from May 1, 2020, to April 1, 2022.

We divide the in-sample data of 782 weeks into 34 periods, i.e., $T_{IN} = 34$, for a total of 34 intervals of 23 weeks each, i.e., $N_t := 23, t = 1, 2, \ldots, T$. Investors make decisions at the beginning of each period and reallocated their wealth to continue investing at the end of each period, and each period consists of about six months.

4.1. In-sample analysis

For the proposed multi-period DRMV model, we need to choose two parameters: the Lagrangian multiplier γ_t and the radius θ_t of the Wasserstein ball. In particular, we let the same value of γ_t for each period in the proposed model to simplify the computation. For γ_t, we choose $\gamma_t = 0.15$ and $\hat{\gamma}_t = \gamma_t,\ t = 1, 2, \ldots, T_{IN}$, after debugging more than once and considering γ_t was greater than 0. For θ_t, we perform the bootstrap method proposed by Kang et al. [19]. This approach is intuitive and involves replotting historical observations with replacements. We assume that return rates were independent and homogeneously distributed but impose no other assumptions on the distribution. Moreover, it is crucial to consider the range of θ_t, a θ_t that is too small could have lost the meaning of robustness, and an excessively large θ_t could have led to an overly conservative strategy. The procedure for computing an optimal value of θ_t using the bootstrap method, for $t = 1, 2, \ldots, T_{IN}$, can be seen in Algorithm 1.

Algorithm 1 Bootstrap procedure for the radius θ_t of the Wasserstein ambiguity set.

Step 1. For each period t, ~$t = 1, 2, \ldots, T$, we select the tth historical return rates of N_t weeks for n assets as the discrete empirical distribution $\hat{\xi}_t$. To simulate the distributional uncertainty, we choose b potential true distribution $\xi_{t,1}, \xi_{t,2}, \ldots, \xi_{t,b}$ by randomly sampling b times with replacement from the weekly return rates on the empirical distribution $\hat{\xi}_t$.Next, we compute the Wasserstein distance $D_{W_{t,j}}$ between $\xi_{t,j}$ and $\hat{\xi}_t$ with $j = 1, 2, \ldots, b$.

Step 2. We reorganize the return rates in $\xi_{t,j}$ and $\hat{\xi}_t$ in ascending order for each asset, and denote the reorganized ones by $\tilde{\xi}_{t,j}$ and $\tilde{\hat{\xi}}_t$, $i = 1, 2, \ldots, b$, respectively. In the following, we assume that the probability $P_{D_{W_{t,j}}}$ of turning all the components of $\tilde{\xi}_{t,j}$ to that of $\tilde{\hat{\xi}}_t$ is the same, that is $P_{D_{W_{t,j}}} = P_{D_W} = \frac{1}{nN_t}$.Then, we have

$$D_{W_{t,j}} = \sum_{r=1}^{n} \sum_{s=1}^{N_t} \left[(\tilde{\xi}_{t,j})_{rs} - (\tilde{\hat{\xi}}_t)_{rs} \right]^2 P_{D_{W_{t,j}}}, \quad j = 1, 2, \ldots, b.$$

Thus, we obtain b Wasserstein distances.

Step 3. We choose the maximum of b distances to guarantee that the value of θ_t is not too small, i.e., the θ_t is chosen to be

$$\hat{\theta}_t = \max_j \left\{ D_{W_{t,j}} \right\}, \quad j = 1, 2, \ldots, b.$$

We compute the basic statistics of the return rates from 15 stocks using in-sample data, i.e., $n = 15$, as shown in Table 1. From a general view, it could be seen from kurtosis and skewness that the return rates are fat-tiled and not symmetrically distributed with large leptokurtosis. We find that for several stocks, the kurtosis is higher than that of other stocks. For example, **CME, EL, EW**, and **XRAY** have a higher kurtosis, implying the return rates are more likely to be extreme. We then

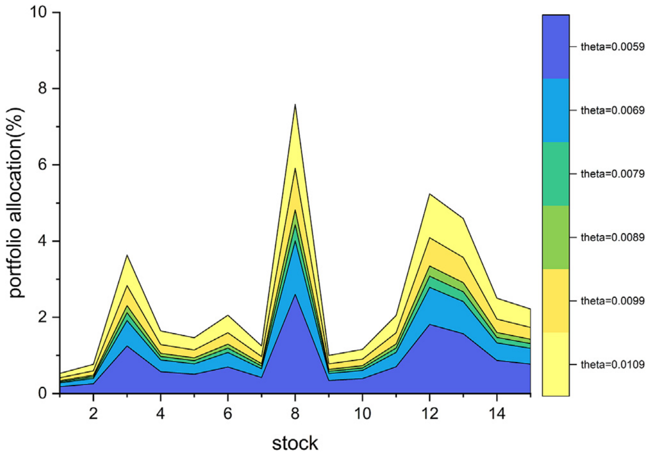

Fig. 1. the staked area of portfolio strategy under different θ.

perform *Kolmogorov – Smirnov* test in MATLAB. The null hypothesis is defined as H_0: The return rates of these stocks come from a normal distribution. The $K-S$ statistic and P-value in Table 1 show that the $K-S$ test rejected the hypothesis at the 5% significance level. Therefore, we conclude that the return rates with in-sample data are from a significantly non-normal distribution.

After estimating γ_t and θ_t, we further examine the specific impact of the changes in $\hat{\gamma}_t$ and $\hat{\theta}_t$. First, we consider the influence of $\hat{\theta}_t$ separately. we compute the average $\hat{\theta}_t$, with t is from moment 1 to T_{IN}, denoted as $\bar{\theta}$, and introduce the relative difference of portfolio strategy ($RD-$weight):

$$RD-\text{weight} = \frac{|u(\theta) - u(\bar{\theta})|}{u(\bar{\theta})} \times 100\%, \quad t = 1, 2, \ldots, T_{IN},$$

where θ is used for each moment t, that is $\theta = \theta_t,\ t = 1, \ldots, T$, and $u(\theta)$ and $u(\bar{\theta})$ are the portfolio weights obtained through θ and $\bar{\theta}$, respectively. Then, the range of θ is as follows,

$$\frac{|\theta - \bar{\theta}|}{\bar{\theta}} \leq 30\%. \tag{11}$$

Therefore, the value of θ varies in the range [0.0059,0.0109].

Figure 1 shows the change in ($RD-$weight) under different θ. we can see the portfolio strategy for 15 stocks performed differently after 34 periods. As a whole, the portfolio strategy of most stocks such as the first stock (**AMZN**) barely changes as θ changed except the eighth stock (**GE**). For some insights, we find that the stocks with particular performance in Table 1 also have significant changes in portfolio strategy. For example, the third stock (**CME**) which shows the largest kurtosis among the 15 stocks has a noticeable change in portfolio change. Additionally, the portfolio strategy of the eighth stock (**GE**) changes acutely, most likely because only **GE** has a negative mean in Table 1. This phenomenon further illustrates that extreme cases in stock returns also have significant impact on portfolio strategy. Hence, we next consider whether the parameters also have visible effects on the return and risk of the portfolio.

In addition, we denote $\bar{\gamma} = \hat{\gamma}_t$ because we previously made $\hat{\gamma}$ the same for each t, $t = 1, 2, \ldots, T_{IN}$. Moreover, we define four evaluation criteria: final wealth of portfolios (W_F), standard deviation (Std), Sharpe ratio ($Sharpe$) and maximum drawdown rate (MDR) of portfolio returns [45] which reflects the stability of portfolio returns. The specific formulas of these four criteria were as follows:

$$W_F = \sum_{i=1}^{T_{IN}} u^i_{T_{IN}} \mathbb{E}_{\mathbb{P}_{T_{IN},N_{T_{IN}}}}(\xi^i_{T_{IN}}),$$

$$Std = \sqrt{\frac{1}{T_{IN}-1}\sum_{t=1}^{T_{IN}}\left[\frac{u_t^\top \mathbb{E}_{\mathbb{P}_{t,N_t}}(\xi_t)}{\sum_{i=1}^n u_t^i} - \frac{1}{T_{IN}}\sum_{t=1}^{T_{IN}}\frac{u_t^\top \mathbb{E}_{\mathbb{P}_{t,N_t}}(\xi_t)}{\sum_{i=1}^n u_t^i}\right]^2},$$

$$Sharpe = \frac{\frac{1}{T_{IN}}\sum_{t=1}^{T_{IN}}\frac{u_t^\top \mathbb{E}_{\mathbb{P}_{t,N_t}}(\xi_t)}{\sum_{i=1}^n u_t^i}}{\text{Std}} = \frac{1}{T_{IN}}\sum_{t=1}^{T_{IN}}\frac{u_t^\top \mathbb{E}_{\mathbb{P}_{t,N_t}}(\xi_t)}{\text{Std}\sum_{i=1}^n u_t^i},$$

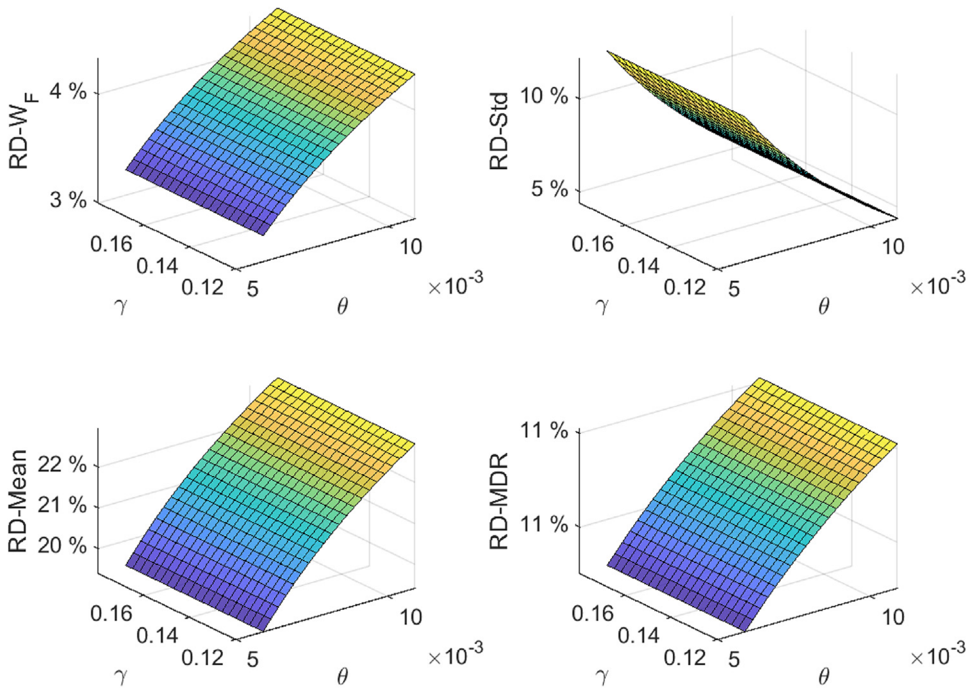

Fig. 2. The relative differences (RD) of Final wealth (W_F), Standard deviation (Std) of portfolio returns, Sharpe ratio ($Sharpe$) and Maximum drawdown rate (MDR) using multi-period DRMV model for varying θ and γ. The bootstrap sample $b = 2000$, the estimated parameters $\bar{\theta} = 0.0084$ and $\bar{\gamma} = 0.15$.

$$MDR = \max_{1<j<T_{IN}} \max_{1<k<j} \frac{w_j - w_k}{w_k},$$

where i denotes the ith stock of n stocks. w_j and w_k denote the corresponding cumulative wealth at moment j and k, respectively, and $W_F = w_{T_{IN}}$. To intuitively show the effect of θ and γ on these evaluation criteria, we use the relative differences (RD) introduced in Chen et al. [45]:

$$RD = \frac{|E(\gamma, \theta) - E(\bar{\gamma}, \bar{\theta})|}{E(\bar{\gamma}, \bar{\theta})} \times 100\% \quad t = 1, 2, \ldots, T_{IN},$$

where $E(\gamma, \theta)$ and $E(\bar{\gamma}, \bar{\theta})$ are the four evaluation criteria obtained by (γ, θ) and (γ, θ), respectively. The setting of θ remains the same as that in (11). Then, the range of γ is

$$\frac{|\gamma - \bar{\gamma}|}{\bar{\gamma}} \leq 20\%.$$

Because the estimated $\bar{\theta}$ is slightly small, we make the change in θ slightly larger than that of γ to make the impact of θ more visible on portfolios. In particular, the values of γ and θ varied in the range of $[0.1200, 0.1800] \times [0.0059, 0.0109]$.

Figure 2 shows the relative differences among the four evaluation criteria. We can see that the 20% and 30% changes in θ and γ lead to a change of 22% and 10% in the mean and standard deviation of our portfolio, but only lead to a change of 4% in final wealth (W_F). In other words, changes in the parameters do not have a significant impact on the final returns and risks of portfolios.

4.2. Out-of-sample analysis

For comparison, we use the historical data in rolling window procedure, following [45] and DeMiguel et al. [41]. First, we choose a window of length N_{IN} over which to perform the estimation and defined the prediction period as N_{OUT}, where $N = N_{IN} + N_{OUT}$ denotes the total number of data, which is $N = 883$. Second, we use the first N_{IN} dataset to estimate of the optimal portfolio strategy through T_{IN} investments and apply this strategy on the first prediction window ρ ($\rho < N_{OUT}$). Third, we repeat this procedure by dropping the earliest ρ in the first estimation window and adding a new ρ at the end of the first estimation window. These processes are repeated $H = \frac{N_{OUT}}{\rho}$ times; therefore, we have H optimal portfolios for 15 stocks (Fig. 3).

To clearly demonstrate the favorable performance of our models, we compare the models we introduced, multi-period mean-variance model with known distribution(MV), Olivares-Nadal-DeMiguel Model(ODM) [39] and the two-stage stochastic convex programming model (Two-Stage) [40]. Specific descriptions of the three models are provided below for comparison:

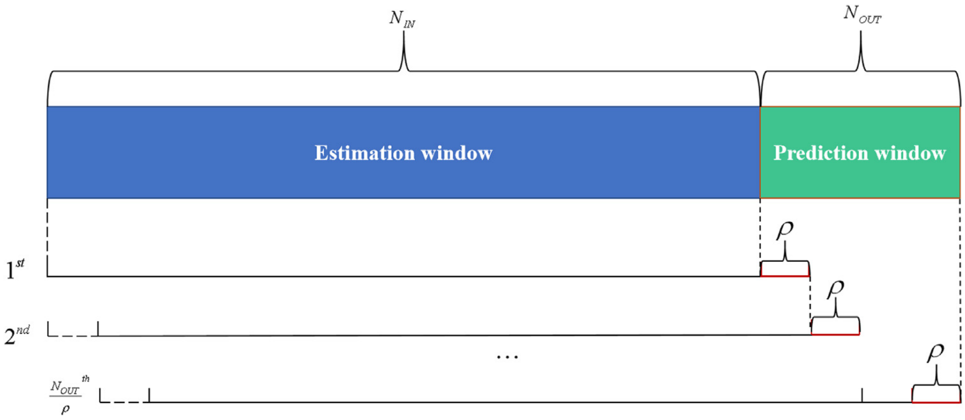

Fig. 3. Schematic diagram of Rolling window method.

(i) The multi-period mean-variance model is obtained when Wasserstein distance of our multi-period DRMV model is equal to 0 ($\theta_t = 0, t = 1, 2, \dots, T$).

(ii) Olivares-Nadal-DeMiguel Model is equals to a mean-variance model with a quadratic transaction cost

$$\gamma_{ODM} u^\top \Sigma u + \kappa \left\| \Sigma^{\frac{1}{2}} (u - u_0) \right\|_2^2,$$

where $\Sigma = \mathrm{var}_{\mathbb{P}}(\xi)$ and the optimal γ_{ODM} was set for comparison. As described in 3.1 section in DeMiguel and Olivares-Nadal [39], we separated the in-sample data from 10 parts (as uniformly as possible) for 10-fold-cross method to estimate the optimal parameter τ. We selected τ from the set $\{0\%, 0.5\%, 1\%, 2.5\%, 5\%, 10\%\}$ by choosing the minimum variance of portfolio returns using 10-fold-cross method, and the coefficient of transaction cost κ was calculated using τ. Thus, we obtained the optimal portfolio weights based on the algorithm proposed by Blanchet et al. [44].

(iii) The two-stage stochastic convex programming model that we compared is of minimizing the sum of a cost function and an expectation of the linear function, which can be read as

$$\min_u c(u_t) + \sup_{P \in \mathcal{P}} \mathbb{E}_P[Q(u_t, \xi_t)], \quad t = 1, 2, \dots, T,$$

where the first stage $c(u_t)$ is the risk of the investment and the second is the return with uncertainty. We use the estimated value of covariance matrix of ξ_t for the first known part, and take the conic optimization problem which is equivalent to the second part of expectation above proved in Theorem 1 of Li et al. [40] for the uncertainty. Moreover, the moment information of random vector ξ_t is estimated by the Monte Carlo sampling procedure described in Li et al. [40]. Ultimately, a set of solutions of resulting conic optimization problem is solved by SeDuMi which is a software for optimization over symmetric cones.

Considering the weekly data of n stocks in out-of-sample data is $\xi_i = \left(\xi_i^1, \xi_i^2, \dots, \xi_i^n\right)^\top$, $i = 1, 2, \dots, H$, where the out-of-sample data contains H prediction windows that consists of ρ samples. After H times of rolling, the estimated portfolio strategy is $u_i = \left(u_i^1, u_i^2, \dots, u_i^n\right)^\top$, $i = 1, 2, \dots, H$. At the beginning of the first rolling, we set the initial wealth as $w_0 = 1$, and the investor cumulative wealth as

$$w_i = \frac{w_{i-1} u_i^\top \xi_i}{\sum_{k=1}^n u_i^k}, \quad i = 1, 2, \dots, H.$$

In the following, we redefine the previous evaluation criteria based on rolling window method as prediction of criteria: final wealth (W_F^p) of portfolios, weighted average (μ^p), standard deviation (Std^p) and Sharpe ratio ($Sharpe^p$) of portfolio returns (we assume the risk-free interest rate is 0).

$$\begin{aligned}
W_F^p &= u_H{}^\top \mathbb{E}_{\mathbb{P}_{H,N_H}}(\xi_H), \\
\mu^p &= \frac{1}{H} \sum_{i=1}^H \frac{u_i^\top \mathbb{E}_{\mathbb{P}_{i,N_i}}(\xi_i)}{\sum_{k=1}^n u_i^k}, \\
Std^p &= \sqrt{\frac{1}{H-1} \sum_{i=1}^H \left[\frac{u_i^\top \mathbb{E}_{\mathbb{P}_{i,N_i}}(\xi_i)}{\sum_{k=1}^n u_i^k} - \mu^p \right]^2}, \\
Sharpe^p &= \frac{\mu^p}{Std^p}.
\end{aligned}$$

Table 2
Statistics of out-of-sample return series got under four models.

	$W_F{}^p$	$Mean^p(10^{-3})$	$Std^p(10^{-2})$	$Sharpe^p$
DRMV	**1.6900**	**5.4472**	**2.1966**	**2.4922**
MV($\theta_t = 0$)	1.5714	4.7279	2.2198	2.1405
ODM	1.5140	4.3571	2.2127	1.9790
Two-Stage	1.5625	4.6708	2.2133	2.1208

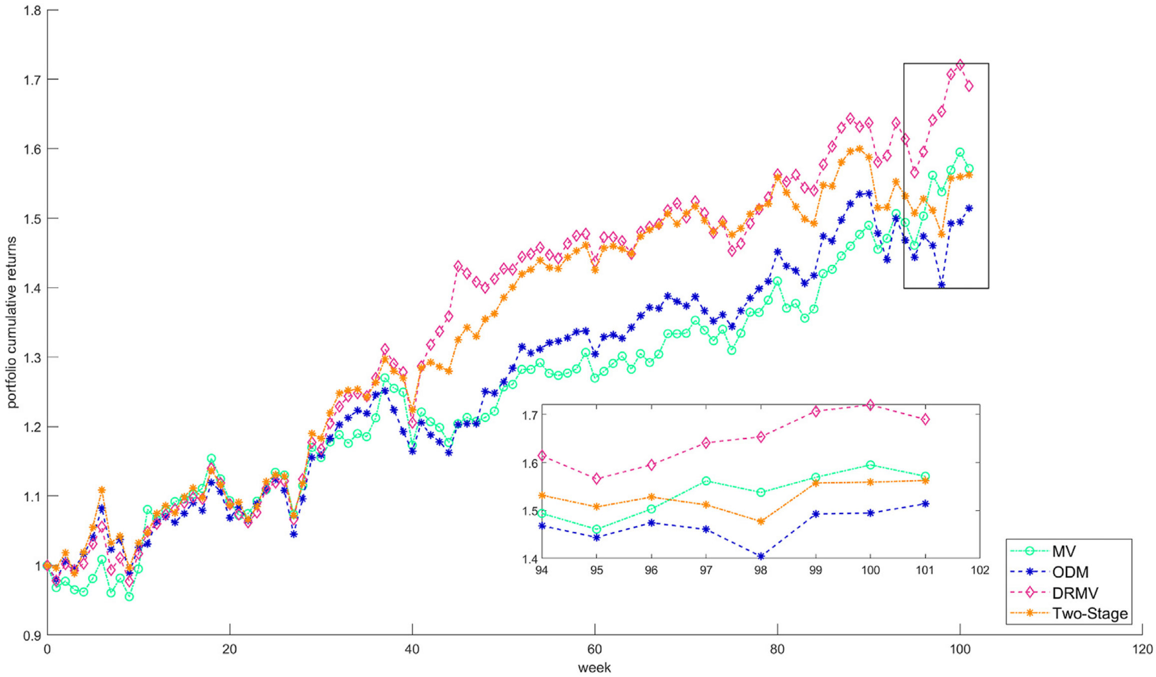

Fig. 4. The cumulative wealth of four models during 101 out-of-sample weeks and partial enlargement from 94 to 101 weeks.

For the above four models, we estimate the corresponding four criteria according to rolling window procedure. We set $\rho = 1$, meaning that the length of a prediction window for each rolling was one week. Table 2 shows the statistics obtained from out-of-sample data for the four models. We can clearly see that the DRMV model performs better than the other three models for the final wealth, mean and standard deviation. In other words, our model has the highest return and lowest risk of portfolio selection among four models, which also demonstrates its robustness to risk control and superiority in portfolio returns. The DRMV model had the highest Sharpe ration among the four models. The Sharpe ratio is a return measure often used to compare the performance of managers by adjusting for risk, and a higher Sharpe ratio represents a higher excess return with a fixed risk.

In contrast, we can see in Fig. 4 that the cumulative returns of the four models are consistent in the first 30 weeks. However, the DRMV trend increases gradually and shows higher returns than the MV, ODM and Two-Stage models. Hence, we conclude that, in the long term, the cumulative returns of DRMV model perform better and demonstrate greater stability than those of the other three models.

5. Conclusion

In this study, we established a new multi-period distributionally robust mean-variance model with the Wasserstein metric for portfolio selection and derived its dual problem for tractability. Although the Wasserstein DRO technique is widely applied to many uncertainty problems, the combination of the Wasserstein DRO and multi-period mean-variance model is relatively new. A series of empirical experiments demonstrated that the model is robust and efficient. Indeed, we did not intend to prove that our model performs best when compared with others, but instead that it could provide guidance for practitioners on medium- and long-term investments based on the DRO approach. Nevertheless, our study has some limitations, such as the fact that our algorithm is not efficient enough, which leads to long solution times in the multi-period condition. Our future research will focus on more practical constraints and the effective algorithm.

Data Availability

Data will be made available on request.

Acknowledgments

This work was partly supported by the National Natural Science Foundation of China (No. 12001286), the Project funded by China Postdoctoral Science Foundation (No. 2022M711672), and the Philosophy and Social Science Foundation of the Jiangsu Higher Education Institutions of China (No. 2020SJA0163).

Appendix

The proof of Theorem 1

Next, we need to solve the following problem firstly

$$\begin{cases} \max\limits_{\mathbb{P}_t \in \mathcal{Q}_{\theta_t}(\mathbb{P}_{t,N_t})} & u_t^\top \mathbb{E}_{\mathbb{P}_t}\left(\xi_t \xi_t^\top\right) u_t \\ \text{s.t.} & u_t^\top \mathbb{E}_{\mathbb{P}_t}(\xi_t) = \hat{\sigma}_t. \end{cases} \tag{12}$$

By the definition of $\mathcal{Q}_{\theta_t}$ in (2), we can transform (12) into the following problem

$$\begin{cases} \max\limits_{\mathbb{P}_t^i \in \mathcal{M}(\Xi_t)} & \frac{1}{N_t}\sum_{i=1}^{N_t} \int_{\Xi_t} \left(u_t^\top \xi_t\right)^2 \mathbb{P}_t^i(d\xi_t) \\ \text{s.t.} & \frac{1}{N_t}\sum_{i=1}^{N_t} \int_{\Xi_t} u_t^\top \xi_t \mathbb{P}_t^i(d\xi_t) = \hat{\sigma}_t, \\ & \frac{1}{N_t}\sum_{i=1}^{N_t} \int_{\Xi_t} \left\|\xi_t - \hat{\xi}_{t,i}\right\|^2 \mathbb{P}_t^i(d\xi_t) \le \theta_t. \end{cases} \tag{13}$$

Correspondingly, the Lagrangian dual function of (13) is

$$\begin{aligned} &\max_{\mathbb{P}_t^i \in \mathcal{M}(\Xi_t)} \min_{\gamma_t \in \mathbb{R}, \lambda_t \ge 0} L_t(\xi_t, \gamma_t, \lambda_t) \\ &= \max_{\mathbb{P}_t^i \in \mathcal{M}(\Xi_t)} \min_{\gamma_t \in \mathbb{R}, \lambda_t \ge 0} \\ &\left\{ \frac{1}{N_t}\sum_{i=1}^{N_t} \int_{\Xi_t} \left(u_t^\top \xi_t\right)^2 - \gamma_t u_t^\top \xi_t - \lambda_t \|\xi_t - \hat{\xi}_{t,i}\|^2 \mathbb{P}_t^i(d\xi_t) + \gamma_t \hat{\sigma}_t + \lambda_t \theta_t \right\} \\ &\le \min_{\gamma_t \in \mathbb{R}, \lambda_t \ge 0} \{\gamma_t \hat{\sigma}_t + \lambda_t \theta_t \\ &+ \frac{1}{N_t}\sum_{i=1}^{N_t} \max_{\mathbb{P}_t^i \in \mathcal{M}(\Xi_t)} \left\{ \int_{\Xi_t} \left(u_t^\top \xi_t\right)^2 - \gamma_t u_t^\top \xi_t - \lambda_t \|\xi_t - \hat{\xi}_{t,i}\|^2 \mathbb{P}_t^i(d\xi_t) \right\} \Bigg\} \\ &= \min_{\gamma_t \in \mathbb{R}, \lambda_t \ge 0} \left\{ \gamma_t \hat{\sigma}_t + \lambda_t \theta_t + \frac{1}{N_t}\sum_{i=1}^{N_t} S_{t,i} \right\}, \end{aligned} \tag{14}$$

where $S_{t,i} \ge \max\limits_{\xi_t \in \Xi_t} \left\{ \left(u_t^\top \xi_t\right)^2 - \gamma_t u_t^\top \xi_t - \lambda_t \left\|\xi_t - \hat{\xi}_{t,i}\right\|^2 \right\}$.

Next, we need to cope with the maximization problem as follows

$$\max_{\xi_t \in \Xi_t} \left\{ \left(u_t^\top \xi_t\right)^2 - \gamma_t u_t^\top \xi_t - \lambda_t \left\|\xi_t - \hat{\xi}_{t,i}\right\|^2 \right\}.$$

Let $\Delta_t = \xi_t - \hat{\xi}_{t,i}$, we have

$$
\begin{aligned}
&\max_{\xi_t \in \Xi_t} \left[\left(u_t^\top \xi_t\right)^2 - \gamma_t u_t^\top \xi_t - \lambda_t \left\| \xi_t - \hat{\xi}_{t,i} \right\|^2 \right] \\
&= \max_{\Delta_t} \left[\|u_t\|^2 \|\Delta_t\|^2 + \left(u_t^\top \hat{\xi}_{t,i}\right)^2 \right. \\
&\left. +2\left|u_t^\top \hat{\xi}_{t,i}\right| \|u_t\| \|\Delta_t\| - |\gamma_t| \|u_t\| \|\Delta_t\| - \gamma_t u_t^\top \hat{\xi}_{t,i} - \lambda_t \|\Delta_t\|^2 \right] \\
&\leq \left(u_t^\top \hat{\xi}_{t,i}\right)^2 - \gamma_t u_t^\top \hat{\xi}_{t,i} \\
&+ \max_{\Delta_t} \left[\left(\|u_t\|^2 - \lambda_t\right) \|\Delta_t\|^2 + \left| 2u_t^\top \hat{\xi}_{t,i} - \gamma_t \right| \|u_t\| \|\Delta_t\| \right],
\end{aligned}
$$

where the inequality follows from $|a| - |b| \leq |a - b|$ for any real numbers a and b. We split the above problem into four cases as follows:

(a) if $\|u_t\|^2 > \lambda_t$, $\quad S_{t,i} = +\infty$;
(b) if $\|u_t\|^2 = \lambda_t$ and $2u_t^\top \hat{\xi}_{t,i} \neq \gamma_t$, $\quad S_{t,i} = +\infty$;
(c) if $\|u_t\|^2 = \lambda_t$ and $2u_t^\top \hat{\xi}_{t,i} = \gamma_t$, $\quad S_{t,i} \geq -\left(u_t^\top \hat{\xi}_{t,i}\right)^2 = -\frac{1}{4}\gamma_t^2$;
(d) if $\|u_t\|^2 < \lambda_t$, $\quad S_{t,i} \geq \left(u_t^\top \hat{\xi}_{t,i}\right)^2 - \gamma_t u_t^\top \xi_{t,i} + \frac{(2u_t^\top \xi_{t,i} - \gamma_t)^2 \|u_t\|^2}{4\left(\lambda_t - \|u_t\|^2\right)}$.

We can see that $S_{t,i}$ is equal to $+\infty$ in both cases (a) and (b), so we only need to consider the latter two cases for problem (14).

In the following, we first discuss the case (c). In this case, according to the above discussions, problem (14) can be further reformulated as

$$
\begin{aligned}
&\min_{\gamma_t, \lambda_t \geq 0} \left\{ \gamma_t \hat{\sigma}_t + \lambda_t \theta_t + \tfrac{1}{N_t} \sum_{i=1}^{N_t} S_{t,i} \right\} \\
= &\min_{\gamma_t = 2u_t^\top \hat{\xi}_{t,i}, \lambda_t = \|u_t\|^2} \left\{ \tfrac{1}{N_t} \sum_{i=1}^{N_t} \left[-\left(u_t^\top \hat{\xi}_{t,i}\right)^2 \right] + \gamma_t \hat{\sigma}_t + \lambda_t \theta_t \right\} \\
= &\min_{\gamma_t} \left\{ -\tfrac{1}{4}\gamma_t^2 + \gamma_t \hat{\sigma}_t + \|u_t\|^2 \theta_t \right\}.
\end{aligned}
$$

We can see that the optimal value of the above problem is $-\infty$. Thus, (d) is nontrivial. In this case, problem (14) can be reformulated as

$$
\begin{aligned}
&\min_{\gamma_t, \lambda_t \geq 0} \left\{ \frac{1}{N_t} \sum_{i=1}^{N_t} S_{t,i} + \gamma_t \hat{\sigma}_t + \lambda_t \theta_t \right\} \\
&= \min_{\gamma_t, \lambda_t > \|u_t\|^2} \left\{ \frac{1}{N_t} \sum_{i=1}^{N_t} \left[\left(u_t^\top \hat{\xi}_{t,i}\right)^2 - \gamma_t u_t^\top \hat{\xi}_{t,i} + \frac{\left(2u_t^\top \hat{\xi}_{t,i} - \gamma_t\right)^2 \|u_t\|^2}{4\left(\lambda_t - \|u_t\|^2\right)} \right] + \gamma_t \hat{\sigma}_t + \lambda_t \theta_t \right\}.
\end{aligned}
\tag{15}
$$

Let

$$
H_t = \frac{1}{N_t} \sum_{i=1}^{N_t} \left[\left(u_t^\top \hat{\xi}_{t,i}\right)^2 - \gamma_t u_t^\top \hat{\xi}_{t,i} + \frac{\left(2u_t^\top \hat{\xi}_{t,i} - \gamma_t\right)^2 \|u_t\|^2}{4\left(\lambda_t - \|u_t\|^2\right)} \right] + \gamma_t \hat{\sigma}_t + \lambda_t \theta_t.
$$

Let the partial derivative in terms of γ_t be zero, we get

$$
\frac{\partial H_t}{\partial \gamma_t} = \hat{\sigma}_t - \frac{1}{N_t} \sum_{i=1}^{N_t} \left[\left(u_t^\top \hat{\xi}_{t,i}\right) + \frac{\left(2u_t^\top \hat{\xi}_{t,i} - \gamma_t\right) \|u_t\|^2}{2\left(\lambda_t - \|u_t\|^2\right)} \right] = 0,
$$

which implies

$$
\gamma_t = 2\hat{\sigma}_t - 2C_t \frac{\lambda_t}{\|u_t\|^2}, \tag{16}
$$

where $C_t = \hat{\sigma}_t - u_t^\top \mathbb{E}_{\mathbb{P}_{t,N_t}}(\xi_t)$ and $\|u_t\|^2 > 0$. Furthermore, γ_t is optimal since we have $\frac{\partial^2 H_t}{\partial \gamma_t^2} > 0$, then we substitute (16) into (15) and get

$$\begin{aligned}
&\min_{\gamma_t,\lambda_t\geq 0}\left\{\frac{1}{N_t}\sum_{i=1}^{N_t} S_{t,i} + \gamma_t\hat{\sigma}_t + \lambda_t\theta_t\right\}\\
=\;& \frac{1}{N_t}\sum_{i=1}^{N_t}\left(u_t^\top\hat{\xi}_{t,i}\right)^2 + \min_{\gamma_t,\lambda_t\geq 0}\left\{\frac{1}{N_t}\sum_{i=1}^{N_t}\frac{\left(2u_t^\top\hat{\xi}_{t,i}-\gamma_t\right)^2\|u_t\|^2}{4\left(\lambda_t-\|u_t\|^2\right)} + \gamma_t C_t + \lambda_t\theta_t\right\}\\
=\;& \frac{1}{N_t}\sum_{i=1}^{N_t}\left(u_t^\top\hat{\xi}_{t,i}\right)^2\\
&+ \min_{\lambda_t>\|u_t\|^2}\left\{\frac{1}{N_t}\sum_{i=1}^{N_t}\frac{\left(u_t^\top\hat{\xi}_{t,i}-\hat{\sigma}_t+C_t\frac{\lambda_t}{\|u_t\|^2}\right)^2\|u_t\|^2}{\lambda_t-\|u_t\|^2} + 2C_t\left(\hat{\sigma}_t - C_t\frac{\lambda_t}{\|u_t\|^2}\right) + \lambda_t\theta_t\right\}.
\end{aligned}$$

We introduce the slack variable ε_t that satisfies $\varepsilon_t \geq 0$ and let $\lambda_t = \|u_t\|^2 + \varepsilon_t$, we obtain

$$\begin{aligned}
&\min_{\gamma_t,\lambda_t\geq 0}\left\{\frac{1}{N_t}\sum_{i=1}^{N_t} S_{t,i} + \gamma_t\hat{\sigma}_t + \lambda_t\theta_t\right\}\\
=&\frac{1}{N_t}\sum_{i=1}^{N_t}\left(u_t^\top\hat{\xi}_{t,i}\right)^2 + \min_{\varepsilon_t\geq 0}\left\{\frac{1}{N_t}\sum_{i=1}^{N_t}\frac{\left[u_t^\top\hat{\xi}_{t,i}-\hat{\sigma}_t+C_t\left(1+\frac{\varepsilon_t}{\|u_t\|^2}\right)\right]^2\|u_t\|^2}{\varepsilon_t}\right.\\
&\left.+2C_t\left[\hat{\sigma}_t - C_t\left(1+\frac{\varepsilon_t}{\|u_t\|^2}\right)\right] + \theta_t\left(\|u_t\|^2+\varepsilon_t\right)\right\}\\
=&\frac{1}{N_t}\sum_{i=1}^{N_t}\left(u_t^\top\hat{\xi}_{t,i}\right)^2 + \theta_t\|u_t\|^2 + 2C_t u_t^\top\mathbb{E}_{\mathbb{P}_{t,N_t}}(\xi_t)\\
&+\min_{\varepsilon_t\geq 0}\left\{\frac{1}{N_t}\sum_{i=1}^{N_t}\frac{\left[\left(u_t^\top\hat{\xi}_{t,i}\right)^2-\left(u_t^\top\mathbb{E}_{\mathbb{P}_{t,N_t}}(\xi_t)\right)^2\right]\|u_t\|^2}{\varepsilon_t} + \left(\theta_t - \frac{C_t^2}{\|u_t\|^2}\right)\varepsilon_t\right\}.
\end{aligned}\tag{17}$$

If $\theta_t - \frac{C_t^2}{\|u_t\|^2} < 0$, the optimal value of the precedent problem is $-\infty$, which means problem (14) is infeasible, if $\theta_t - \frac{C_t^2}{\|u_t\|^2} \geq 0$, then the optimal value of problem (17) is

$$\begin{aligned}
&\frac{1}{N_t}\sum_{i=1}^{N_t}\left(u_t^\top\hat{\xi}_{t,i}\right)^2 + \theta_t\|u_t\|^2 + 2\mathbb{E}_{\mathbb{P}_{t,N_t}}(\xi_t)u_t^\top C_t\\
&+2\sqrt{\left(\theta_t - C_t^2/\|u_t\|^2\right)}\sqrt{\frac{1}{N_t}\sum_{i=1}^{N_t}\left[\left(u_t^\top\hat{\xi}_{t,i}\right)^2 - u_t\mathbb{E}_{\mathbb{P}_{t,N_t}}(\xi_t)^2\right]\|u_t\|^2}\\
=\;& \mathbb{E}_{\mathbb{P}_{t,N_t}}\left(u_t^\top\xi_t\right)^2 + \theta_t\|u_t\|^2 + 2\mathbb{E}_{\mathbb{P}_{t,N_t}}(\xi_t)u_t^\top\left[\hat{\sigma}_t - u_t^\top\mathbb{E}_{\mathbb{P}_{t,N_t}}(\xi_t)\right]+\\
&2\sqrt{u_t^\top \mathrm{Var}_{\mathbb{P}_{t,N_t}}(\xi_t)u_t}\sqrt{\theta_t\|u_t\|^2 - \left(\hat{\sigma}_t - u_t^\top\mathbb{E}_{\mathbb{P}_{t,N_t}}(\xi_t)\right)^2}.
\end{aligned}$$

Denote

$$\begin{aligned}
h_t\left(\hat{\sigma}_t,u_t\right) =\;& \mathbb{E}_{\mathbb{P}_{t,N_t}}\left(u_t^\top\xi_t\right)^2 + \|u_t\|^2\theta_t + 2\mathbb{E}_{\mathbb{P}_{t,N_t}}(\xi_t)u_t^\top\left[\hat{\sigma}_t - u_t^\top\mathbb{E}_{\mathbb{P}_{t,N_t}}(\xi_t)\right]\\
&+2\sqrt{u_t^\top \mathrm{Var}_{\mathbb{P}_{t,N_t}}(\xi_t)u_t}\sqrt{\theta_t\|u_t\|^2 - \left(\hat{\sigma}_t - u_t^\top\mathbb{E}_{\mathbb{P}_{t,N_t}}(\xi_t)\right)^2}.
\end{aligned}$$

Then, we have

$$\begin{aligned}
&h_t\left(\hat{\sigma}_t,u_t\right) - \hat{\sigma}_t^2\\
&= \mathbb{E}_{\mathbb{P}_{t,N_t}}\left(u_t^\top\xi_t\right)^2 + \|u_t\|^2\theta_t - \hat{\sigma}_t^2 + 2\mathbb{E}_{\mathbb{P}_{t,N_t}}(\xi_t)u_t^\top\left[\hat{\sigma}_t - u_t^\top\mathbb{E}_{\mathbb{P}_{t,N_t}}(\xi_t)\right]\\
&\quad+2\sqrt{u_t^\top \mathrm{Var}_{\mathbb{P}_{t,N_t}}(\xi_t)u_t}\sqrt{\theta_t\|u_t\|^2 - \left(\hat{\sigma}_t - u_t^\top\mathbb{E}_{\mathbb{P}_{t,N_t}}(\xi_t)\right)^2}\\
&= u_t^\top\left[\mathbb{E}_{\mathbb{P}_{t,N_t}}\left(\xi_t\xi_t^\top\right) - \left[\mathbb{E}_{\mathbb{P}_{t,N_t}}(\xi_t)\right]^2\right]u_t + \|u_t\|^2\theta_t - \left(\hat{\sigma}_t - u_t^\top\mathbb{E}_{\mathbb{P}_{t,N_t}}(\xi_t)\right)^2\\
&\quad+2\sqrt{u_t^\top \mathrm{Var}_{\mathbb{P}_{t,N_t}}(\xi_t)u_t}\sqrt{\theta_t\|u_t\|^2 - \left(\hat{\sigma}_t - u_t^\top\mathbb{E}_{\mathbb{P}_{t,N_t}}(\xi_t)\right)^2}\\
&= \left(\sqrt{u_t^\top \mathrm{Var}_{\mathbb{P}_{t,N_t}}(\xi_t)u_t} + \sqrt{\|u_t\|^2\theta_t - \left(\hat{\sigma}_t - u_t^\top\mathbb{E}_{\mathbb{P}_{t,N_t}}(\xi_t)\right)^2}\right)^2.
\end{aligned}$$

Based on the above relations and discussions, we can reformulate the problem (6) as follows

$$
\begin{aligned}
&\min_{u} \sum_{t=1}^{T} \left\{ \max_{\hat{\sigma}_t \geq \sigma_t} \left[\max_{\substack{\mathbb{P}_t \in \mathcal{Q}_{\theta_t}\left(\mathbb{P}_{t,N_t}\right) \\ \mathbb{E}_{\mathbb{P}_t}\left(u_t^\top \xi_t\right) = \hat{\sigma}_t}} \left\{ u_t^\top \mathbb{E}_{\mathbb{P}_t}\left(\xi \xi^\top\right) u_t \right\} - \hat{\sigma}_t^2 \right] \right\} \\
&= \min_{u} \sum_{t=1}^{T} \left\{ \max_{\hat{\sigma}_t \geq \sigma_t} \left\{ h_t\left(\hat{\sigma}_t, u_t\right) - \hat{\sigma}_t^2 \right\} \right\} \\
&= \min_{u} \sum_{t=1}^{T} \left\{ \max_{\hat{\sigma}_t \geq \sigma_t} \left\{ \left(\sqrt{u_t^\top \mathrm{Var}_{\mathbb{P}_{t,N_t}}\left(\xi_t\right) u_t} + \sqrt{\|u_t\|^2 \theta_t - \left(\hat{\sigma}_t - u_t^\top \mathbb{E}_{\mathbb{P}_{t,N_t}}\left(\xi_t\right)\right)^2} \right)^2 \right\} \right\} \\
&= \min_{u} \sum_{t=1}^{T} \left(\sqrt{u_t^\top \mathrm{Var}_{\mathbb{P}_{t,N_t}}\left(\xi_t\right) u_t} + \sqrt{\|u_t\|^2 \theta_t} \right)^2 .
\end{aligned}
\tag{18}
$$

where the last equality holds when $\hat{\sigma}_t = u_t^\top \mathbb{E}_{\mathbb{P}_{t,N_t}}(\xi_t) \geq \sigma_t$, which is optimal for $\hat{\sigma}_t$. In addition, the optimal solution of (18) is the same to that of its square root form. This completes the proof of Theorem 1.

References

[1] H. Markowitz, The Optimization of a Quadratic Function Subject to Linear Constraints, Technical Report, RAND CORP SANTA MONICA CA, 1955.
[2] N.L. Jacob, A limited-diversification portfolio selection model for the small investor, J. Finance 29 (3) (1974) 847–856.
[3] A.F. Perold, Large-scale portfolio optimization, Manage. Sci. 30 (10) (1984) 1143–1160.
[4] T.-J. Chang, N. Meade, J.E. Beasley, Y.M. Sharaiha, Heuristics for cardinality constrained portfolio optimisation, Comput. Oper. Res. 27 (13) (2000) 1271–1302.
[5] D. Bertsimas, R. Cory-Wright, A scalable algorithm for sparse portfolio selection, INFORMS J. Comput. 34 (3) (2022) 1489–1511.
[6] X. Cui, J. Gao, X. Li, Y. Shi, Survey on multi-period mean–variance portfolio selection model, J. Oper. Res. Soc.China (2022) 1–24.
[7] D. Li, W.-L. Ng, Optimal dynamic portfolio selection: multiperiod mean-variance formulation, Math. Finance 10 (3) (2000) 387–406.
[8] X. Cui, J. Gao, X. Li, D. Li, Optimal multi-period mean–variance policy under no-shorting constraint, Eur. J. Oper. Res. 234 (2) (2014) 459–468.
[9] X. Cui, X. Li, D. Li, Unified framework of mean-field formulations for optimal multi-period mean-variance portfolio selection, IEEE Trans. Automat. Control 59 (7) (2014) 1833–1844.
[10] H. Xiao, Z. Zhou, T. Ren, Y. Bai, W. Liu, Time-consistent strategies for multi-period mean-variance portfolio optimization with the serially correlated returns, Commun. Stat.-Theory Methods 49 (12) (2020) 2831–2868.
[11] S. Corsaro, V. De Simone, Z. Marino, Fused LASSO approach in portfolio selection, Ann. Oper. Res. 299 (1) (2021) 47–59.
[12] X. Li, A.S. Uysal, J.M. Mulvey, Multi-period portfolio optimization using model predictive control with mean-variance and risk parity frameworks, Eur. J. Oper. Res. 299 (3) (2022) 1158–1176.
[13] B. Li, Y. Zhu, Y. Sun, G. Aw, K.L. Teo, Multi-period portfolio selection problem under uncertain environment with bankruptcy constraint, Appl. Math. Model. 56 (2018) 539–550.
[14] B. Li, K.L. Teo, Portfolio optimization in real financial markets with both uncertainty and randomness, Appl. Math. Model. 100 (2021) 125–137.
[15] R. Chen, I.C. Paschalidis, Distributionally robust learning, Foundations Trends®Optim. 4 (1–2) (2020) 1–243.
[16] H. Rahimian, S. Mehrotra, Distributionally robust optimization: a review, arXiv preprint arXiv:1908.05659(2019).
[17] J. Zhen, D. Kuhn, W. Wiesemann, Mathematical foundations of robust and distributionally robust optimization, arXiv preprint arXiv:2105.00760(2021).
[18] E. Delage, Y. Ye, Distributionally robust optimization under moment uncertainty with application to data-driven problems, Oper. Res. 58 (3) (2010) 595–612.
[19] Z. Kang, X. Li, Z. Li, S. Zhu, Data-driven robust mean-CVaR portfolio selection under distribution ambiguity, Quant. Finance 19 (1) (2019) 105–121.
[20] J. Liu, Z. Chen, A. Lisser, Z. Xu, Closed-form optimal portfolios of distributionally robust mean-CVar problems with unknown mean and variance, Appl. Math. Optim. 79 (3) (2019) 671–693.
[21] B. Li, Y. Tan, A. Wu, G. Duan, A distributionally robust optimization based method for stochastic model predictive control, IEEE Trans. Automat. Control 67 (11) (2021) 5762–5776.
[22] Z. Gong, C. Liu, K.L. Teo, J. Sun, Distributionally robust parameter identification of a time-delay dynamical system with stochastic measurements, Appl. Math. Model. 69 (2019) 685–695.
[23] Z. Gong, C. Liu, J. Sun, K.L. Teo, Distributionally robust L_1-estimation in multiple linear regression, Optim. Lett. 13 (4) (2019) 935–947.
[24] B. Li, Y. Rong, J. Sun, K.L. Teo, A distributionally robust linear receiver design for multi-access space-time block coded MIMO systems, IEEE Trans. Wirel. Commun. 16 (1) (2016) 464–474.
[25] B. Li, Y. Rong, J. Sun, K.L. Teo, A distributionally robust minimum variance beamformer design, IEEE Signal Process. Lett. 25 (1) (2017) 105–109.
[26] P. Mohajerin Esfahani, D. Kuhn, Data-driven distributionally robust optimization using the Wasserstein metric: performance guarantees and tractable reformulations, Math. Program. 171 (1) (2018) 115–166.
[27] Z. Hu, L.J. Hong, Kullback-Leibler divergence constrained distributionally robust optimization, Available Optim. Online (2013) 1695–1724.
[28] D. Kuhn, P.M. Esfahani, V.A. Nguyen, S. Shafieezadeh-Abadeh, Wasserstein distributionally robust optimization: theory and applications in machine learning, in: Operations Research & Management Science in the Age of Analytics, Informs, 2019, pp. 130–166.
[29] M.S. Lobo, S. Boyd, The worst-case risk of a portfolio, Unpublished manuscript. Available from http://faculty.fuqua.duke.edu/%7Emlobo/bio/researchfiles/rsk-bnd.pdf (2000).
[30] N. Gülpınar, B. Rustem, Worst-case robust decisions for multi-period mean–variance portfolio optimization, Eur. J. Oper. Res. 183 (3) (2007) 981–1000.
[31] G. Pflug, D. Wozabal, Ambiguity in portfolio selection, Quant. Finance 7 (4) (2007) 435–442.
[32] D. Wozabal, A framework for optimization under ambiguity, Ann. Oper. Res. 193 (1) (2012) 21–47.
[33] Y. Sun, E.L.G. Aw, B. Li, K.L. Teo, J. Sun, CVaR-based robust models for portfolio selection, J. Ind. Manage. Optim. 16 (4) (2020) 1861.
[34] N. Du, Y. Liu, Y. Liu, A new data-driven distributionally robust portfolio optimization method based on Wasserstein ambiguity set, IEEE Access 9 (2020) 3174–3194.
[35] J. Liu, Z. Chen, Time consistent multi-period robust risk measures and portfolio selection models with regime-switching, Eur. J. Oper. Res. 268 (1) (2018) 373–385.
[36] L. Jiang, C. Wu, S. Wang, Distributionally robust multi-period portfolio selection subject to bankruptcy constraints, J. Ind. Manage. Optim. (2021). https://doi.org/10.3934/jimo.2021218
[37] G.C. Calafiore, Ambiguous risk measures and optimal robust portfolios, SIAM J. Optim. 18 (3) (2007) 853–877.

[38] G.A. Hanasusanto, D. Kuhn, Conic programming reformulations of two-stage distributionally robust linear programs over Wasserstein balls, Oper. Res. 66 (3) (2018) 849–869.
[39] V. DeMiguel, A.V. Olivares-Nadal, Technical note–a robust perspective on transaction costs in portfolio optimization, Oper. Res. (2018). (Forthcoming)
[40] B. Li, X. Qian, J. Sun, K.L. Teo, C. Yu, A model of distributionally robust two-stage stochastic convex programming with linear recourse, Appl. Math. Model. 58 (2018) 86–97.
[41] V. DeMiguel, L. Garlappi, F.J. Nogales, R. Uppal, A generalized approach to portfolio optimization: improving performance by constraining portfolio norms, Manage. Sci. 55 (5) (2009) 798–812.
[42] Z. Chen, G. Li, J. Guo, Optimal investment policy in the time consistent mean–variance formulation, Insurance Math. Econ. 52 (2) (2013) 145–156.
[43] S. Corsaro, V. De Simone, Z. Marino, Split Bregman iteration for multi-period mean variance portfolio optimization, Appl. Math. Comput. 392 (2021) 125715.
[44] J. Blanchet, L. Chen, X.Y. Zhou, Distributionally robust mean-variance portfolio selection with Wasserstein distances, Manage. Sci. (2021), doi:10.1287/mnsc.2021.4155. Online first
[45] D. Chen, Y. Wu, J. Li, X. Ding, C. Chen, Distributionally robust mean-absolute deviation portfolio optimization using Wasserstein metric, J. Global Optim. (2022) 1–23.