JUSTC

Article

http://justc.ustc.edu.cn

Received: February 28, 2024; Accepted: April 25, 2024

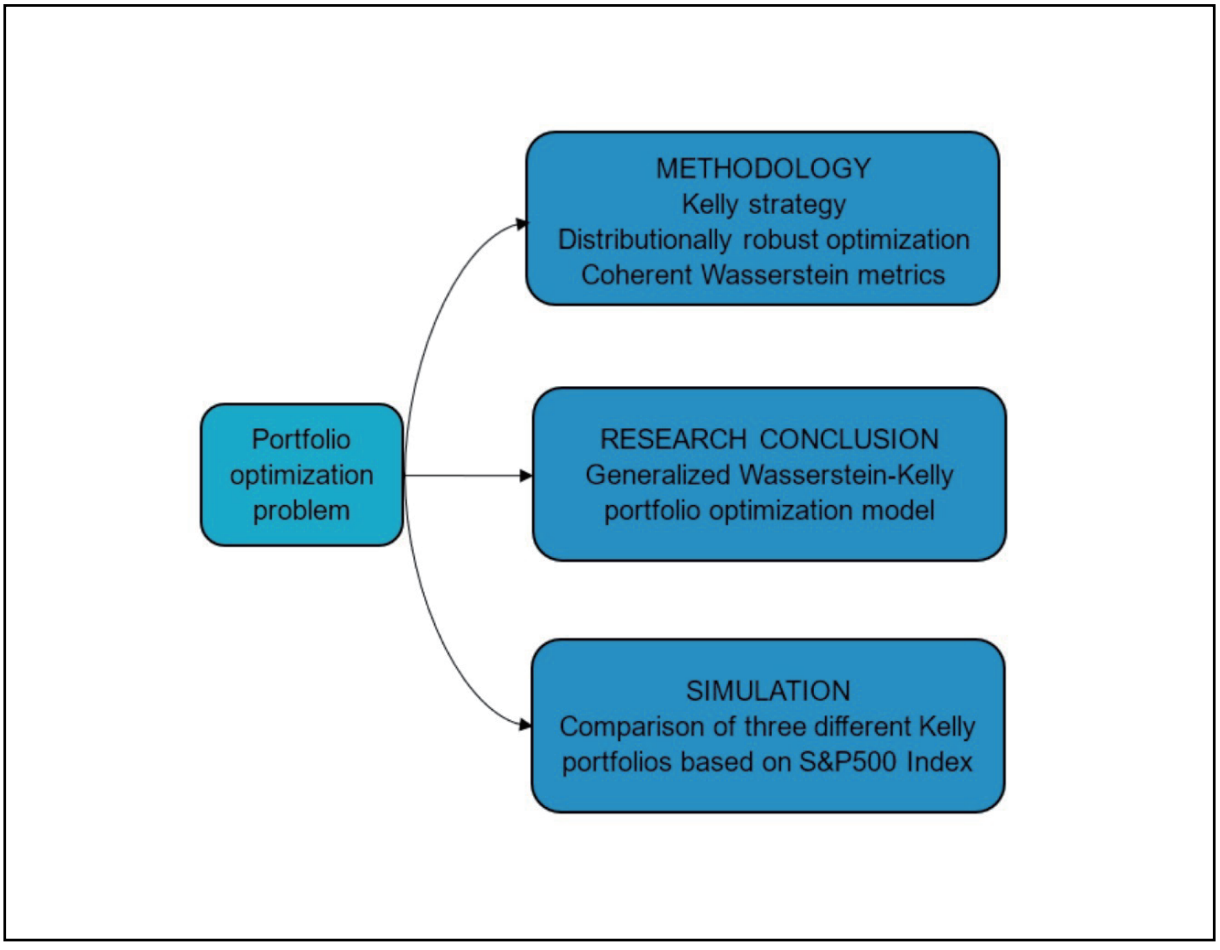
Data-driven distributionally robust Kelly portfolio optimization based on coherent Wasserstein metrics

Yimeng Sun, and Zhenfeng Zou ✉

Department of Statistics and Finance, School of Management, University of Science and Technology of China, Hefei 230026, China

✉Correspondence: Zhenfeng Zou, E-mail: zfzou@ustc.edu.cn

Graphical abstract



Research framework of coherent Wasserstein–Kelly portfolio optimization.

Public summary

■ We study the Kelly portfolio optimization problem based on coherent Wasserstein metrics.

■ We propose the generalized Wasserstein–Kelly portfolio optimization model, which is robust and data-driven.

■ We conduct real and meaningful numerical simulations that demonstrate the validity and superiority of our new models.

Citation: Sun Y M, Zou Z F. Data-driven distributionally robust Kelly portfolio optimization based on coherent Wasserstein metrics. *JUSTC*, **2025**, 55(8): 0805. CSTR: 32290.14.JUSTC-2024-0030. DOI: 10.52396/JUSTC-2024-0030

JUSTC

Article

http://justc.ustc.edu.cn

Received: February 28, 2024; Accepted: April 25, 2024

Data-driven distributionally robust Kelly portfolio optimization based on coherent Wasserstein metrics

Yimeng Sun, and Zhenfeng Zou ✉

Department of Statistics and Finance, School of Management, University of Science and Technology of China, Hefei 230026, China

✉Correspondence: Zhenfeng Zou, E-mail: zfzou@ustc.edu.cn

Cite This: *JUSTC*, **2025**, 55(8): 0805 (10pp)

Read Online

Abstract: The Kelly strategy is a common approach in portfolio optimization problems that aims to maximize the expected portfolio growth rate in the long term. Its computation requires complete knowledge of the asset return distribution, which is obviously not observable, but can be inferred from sample data. Motivated by recent developments in data-driven optimization methods, we propose a new class of coherent Wasserstein data-driven Kelly portfolio optimization models. In particular, we establish a class of ambiguity sets based on coherent Wasserstein metrics, and these new metrics can strike a good balance between robustness and data-drivenness, thus providing richer choices for ambiguity set design. The Kelly portfolio optimization model, which is data-driven and based on coherent Wasserstein balls, can be solved efficiently as a finite-dimensional convex program. This model also provides a robust data-driven solution. In addition, we numerically investigate the proposed model and find that it outperforms the type-1 Wasserstein–Kelly portfolio, especially the classical Kelly portfolio. Moreover, it indicates that we can obtain a portfolio with higher final value and stability, especially in controlling volatility and maximum drawdown.

Keywords: distributionally robust optimization; Kelly strategy; coherent Wasserstein metrics

CLC number: F224; F832.51 **Document code:** A

2020 Mathematics Subject Classification: 65K10; 90G17; 91G10

1 Introduction

First introduced by Kelly in 1956, the Kelly strategy refers to how to choose the optimal bets in a repetitive betting game to maximize wealth accumulation[1]. Since then, the Kelly strategy has attracted considerable attention and has been applied in diverse fields such as optimal allocation[2], portfolio selection[3,4], and fund management[5]. Specifically, the Kelly strategy maximizes the expected portfolio growth rate, which is defined as the natural logarithm of the geometric mean of the total returns over the entire investment horizon. Luenberger[6] noted that if the asset returns are mutually independent and identically distributed, the portfolio growth rate over an infinite investment horizon is equal to the expected logarithm of the total portfolio return over any single rebalancing period according to the strong law of large numbers. The Kelly strategy has many desirable properties. For example, Mossin[7] and Hakansson[8] noted an important practical result with respect to Kelly's strategy, i.e., it has a myopic property. This property means that we can obtain the optimal solution over a multiperiod investment process by solving a single-period optimization problem, and then using convex optimization techniques[9] to solve numerically. In addition, MacLean et al.[10] showed that the portfolio invested using the Kelly strategy has a probability of 1 of outperforming any other portfolio over the long term.

However, the Kelly strategy has an obvious weakness, which requires an adequate understanding of the distribution of joint asset return. Regrettably, the true distribution of returns is typically unknown, or at best only partially known in practice. Previous research on the Kelly strategy has often assumed the distribution of joint asset return and fitted it to sample data. However, this approach is subject to significant misspecification bias. Thus, this method often leads to disappointing performance in out-of-sample experiments[11]. Another natural idea is directly to use the empirical return distribution, which is fully data-driven. It can avoid bias, but this is an over-fitting method, and the gap between the empirical in-sample and the out-of-sample distributions still leads to poor performance.

To overcome these shortcomings, two types of distributionally robust optimization (DRO) methods have emerged in the literature. One is the moment-based uncertainty set, which considers distributions satisfying certain constraints on moments, e.g., the first and second-order moments. Another one is the distance-based uncertainty set, which considers distributions that are within a distance from a reference distribution. A well-known distance-based ambiguity set is defined by the Wasserstein metric[12], which contains both discrete and continuous distributions. With the Wasserstein metric, the DRO model also becomes the well-known Wasserstein DRO problem, which can be effectively tackled. Specifically, Esfahani and Kuhn[13] showed that the problem of Wasserstein DRO can be transformed efficiently into a convex program

CSTR: 32290.14.JUSTC-2024-0030
DOI: 10.52396/JUSTC-2024-0030
JUSTC, **2025**, 55(8): 0805

and enjoy finite-sample performance guarantees, these results can also be seen in Refs. [14, 15]. Further research has explored additional benefits of Wasserstein DRO, including generalization bounds and regularization effects[16,17]. Wasserstein metrics contain a series of metrics with different orders, i.e., the type-p Wasserstein metric, $p\in[1,\infty)$[18]. Regarding the application of the type-p Wasserstein metric in DRO, we recommend Refs. [16, 19]. It is well known that the type-p Wasserstein ball, which is the ambiguity set defined based on the type-p Wasserstein metric, contains only distributions with finite pth-order moments. As discussed by Li and Mao[20]: A higher-order Wasserstein ball is less conservative or can be considered more data-driven. Thus, it needs to balance robustness and data-drivenness when designing ambiguity sets. To address this trade-off, they presented a general framework termed coherent Wasserstein metrics, which not only satisfy the desirable properties but also provide a powerful tool for reconciling the type-1 and type-∞ Wasserstein metrics. Furthermore, the DRO problem based on coherent Wasserstein metrics can be tractably solved as a convex program.

Recent studies have concentrated on solving the Kelly strategy from the perspective of robust optimization or DRO. Rujeerapaiboon et al.[21] designed fixed-mix strategies for solving a data-driven model, whose ambiguity set is composed of first and second-order moments using the empirical distribution. They showed that the problem is indeed a tractable cone program that can be computed efficiently. Sun and Boyd[22] used many different classes of ambiguity sets including polyhedral sets, f-divergence balls, and Wasserstein balls, and ultimately derived the disciplined convex programming forms of distributionally robust Kelly problems. Li[23] employed the classic type-p Wasserstein metric to construct an ambiguity set, resulting in the Wasserstein–Kelly portfolio optimization model. This model can be expressed as a convex program with finite dimensions. Hsieh[24] rewrote the original distributional robust log-optimal portfolio problem into a linear program using a supporting hyperplane approximation approach.

Motivated by Refs. [20, 23], our paper incorporates two merits, i.e., robustness and data-drivenness, in designing uncertainty sets. To be more concrete, we attempt to apply coherent Wasserstein metrics to the Kelly portfolio optimization problem. This method leads to a generalized Wasserstein–Kelly portfolio optimization model, which can be solved efficiently as a convex program. In numerical simulations, we find that such a portfolio based on our new model performs well in out-of-sample experiments and outperforms other portfolios in terms of computational speed, stability, and final portfolio values, which provides innovative ideas for solving the data-driven distributionally robust portfolio optimization problem.

The rest of the paper is organized as follows. Section 2 presents the preparation work, which introduces some necessary background and coherent Wasserstein metrics. Section 3 presents the generalized Wasserstein–Kelly portfolio optimization model, and gives its convex optimization program. Section 4 conducts a real meaningful numerical study based on the S&P 500 index, taking the CVaR Wasserstein–Kelly portfolio as an example. We also compare the numerical results with those of other strategies and highlight the advantages of our model. Section 5 concludes the paper.

Notation. Positive integers and non-negative real numbers are denoted by $\mathbb{N}$ and $\mathbb{R}_+$, respectively. Denote $[n]=\{1,\cdots,n\}$ for any $n\in\mathbb{N}$. For two vectors $a,b\in\mathbb{R}^m$, the inner product is given by $\langle a,b\rangle=\sum_{i=1}^{m}a_ib_i$. Then the dual norm is defined through $\|z\|_*:=\sup_{\|\xi\|\le1}\langle z,\xi\rangle$ for a given norm $\|\cdot\|$ on $\mathbb{R}^m$. We use $A^\top$ to denote the transpose of the matrix or vector A. Consider an atomless probability space $(\Omega,\mathcal{F},\mathbb{P})$, denoted by δ_R the Dirac distribution concentrating unit mass at $R\in\mathbb{R}$. Moreover, $X\overset{\mathrm{d}}{=}Y$ is used to denote that X and Y have the same distribution.

2 Preliminaries

In this section, we first review the settings of the Kelly strategy in Subsection 2.1. Then the definitions of coherent Wasserstein metrics and their properties are collected from Li and Mao[20] in Subsection 2.2. Readers who are familiar with the Kelly strategy and coherent Wasserstein metrics can skip this section.

2.1 Kelly strategy

The Kelly strategy is a good way to calculate the optimal proportion of all bets to be placed each time in a repetitive gambling game, based on the probability of winning or losing. This strategy has been extended to more general application scenarios and is often used to solve portfolio optimization problems. We begin by introducing the basic setup in this subsection.

Now suppose that a long-term investment market is made in $n\in\mathbb{N}$ assets. A decision maker (DM) can adjust his asset allocations only at trading time points $t\in[T]$ with $T\in\mathbb{N}$. Define the prices of these assets at various trading times $t\in[T]$ as $\vec{S}_t:=(S_{t,1},\cdots,S_{t,n})$, and their returns can be denoted by $\vec{R}_t:=(R_{t,1},\cdots,R_{t,n})$, i.e.,

$$R_{t,i}=\frac{S_{t,i}-S_{t-1,i}}{S_{t-1,i}},\quad t\in[T],\ i\in[n].$$

Besides, a series of vectors $w_t=(w_{t,1},\cdots,w_{t,n})$ is the wealth allocation to the n assets at different trading time points t, reflecting the investor's decisions about the proportion of total capital in these n different assets. Naturally, we require w_t to satisfy the properties of non-negativity and that the sum is equal to 1. For simplicity, assume that $w_t\in\mathcal{W}$, where $\mathcal{W}:=\{w\in\mathbb{R}_+^n\mid\vec{\mathbf{1}}^\top w=1\}$.

The DM determines his optimal allocation $\{w_t\}_{t=1}^T$ based on the market information available at time $t\in[T]$. Under the assumption that the initial capital of the DM is 1, the final value of the portfolio at point T can be expressed as

$$V_T=\prod_{t=1}^{T}(1+\vec{R}_t^\top w_t).$$

Thus, the growth rate of the portfolio can be expressed as the natural logarithm of the geometric mean of the total absolute returns, i.e.,

CSTR: 32290.14.JUSTC-2024-0030
DOI: 10.52396/JUSTC-2024-0030
JUSTC, **2025**, 55(8): 0805

$$\gamma_T = \log\left(\prod_{t=1}^{T}(1+\vec{R}_t^{\top}w_t)\right)^{1/T} = \frac{1}{T}\sum_{t=1}^{T}\log(1+\vec{R}_t^{\top}w_t).$$

The reverse formula $V_T = \exp\{T\gamma_T\}$ indicates that there is a monotonic relationship between the final value and the growth rate of the portfolio, so maximizing the final value of the portfolio is equivalent to maximizing the growth rate.

According to the theory of efficient market hypothesis (EMH)[25,26], if the market is weakly efficient, all the existing historical information is reflected by the current market price at any moment. Thus, it is impossible to obtain abnormal profits from investing in these n assets. Then the asset returns $\vec{R}_t$, from a statistical point of view, can be regarded as a white noise process, i.e., this process is independent and identically distributed. Based on EMH, we can use $\vec{R}$ to denote a random vector that has the same distribution as $\vec{R}_t$ for any $t \in [T]$. Also, there are several optimal investment strategies in the literature for maximizing the growth rate of a portfolio. One of the most popular and important portfolio mechanisms is the fixed-mix strategy, also known as the constant proportions strategy, which means keeping the portfolio composition fixed throughout the entire investment horizon, we refer to Ref. [27] for the performance of the constant proportions strategy. Therefore, we only need to make appropriate adjustments to the asset allocation at the trading time points. In this setting, the fixed-mix strategy still requires dynamic adjustments because the market changes randomly between trading periods. This necessitates periodic rebalancing to maintain the target portfolio allocation.

Based on the above statement, we can set the portfolio composition to a fixed value $w \in \mathcal{W}$. This value remains constant throughout the investment interval, i.e., $w_t = w$ for all $t \in [T]$. The asymptotic growth rate of the portfolio, combined with the strong law of large numbers, is as follows:

$$\lim_{T\to\infty}\gamma_T = \mathbb{E}\left[\log(1+\vec{R}^{\top}w)\right]. \tag{1}$$

This suggests that the problem of maximizing the growth rate over the whole investment plan can be reduced to a single-period optimization problem. In particular, the Kelly strategy is a mix-fixed strategy that maximizes the right-hand side of (1), i.e.,

$$\max_{w\in\mathcal{W}}\mathbb{E}\left[\log(1+\vec{R}^{\top}w)\right]. \tag{2}$$

In fact, the Kelly strategy has excellent performance and has been shown to outperform all other causal portfolio strategies, including but not limited to the fixed-mix strategy, in maximizing the growth rate problem. In addition, convex optimization techniques can be used to solve this single-period Kelly portfolio optimization model[9].

2.2 Coherent Wasserstein metrics

As shown in (2), it usually assumes that the DM has full knowledge about the distribution of asset returns, denoted by $\vec{R} \sim F_R$. Actually, obtaining an accurate distribution of asset returns is a challenging task, especially for multivariate distributions. The available information for the investor to infer the true distribution is partially a finite set of historical samples, denoted by $\hat{R}_1, \cdots, \hat{R}_N$. Therefore, a natural idea is the direct replacement of the true distribution F_R in (2) with its empirical distribution $\hat{F}_R := \frac{1}{N}\sum_{j=1}^{N}\delta_{\hat{R}_j}$, and then to solve instead the following optimization problem

$$\max_{w\in\mathcal{W}}\frac{1}{N}\sum_{j=1}^{N}\log(1+\hat{R}_j^{\top}w). \tag{3}$$

This approach is also referred to as sample average approximation (SAA). It is an optimization method based entirely on the empirical distribution. It is obvious that SAA is very sensitive to sample errors, as well as an over-fitting approach that leads to very poor out-of-sample performance.

In recent years, several studies have attempted to reduce the impact of sample errors using DRO, which essentially involves building an uncertainty set, based on the empirical distribution $\hat{F}_R$. Then, the DM optimizes against the worst-case distribution over the ambiguity set, thus obtaining a more robust result:

$$\max_{w\in\mathcal{W}}\inf_{F_R\in\mathcal{B}(\hat{F}_R)}\mathbb{E}_{F_R}\left[\log(1+\vec{R}^{\top}w)\right]. \tag{4}$$

The set $\mathcal{B}(\hat{F}_R)$ is the constructed uncertainty set that most likely contains the true distribution F_R. A natural idea of the set $\mathcal{B}(\hat{F}_R)$ is of the general form:

$$\mathcal{B}(\hat{F}_R) := \left\{F \in \mathcal{M}(\mathbb{R}^n) : d(F,\hat{F}_R) \le \varepsilon\right\},$$

where $\mathcal{M}(\mathbb{R}^n)$ is the set of all distributions supported on $\mathbb{R}^n$, $d(F_1,F_2)$ is a probability metric used to measure the distance between any two distributions $F_1, F_2 \in \mathcal{M}(\mathbb{R}^n)$, and ε is the radius of the ball at the center of the empirical distribution $\hat{F}_R$.

When dealing with (4), it is crucial for the DM to determine the structure of the ambiguity set $\mathcal{B}(\hat{F}_R)$, which depends on choosing the probability metric d. The type-p Wasserstein metric is a common approach for ambiguity sets among several popular probability metrics, because the set constructed by the type-p Wasserstein metric contains both discrete and continuous distributions. First, recall the definition of the Wasserstein metric with order p,

$$d_p(F_1,F_2) := \inf_{\Pi}\left\{(\mathbb{E}[\|\xi_1-\xi_2\|^p])^{1/p}\,\middle|\,\begin{array}{c}\Pi \text{ is a joint distribution}\\ \text{of } \xi_1 \text{ and } \xi_2\\ \text{with marginals } F_1\\ \text{and } F_2, \text{ respectively}\end{array}\right\},$$

for any $p \in [1,\infty)$, and $d_\infty(F_1,F_2)$ is defined as $\lim_{p\to\infty} d_p(F_1,F_2)$ and corresponds to essential supremum, i.e.,

$$d_\infty(F_1,F_2) := \inf_{\Pi}\left\{\text{ess-sup}\,[\|\xi_1-\xi_2\|]\,\middle|\,\begin{array}{c}\Pi \text{ is a joint distribution}\\ \text{of } \xi_1 \text{ and } \xi_2\\ \text{with marginals } F_1\\ \text{and } F_2, \text{ respectively}\end{array}\right\}.$$

Suppose that $\hat{\xi}_1, \cdots, \hat{\xi}_N$ represent N random samples drawn from the distribution F. Then the empirical distribution of F is $\hat{F} := \frac{1}{N}\sum_{j=1}^{N}\delta_{\hat{\xi}_j}$. Therefore, the ambiguity set induced by the

CSTR: 32290.14.JUSTC-2024-0030
DOI: 10.52396/JUSTC-2024-0030
JUSTC, **2025**, 55(8): 0805

type-p Wasserstein metric, called the Wasserstein ball, is given by

$$\mathcal{B}_p(\hat{F}) := \left\{F \in \mathcal{M}(\mathbb{R}^n) : d_p(F,\hat{F}) \le \varepsilon\right\}.$$

The most commonly used type-p Wasserstein metric is the type-1 and type-∞ cases, i.e., $p=1$ or $p=\infty$.

As demonstrated by Li and Mao[20], $\mathcal{B}_1(\hat{F})$ is a good choice from a robustness perspective because it will contain all distributions with finite first-order moments, but this point also makes $\mathcal{B}_1(\hat{F})$ overly conservative because it may include many distributions that are very different from $\hat{F}$. A useful contrast to highlight this limitation of the type-1 Wasserstein ball, $\mathcal{B}_\infty(\hat{F})$ can make full use of the known information to control the distributions with bounded support. Thus, the ambiguity sets constructed by the type-1 and type-∞ Wasserstein metrics, respectively, can be found to be in conflict, so Li and Mao[20] introduced a more mild Wasserstein metric that can achieve robustness and data-drivenness at the same time.

Below, we continue to recall the definition of the coherent Wasserstein metric, which is a key concept of this paper.

Definition 1. (Coherent Wasserstein metric) A metric $d_{\mathcal{M}}(\mathbb{R}^n)\times\mathcal{M}(\mathbb{R}^n)\to\mathbb{R}_+$ is called a coherent Wasserstein metric if it takes the form of

$$d_\rho(F_1,F_2) := \inf_{\Pi}\left\{ \rho(\|\xi_1-\xi_2\|) \;\middle|\; \begin{array}{c}\Pi \text{ is a joint distribution}\\ \text{of } \xi_1 \text{ and } \xi_2\\ \text{with marginals } F_1\\ \text{and } F_2, \text{ respectively}\end{array} \right\}, \tag{5}$$

where $\rho:\mathcal{M}(\mathbb{R}^n)\to\mathbb{R}\cup\{+\infty\}$ is a functional that satisfies the following properties:

(P1) (Translation invariance) $\rho(X+c)=\rho(X)+c$ for any $c\ge 0$;

(P2) (Monotonicity) $\rho(X_1)\ge\rho(X_2)$ for any $X_1\ge X_2$;

(P3) (Subadditvity) $\rho(X_1+X_2)\le\rho(X_1)+\rho(X_2)$;

(P4) (Positive homogeneity) $\rho(\lambda X)=\lambda\rho(X)$ for any $\lambda>0$;

(P5) (Law invariance) $\rho(X_1)=\rho(X_2)$ for any $X_1\overset{\mathrm{d}}{=}X_2$;

(P6) (Normalize) $\rho(0)=0$.

In fact, ρ satisfying Properties (P1)–(P6) in Definition 1 is a law-invariant coherent risk measure in the field of risk management, which represents the minimal amount of capital that has to be raised and injected in a financial position to ensure that the position is acceptable (safe)[28]. However, Subsection 2.1 considers the return random vectors, so ρ, satisfying properties (P1)–(P6) of Definition 1, is only a functional and cannot be interpreted as the minimal amount of capital. It is easy to see that the expectation $\mathbb{E}$ and the ess-sup satisfy properties (P1)–(P6) listed in Definition 1, so considering the role of the general functional ρ in Wasserstein metrics is a natural forward step. Also, coherent Wasserstein metrics can be viewed as natural risk-averse formulations of the classical optimal transport problem[18]. These new coherent Wasserstein metrics, similar to classical Wasserstein metrics, are valid distance metrics with the following properties: identity of indiscernibles, non-negativity, symmetry, and triangle inequality[20].

This family of metrics combines the excellent qualities of Wasserstein metrics with orders 1 and ∞, that is, robustness and data-drivenness[20] (see Definitions 2 and 3 below for their formal definitions). For a simpler representation, we use a random variable X called a transportation random variable to replace $\|\xi_1-\xi_2\|$ if there exists $X\overset{\mathrm{d}}{=}\|\xi_1-\xi_2\|$. Denote A as an index set.

Definition 2. (Robustness) A family of metrics $\{d_{\rho_\alpha}\}_{\alpha\in A}$ is said to have the property of robustness, which means that for each $\alpha\in A$, ρ_α is well-defined (takes finite value) for any transportation random variable X with finite first moment, i.e., $X\in L_1$.

Definition 3. (Data-drivenness) A family of metrics $\{d_{\rho_\alpha}\}_{\alpha\in A}$ is said to have the property of data-drivenness, which means that there exists a series of indices $\alpha_n\in A, n\in\mathbb{N}$, such that ρ_{α_n} converges to ess-sup, i.e., the worst-case risk measure. The Wasserstein ball $\mathcal{B}_{\rho_{\alpha_n}}(\hat{F})$ converges to $\mathcal{B}_\infty(\hat{F})$ as $n\to\infty$.

Li and Mao[20] argued that the flexibility needed to build a new family of metrics is provided by the coherent Wasserstein metrics. These new metrics $\{d_{\rho_\alpha}\}_{\alpha\in A}$ satisfy the robustness and data-drivenness. They also provided an equivalent proposition for characterizing robustness and data-drivenness using the dual representation of ρ_α.

Below, we introduce two families of coherent Wasserstein metrics based on two popular risk measures, namely conditional value-at-risk (CVaR) and expectile. These two families of metrics not only conform to the definition of the coherent Wasserstein metric, but can also be easily proven to satisfy robustness and data-drivenness requirements. Recall that the value-at-risk (VaR) at level $\alpha\in(0,1)$ of X is defined by

$$\mathrm{VaR}_\alpha(X)=F^{-1}(\alpha)=\inf\{x:F(x)\ge\alpha\},\quad X\sim F.$$

Example 1. (CVaR-Wasserstein metric) Let's first review the definition of CVaR as follows:

$$\rho(X)=\mathrm{CVaR}_\alpha(X)=\frac{1}{1-\alpha}\int_\alpha^1 \mathrm{VaR}_u(X)\mathrm{d}u,\quad \alpha\in[0,1).$$

By Definition 1, it is easy to see that

$$d_{\mathrm{CVaR}_\alpha}(F_1,F_2) := \inf_{\Pi}\{\mathrm{CVaR}_\alpha(\|\xi_1-\xi_2\|):\Pi\in\Pi(F_1,F_2)\},$$

and the CVaR-Wasserstein ball is defined by

$$\mathcal{B}^\alpha_{(1)}(\hat{F}) := \left\{F\in\mathcal{M}(\mathbb{R}^n): d_{\mathrm{CVaR}_\alpha}(F,\hat{F})\le\varepsilon\right\},\quad \alpha\in[0,1).$$

Example 2. (Expectile-Wasserstein metric) Consider a random variable X with $\mathbb{E}[X^2]<\infty$. Then the expectile $e_\alpha(X)$ at level $\alpha\in[0,1)$ of X is the minimizers of an asymmetric quadratic loss,

$$e_\alpha(X)=\operatorname*{argmin}_{x\in\mathbb{R}}\{\alpha\mathbb{E}[((X-x)_+)^2]+(1-\alpha)\mathbb{E}[((X-x)_-)^2]\},$$

where $a_+=\max\{a,0\}$ and $a_-=\max\{-a,0\}$. Actually, by the first-order-condition, we can simply require $X\in L_1$ since $e_\alpha(X)$ is defined as the unique solution to

$$\alpha\mathbb{E}[(X-x)_+]=(1-\alpha)\mathbb{E}[(X-x)_-],$$

where $x\in\mathbb{R}$. It is well known that $e_\alpha(X)$ is coherent for any $\alpha\in[1/2,1)$ and reduces to the expectation $\mathbb{E}$ when $\alpha=1/2$. Also, expectile converges to the worst-case risk measure ess-sup as $\alpha\to1$; see Ref. [29]. For a level $\alpha\in[1/2,1)$, let $\rho(X)=e_\alpha(X)$, then we obtain the following metric by

CSTR: 32290.14.JUSTC-2024-0030
DOI: 10.52396/JUSTC-2024-0030
JUSTC, **2025**, 55(8): 0805

Definition 1:

$$d_{e_\alpha}(F_1,F_2):=\inf_{\Pi}\{e_\alpha(\|\xi_1-\xi_2\|):\Pi\in\Pi(F_1,F_2)\},$$

and the expectile-Wasserstein ball

$$\mathcal{B}^{\alpha}_{(2)}(\hat{F}):=\left\{F\in\mathcal{M}(\mathbb{R}^n):d_{e_\alpha}(F,\hat{F})\le\varepsilon\right\},\quad \alpha\in[1/2,1).$$

3 Main results

In this section, we first present how we can apply the coherent Wasserstein metric to the Kelly portfolio optimization problem. Note that the support set of F_R from the coherent Wasserstein ball may be unbounded. This may cause the absolute return $1+\vec{R}^{\top}w$ to be negative, which conflicts with the feasible range of the logarithmic function. To overcome this deficiency, Li[23] introduced the following transformation:

$$1+\vec{R}_t^{\top}w=\exp(\vec{r}_t)^{\top}w,$$

where

$$\vec{r}_t:=(r_{t,1},\cdots,r_{t,n})=\left(\ln\frac{S_{t,1}}{S_{t-1,1}},\cdots,\ln\frac{S_{t,n}}{S_{t-1,n}}\right).$$

In fact, the composition of $\vec{r}_t$ is consistent with the stochastic behavior models of stock prices often used in the financial field. Now, let $\hat{r}_1,\cdots,\hat{r}_N$ represent N samples of log-returns and the empirical distribution of log-returns be $\hat{F}_r:=\frac{1}{N}\sum_{j=1}^{N}\delta_{\hat{r}_j}$. The ball based on the coherent Wasserstein metric for the distribution of log-returns is defined by

$$\mathcal{B}_{\rho_\alpha}(\hat{F}_r):=\left\{F\in\mathcal{M}(\mathbb{R}^n):d_{\rho_\alpha}(F,\hat{F}_r)\le\varepsilon\right\},$$

where $\{d_{\rho_\alpha}\}_{\alpha\in A}$ is a family of coherent Wasserstein metrics. Thus, the following data-driven coherent Wasserstein Kelly portfolio optimization model is the main object of this paper:

$$\max_{w\in\mathcal{W}}\inf_{F_r\in\mathcal{B}_{\rho_\alpha}(\hat{F}_r)}\mathbb{E}_{F_r}[\log(\exp(\vec{r})^{\top}w)]. \tag{6}$$

(6) provides a new way of thinking about portfolio selection with the Kelly strategy, as well as an application of coherent Wasserstein DRO, so we call it the generalized Wasserstein–Kelly portfolio optimization model. From a computational point of view, it is easy to address because it makes the objective function associated with the variable $\vec{r}$ convex, allowing further reconstruction of the model into a convex program with finite dimensions that can be solved by a standard solver. We leave the proof of Theorem 1 below to the following subsection.

Theorem 1. The following convex optimization program can be used to solve the generalized Wasserstein–Kelly portfolio optimization model (6):

$$\begin{aligned}\max\ &\frac{1}{N}\sum_{j=1}^{N}\left(\hat{r}_j^{\top}v^{(j)}+\sum_{i=1}^{N}v_i^{(j)}\log\left(\frac{w_i}{v_i^{(j)}}\right)\right)-\lambda\varepsilon,\\ \text{subject to}\ &w\in\mathcal{W},\ \lambda\ge0,\ \frac{q}{\lambda}\in A_\rho,\ v^{(j)}\ge0,\\ &q_j\ge\left\|v^{(j)}\right\|_*,\ \sum q_j=N\lambda.\end{aligned} \tag{7}$$

where $\|\cdot\|_*$ represents the dual norm of the norm $\|\cdot\|$, and A_ρ is a subset of a probability simplex, defined by

$$A_\rho=\left\{y\in\mathbb{R}^N:\exists Z\sim\frac{1}{N}\sum_{i=1}^{N}\delta_{y_i},\ Z\in\mathcal{Z}_\rho\right\},$$

and $\mathcal{Z}_\rho$ represents the risk envelope of ρ, i.e., $\mathcal{Z}_\rho=\{Z\ge0:\mathbb{E}[Z]=1,\ \mathbb{E}[ZX]\le\rho(X)\ \text{for all}\ X\}$. It is defined that $q/0\notin A_\rho$ for any $q\ne0$ and $0/0\in A_\rho$.

If $\rho(\cdot)=\mathbb{E}[\cdot]$ in Theorem 1, then $\mathcal{Z}_{\mathbb{E}}=\{1\}$, i.e., $A_{\mathbb{E}}=\{(1,\cdots,1)\}\in\mathbb{R}^N$. The constraint condition $\frac{q}{\lambda}\in A_{\mathbb{E}}$ in (7) reduces to $q_i=q_j$ when $i,j\in[N]$. Therefore, Theorem 1 reduces to

$$\begin{aligned}\max\ &\frac{1}{N}\sum_{j=1}^{N}\left(\hat{r}_j^{\top}v^{(j)}+\sum_{i=1}^{n}v_i^{(j)}\log\left(\frac{w_i}{v_i^{(j)}}\right)\right)-\lambda\varepsilon,\\ \text{subject to}\ &w\in\mathcal{W},\ v^{(j)}\ge0,\ \lambda\ge\left\|v^{(j)}\right\|_*.\end{aligned} \tag{8}$$

It is well known that $\rho(\cdot)=\mathbb{E}[\cdot]$ reduces to the type-1 Wasserstein metric, then our conclusion recovers Proposition 2.2 in Ref. [23].

Note that, the objective functions of (7) and (8) are the same, with only different constraints. Furthermore, if the coherent Wasserstein metric is induced by CVaR or expectile, i.e., the CVaR-Wasserstein metric (Example 1) or expectile-Wasserstein metric (Example 2), we have the following corollaries, which will be used for simulations in Section 4.

Corollary 1. For any $\alpha\in[0,1)$, CVaR_α can be expressed as (see Ref. [28, Theorem 4.52])

$$\mathrm{CVaR}_\alpha(X)=\sup_{Z\in\mathcal{Z}_{\mathrm{CVaR}_\alpha}}\mathbb{E}[ZX]$$

with

$$\mathcal{Z}_{\mathrm{CVaR}_\alpha}=\left\{Z\ge0:\mathbb{E}[Z]=1,\ Z\le\frac{1}{1-\alpha}\right\}.$$

Therefore, we obtain

$$A_{\mathrm{CVaR}_\alpha}=\left\{y\in\mathbb{R}_+^N:\sum_{i=1}^{N}y_i=N,\ y_i\le\frac{1}{1-\alpha}\right\}.$$

Therefore, according to Theorem 1, problem (7) can be solved by

$$\begin{aligned}\max\ &\frac{1}{N}\sum_{j=1}^{N}\left(\hat{r}_j^{\top}v^{(j)}+\sum_{i=1}^{n}v_i^{(j)}\log\left(\frac{w_i}{v_i^{(j)}}\right)\right)-\lambda\varepsilon,\\ \text{subject to}\ &w\in\mathcal{W},\ \lambda\ge0,\ v^{(j)}\ge0,\\ &q_j\ge\left\|v^{(j)}\right\|_*,\ q_j\le\frac{\lambda}{1-\alpha},\ \sum q_j=N\lambda.\end{aligned}$$

Corollary 2. For any $\alpha\in[1/2,1)$, expectile e_α can be expressed as (see Ref. [29, Proposition 8])

$$e_\alpha(X)=\sup_{Z\in\mathcal{Z}_{e_\alpha}}\mathbb{E}[XZ]$$

with

$$\mathcal{Z}_{e_\alpha}=\left\{Z\ge0:\mathbb{E}[Z]=1,\ \frac{\text{ess-sup}\,Z}{\text{ess-inf}\,Z}\le\frac{\alpha}{1-\alpha}\right\}.$$

CSTR: 32290.14.JUSTC-2024-0030
DOI: 10.52396/JUSTC-2024-0030
JUSTC, **2025**, 55(8): 0805

Therefore, we get

$$A_{e_\alpha} = \left\{ y \in \mathbb{R}_+^N : \sum_{i=1}^N y_i = N,\ \frac{\max_i y_i}{\min_i y_i} \le \frac{\alpha}{1-\alpha} \right\}.$$

Therefore, according to Theorem 1, problem (7) can be solved by

$$\begin{aligned} \max\ & \frac{1}{N}\sum_{j=1}^N \left(\hat{r}_j^\top v^{(j)} + \sum_{i=1}^n v_i^{(j)} \log\left(\frac{w_i}{v_i^{(j)}}\right)\right) - \lambda\varepsilon, \\ \text{subject to}\quad & w \in \mathcal{W},\ \lambda \ge 0,\ v^{(j)} \ge 0, \\ & q_j \ge \left\| v^{(j)} \right\|_*,\ q_i \le \frac{\alpha}{1-\alpha} q_j,\ \sum q_j = N\lambda. \end{aligned}$$

3.1 Proof of Theorem 1

The optimization problem (6) is

$$\max_{w\in\mathcal{W}} \inf_{F_r \in \mathcal{B}_\rho(\hat{F}_r)} \mathbb{E}_{F_r}[\log(\exp(\vec{r})^\top w)], \tag{9}$$

where $\mathcal{B}_\rho(\hat{F}_r) := \left\{ F \in \mathcal{M}(\mathbb{R}^n) : d_\rho(F, \hat{F}_r) \le \varepsilon \right\}$, and ρ is a coherent Wasserstein metric. We will solve the optimal solution of the wealth allocation vector w from (9), which is also the result of the optimization problem below:

$$\min_{w\in\mathcal{W}} \sup_{F_r \in \mathcal{B}_\rho(\hat{F}_r)} \mathbb{E}_{F_r}[-\log(\exp(\vec{r})^\top w)]. \tag{10}$$

Let $u(r,w) := \log(\exp(\vec{r})^\top w) = \log(\sum_{i=1}^n e^{r_i} w_i)$. It is observed that $u(r,w)$ is a convex function with respect to r and, hence, $\ell(r,w) = -u(r,w)$ is concave with respect to r. For ease of notation, we ignore the variable w and denote $\ell(r,w)$ by $\ell(r)$.

We first focus on the inner maximization problem of (10), i.e., the worst-case expectation problem. Following the definition of coherent Wasserstein metric, it can also be expressed in terms of the joint distributions Π:

$$\begin{aligned} \sup_{\Pi}\ & \mathbb{E}^\Pi[\ell(r)] \\ \text{subject to}\quad & \rho^\Pi(\|\hat{r} - r\|) \le \varepsilon, \\ & \Pi \text{ is a joint distribution of } \hat{r} \text{ and } r \\ & \text{with marginals } \hat{F}_r \text{ and } F_r, \text{ respectively.} \end{aligned} \tag{11}$$

Since $\ell(r)$ is concave in r, (11) is equivalent to (see Ref. [20, Theorem 1])

$$\begin{aligned} \sup_{r_1,\cdots,r_N}\ & \frac{1}{N}\sum_{j=1}^N \ell(r_j) \\ \text{subject to}\quad & \rho^\Pi(\|\hat{r} - r\|) \le \varepsilon, \\ & \Pi((\hat{r}, r) = (\hat{r}_j, r_j)) = \frac{1}{N},\ r_j \in \mathbb{R}^n,\ j \in [N]. \end{aligned} \tag{12}$$

Notice that given any random variable Y that satisfies $\mathbb{P}(Y = r_j) = 1/N$ for $j \in [N]$, any law-invariant convex risk measure ρ can also be written in terms of (see Ref. [20, Proving (44)])

$$\rho(Y) = \sup_{z\in A_\rho} \frac{1}{N}\sum_{i=1}^N r_j z_j,$$

where A_ρ is a subset of a probability simplex, defined by

$$A_\rho = \left\{ y \in \mathbb{R}^N : \exists Z \sim \frac{1}{N}\sum_{j=1}^N \delta_{y_j},\ Z \in \mathcal{Z}_\rho \right\},$$

and $\mathcal{Z}_\rho$ denotes the risk envelope of ρ, i.e., $\mathcal{Z}_\rho = \{Z \ge 0 : \mathbb{E}[Z] = 1,\ \mathbb{E}[ZX] \le \rho(X) \text{ for all } X\}$.

According to this, we can write problem (12) as

$$\begin{aligned} \sup_{r_1,\cdots,r_N}\ & \frac{1}{N}\sum_{j=1}^N \ell(r_j) \\ \text{subject to}\quad & \sup_{z\in A_\rho} \left\{ \frac{1}{N}\sum_{j=1}^N \left\| \hat{r}_j - r_j \right\| z_j \right\} \le \varepsilon,\ r_j \in \mathbb{R}^n,\ j \in [N], \end{aligned} \tag{13}$$

and its Lagrange dual by

$$\begin{aligned} & \inf_{\lambda\ge 0} \sup_{r_j\in\mathbb{R}^n} \frac{1}{N}\sum_{j=1}^N \ell(r_j) + \lambda\left(\varepsilon - \sup_{z\in A_\rho} \left\{ \frac{1}{N}\sum_{j=1}^N \left\| \hat{r}_j - r_j \right\| z_j \right\} \right) = \\ & \inf_{\lambda\ge 0} \sup_{r_j\in\mathbb{R}^n} \inf_{z\in A_\rho} \frac{1}{N}\sum_{j=1}^N \ell(r_j) + \lambda\left(\varepsilon - \left\{ \frac{1}{N}\sum_{j=1}^N \left\| \hat{r}_j - r_j \right\| z_j \right\} \right) = \\ & \inf_{\lambda\ge 0, z\in A_\rho} \sup_{r_j\in\mathbb{R}^n} \lambda\varepsilon + \frac{1}{N}\sum_{j=1}^N (\ell(r_j) - \lambda \left\| \hat{r}_j - r_j \right\| z_j) = \\ & \inf_{\lambda\ge 0, z\in A_\rho} \lambda\varepsilon + \frac{1}{N}\sum_{j=1}^N \sup_{r\in\mathbb{R}^n} (\ell(r) - \lambda \left\| \hat{r}_j - r \right\| z_j), \end{aligned} \tag{14}$$

where the second equation abides by Sion's minimax theorem, which states that the objective function is convex in z and concave in r_j, and A_ρ is a compact set. When $\varepsilon > 0$, strong duality holds since the slater condition is satisfied. Strong duality also holds when $\varepsilon = 0$, because both the original and the dual degenerate to $\frac{1}{N}\sum_{j=1}^N \ell(\hat{r}_j)$.

Taking $q = \lambda z$, consider the problem

$$\sup_{r\in\mathbb{R}^n} \left(\ell(r) - \left\| \hat{r}_j - r \right\| q_j \right). \tag{15}$$

Note that $g(r) := \left\| \hat{r}_j - r \right\| p_j$ is convex in r. Therefore, we apply the Fenchel duality theorem to (15), which leads to

$$\begin{aligned} \sup_{r\in\mathbb{R}^n} (\ell(r) - \left\| \hat{r}_j - r \right\| q_j) &= \sup_{r\in\mathbb{R}^n} (\ell(r) - g(r)) = \\ & - \inf_{r\in\mathbb{R}^n} (-\ell(r) + g(r)) = \\ & - \max_v (-(-\ell)^*(v) - g^*(-v)) = \\ & - \max_v (-u^*(v) - g^*(-v)) = \\ & \min_v u^*(v) + g^*(-v), \end{aligned} \tag{16}$$

where u^* and g^* represent the convex conjugates of u and g, respectively, i.e.,

$$u^*(v) = -\sum_{i=1}^n v_i \cdot \log\left(\frac{w_i}{v_i}\right),\ v > 0,$$

$$g^*(-v) = \begin{cases} -\hat{r}_j^\top v, & \|v\|_* \le q_j; \\ \infty, & \text{otherwise.} \end{cases}$$

CSTR: 32290.14.JUSTC-2024-0030
DOI: 10.52396/JUSTC-2024-0030
JUSTC, **2025**, 55(8): 0805

Substituting these convex conjugates into (16), we have

$$\min_{v} u^*(v)+g^*(-v)=\min_{v\geq 0, q_j\geq\|v\|_*} -\sum_{i=1}^{n} v_i\log\left(\frac{w_i}{v_i}\right)-\hat{r}_j^{\top}v.$$

Putting this result into (14), we can get the equivalent expression for the worst-case expectation problem, i.e.,

$$\begin{aligned}&\min \frac{1}{N}\sum_{j=1}^{N}\left(-\hat{r}_j^{\top}v^{(j)}-\sum_{i=1}^{n} v_i^{(j)}\log\left(\frac{w_i}{v_i^{(j)}}\right)\right)+\lambda\varepsilon,\\ &\text{subject to}\quad \lambda\geq 0,\ \frac{q}{\lambda}\in A_\rho,\ v^{(j)}\geq 0,\\ &\qquad\qquad q_j\geq\left\|v^{(j)}\right\|_*,\ \sum q_j=N\lambda.\end{aligned}\tag{17}$$

The final step is to replace the inner maximization in (10) with the minimization problem above, we obtain the final formulation as follows:

$$\begin{aligned}&\min \frac{1}{N}\sum_{j=1}^{N}\left(-\hat{r}_j^{\top}v^{(j)}-\sum_{i=1}^{n} v_i^{(j)}\log\left(\frac{w_i}{v_i^{(j)}}\right)\right)+\lambda\varepsilon,\\ &\text{subject to}\quad w\in\mathcal{W},\lambda\geq 0,\ \frac{q}{\lambda}\in A_\rho,\ v^{(j)}\geq 0,\\ &\qquad\qquad q_j\geq\left\|v^{(j)}\right\|_*,\ \sum q_j=N\lambda.\end{aligned}\tag{18}$$

We know that the optimal wealth allocation vector w solved by (9) and (10) are the same, differing only by a minus sign in the result of the objective function, so the reformulation of (9) is obtained by transforming the minimum problem in (18) into a maximum problem, which is the desired result in Theorem 1.

4 Numerical studies

In this section, we compare the generalized Wasserstein–Kelly portfolio optimization model based on the CVaR-Wasserstein metric at a level of 0.9 (i.e., $\alpha=0.9$) with the Kelly strategy based on the SAA method (3) and the classical type-1 Wasserstein–Kelly portfolio optimization model[23] (8) to evaluate the effectiveness of our model (6). The main reason for choosing CVaR is that it not only includes expectation $\mathbb{E}$, but is also a commonly used measure in quantitative risk management. Daily closing prices data for each S&P 500 stock from January 1, 2020 to December 31, 2023 are collected. The following illustrates the effect of the radius parameter ε on portfolio diversification. It is based on a few selected stocks in Subsection 4.1. Then we use many randomly selected stocks from the index to assess the out-of-sample performance of three different portfolios in Subsection 4.2. All the simulations are performed based on the **CVXPY** package for Python.

4.1 Diversification effects

The distance between the true and the empirical distributions is limited by the radius parameter ε under the CVaR-Wasserstein metric. The effects of changing the parameters on the composition of the generalized Wasserstein–Kelly portfolio are studied. To relate this radius with the historical data, we specifically set $\varepsilon=\delta\cdot\bar{r}$, where $\bar{r}$ denotes the average log-return over all assets, and δ can be considered as a scaling parameter, which shows the size of ε relative to $\bar{r}$. We first select 10 major stocks in the S&P 500 index with prices, as shown in Fig. 1.

The generalized Wasserstein–Kelly portfolios are optimized based on the historical data of 2020 (the first 253 trading days in Fig. 1). The trend of the portfolios with different δ values is displayed in Fig. 2, which shows the change in the proportion of each stock. Obviously, when $\delta=0$, it is the Kelly strategy with the SAA method, and as δ increases, it leads to greater portfolio diversification. We find that the proportions of different stocks will gradually converge to $1/N$, which means that the generalized Wasserstein–Kelly portfolio converges to $1/N$ portfolio in the limit. This is also a good illustration of how the generalized Wasserstein–Kelly portfolios achieve robustness through diversification.

4.2 Out-of-sample performances

In this subsection, we use three methods, i.e., SAA, the type-1 Wasserstein metric and the CVaR-Wasserstein metric, on the data-driven Kelly portfolio to compare their out-of-sample performances. For simplicity, we call them the Kelly portfolio, type-1 portfolio, and CVaR portfolio, respectively. Ten stocks from the S&P 500 index are randomly selected, then using historical closing prices from 2020 (253 trading days)

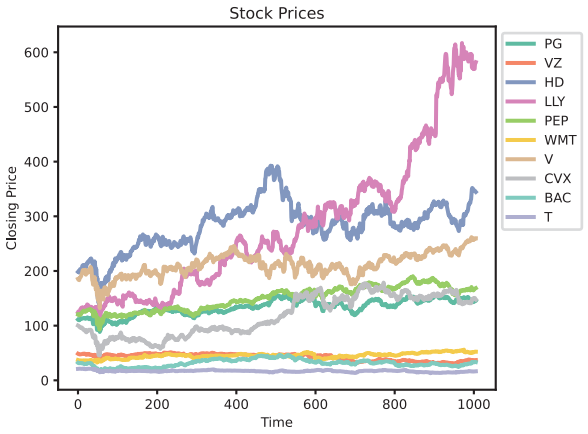


Fig. 1. The closing prices of 10 major stocks.

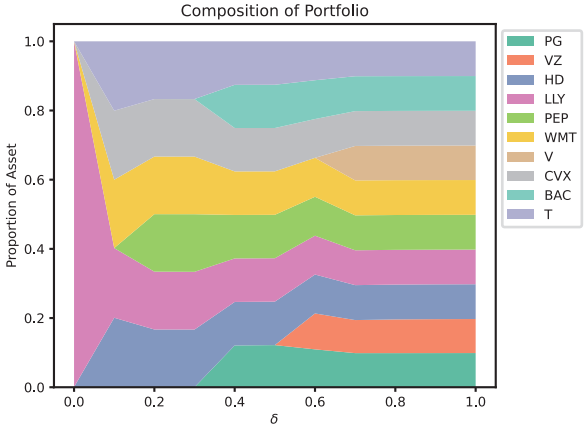
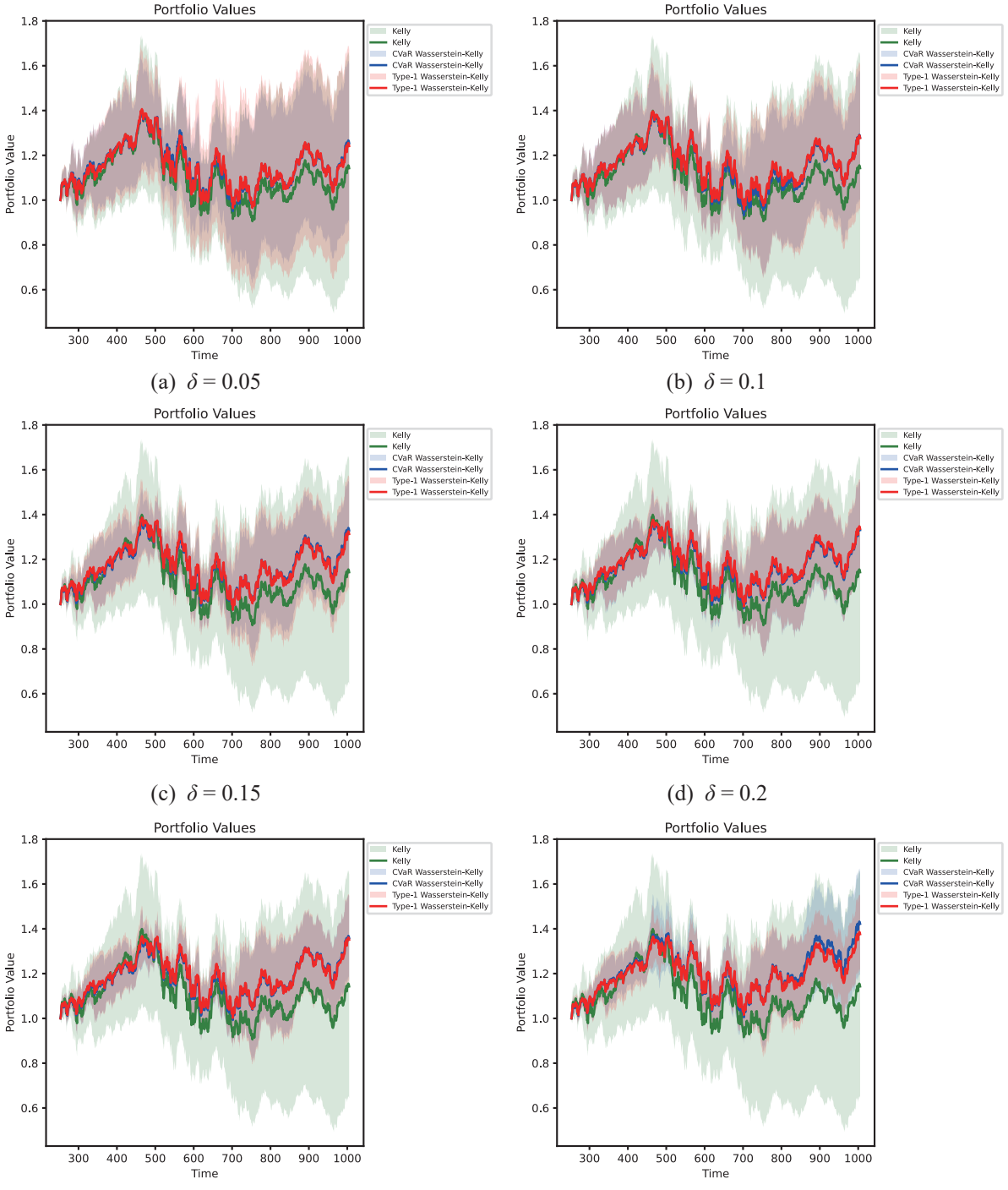


Fig. 2. Composition of the generalized Wasserstein–Kelly portfolio with different scaling parameters δ.

CSTR: 32290.14.JUSTC-2024-0030
DOI: 10.52396/JUSTC-2024-0030
JUSTC, **2025**, 55(8): 0805

to calculate the Kelly portfolio and two Wasserstein–Kelly portfolios. The out-of-sample tests are conducted over the period from 2021-01-01 to 2023-12-31. We simulate the portfolio values of the three methods on each trading day. The above process is repeated 100 times, thus reducing the impact of the subjectivity and extreme cases of stock selection on the experimental results. Below, we set $\delta = 0.05$, 0.1, 0.15, 0.2, 0.25, and 0.3 to further observe the effect of the radius parameter ε on the results of the optimization.

The simulation results with different δ values for the Kelly portfolio and two Wasserstein–Kelly portfolios over the out-of-sample horizons 2021-01-01–2023-12-31 are displayed in Fig. 3. The bold lines represent the average portfolio value of 100 repetitions, and the standard deviation of these repetitions is depicted by the shaded parts. According to the simulation, the CVaR portfolio is the optimal portfolio in terms of



(a) $\delta = 0.05$ (b) $\delta = 0.1$

(c) $\delta = 0.15$ (d) $\delta = 0.2$

(e) $\delta = 0.25$ (f) $\delta = 0.3$

Fig. 3. The development of portfolio values for the Kelly portfolio, the CVaR portfolio, and the type-1 portfolio with δ over the period from 2021-01-01 to 2023-12-31.

CSTR: 32290.14.JUSTC-2024-0030
DOI: 10.52396/JUSTC-2024-0030
JUSTC, **2025**, 55(8): 0805

both the return values and the stability. As δ increases, this method converges quickly. Thus, we can obtain the optimized results in the "neighborhood" of the empirical distribution, which has several advantages in terms of computational complexity and speed.

The following boxplots in Fig. 4 respectively show the annualized return, annualized volatility, Sharpe ratio, maximum drawdown, and log of final portfolio value. We can observe that the CVaR portfolio outperforms the other two portfolios on these performance indicators. Furthermore, the boxplots of the CVaR portfolio show significant "shrinkage" as δ increases. This suggests that our model (6) will be superior in terms of robustness.

5 Conclusions

In this paper, we incorporate coherent Wasserstein metrics into the Kelly strategy, thus providing new ideas for solving the generalized Wasserstein DRO problem. We construct a ball of radius ε centered at the empirical distribution relying on coherent Wasserstein metrics. Then we propose the

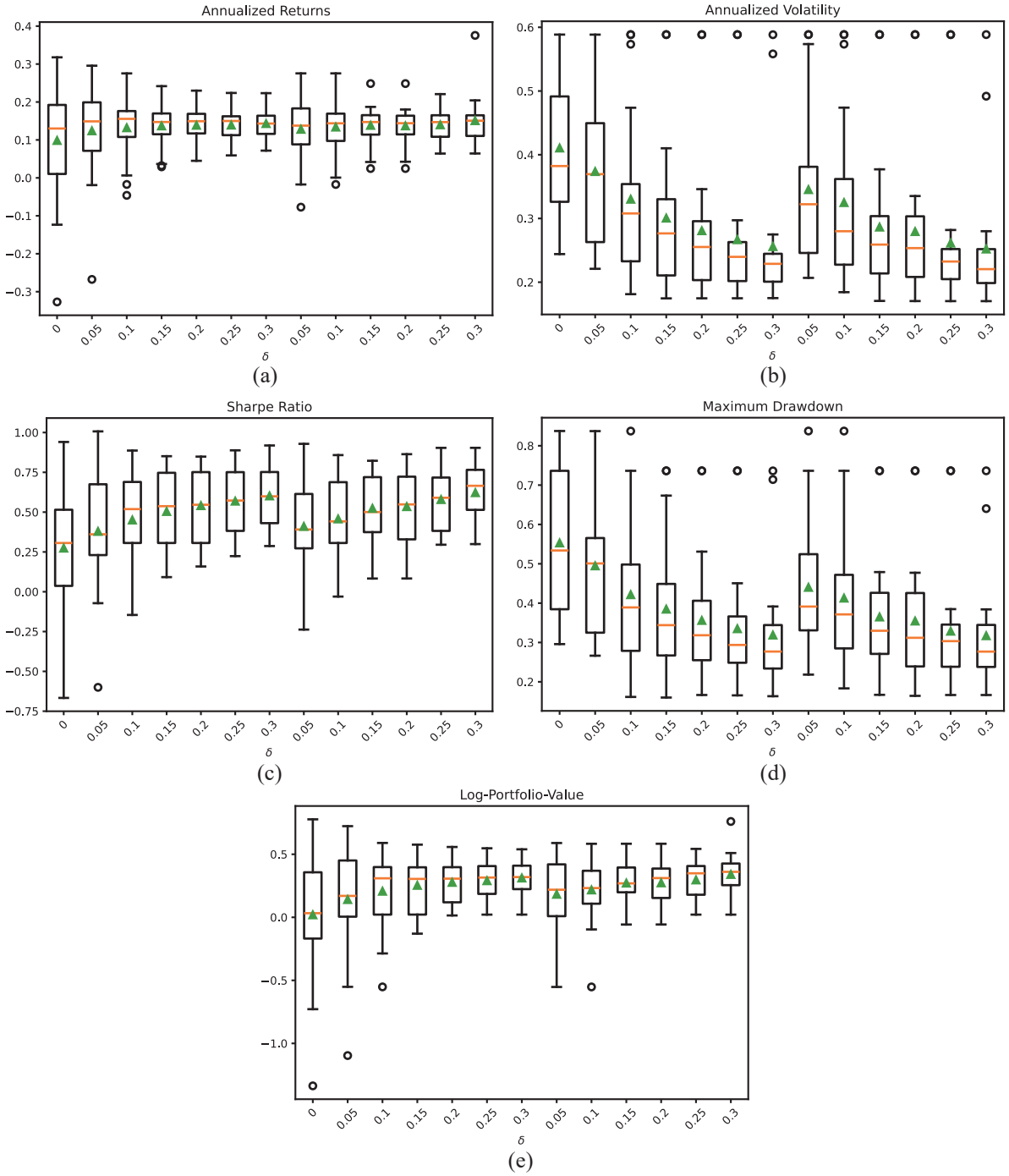


Fig. 4. The boxplots for three portfolios with different δ, and $\delta = 0$ corresponds to the Kelly portfolio, while $\delta = 0.05$, 0.1, 0.15, 0.2, 0.25, and 0.3 correspond to the two Wasserstein–Kelly portfolios, where the former is type-1 portfolio and the latter is CVaR portfolio.

CSTR: 32290.14.JUSTC-2024-0030
DOI: 10.52396/JUSTC-2024-0030
JUSTC, **2025**, 55(8): 0805

generalized Wasserstein–Kelly portfolio optimization model (6). Within convex techniques, it is easy to reformulate the model as a tractable convex program with finite dimensions that can not only be computed efficiently but also obtain solutions that perform well in out-of-sample tests. In particular, we perform meaningful numerical simulations using the family of CVaR-Wasserstein metrics. The numerical results show that it outperforms the Kelly portfolio and the type-1 Wasserstein– Kelly portfolio in both final value and stability. In addition, the CVaR Wasserstein–Kelly portfolio will be a more desirable choice for investors who are more concerned with volatility and maximum drawdown. Overall, our findings provide a new attempt to explore the potential of the generalized Wasserstein DRO in portfolio optimization problems.

Acknowledgements

This work was supported by the National Natural Science Foundation of China (12401625), the China Postdoctoral Science Foundation (2024M753074), the Postdoctoral Fellowship Program of CPSF (GZC20232556), and the Fundamental Research Funds for the Central Universities (WK2040000108).

Conflict of interest

The authors declare that they have no conflict of interest.

Biographies

Yimeng Sun is a master's student at the Department of Statistics and Finance, University of Science and Technology of China. Her research mainly focuses on distributionally robust optimization.

Zhenfeng Zou is currently a postdoctoral researcher at the Department of Statistics and Finance, University of Science and Technology of China (USTC). He received his Ph.D. degree in Statistics from USTC in 2023. His research mainly focuses on quantitative risk management and optimal (re)insurance.

References

[1] Kelly J. A new interpretation of information rate. *Bell System Technical Journal*, **1956**, *35* (4): 917–926.
[2] Latané H. Criteria for choice among risky ventures. *Journal of Political Economy*, **1959**, *67* (2): 144–155.
[3] Algoet P, Cover T. Asymptotic optimality and asymptotic equipartition properties of log-optimum investment. *Annals of Probability*, **1988**, *16* (2): 879–898.
[4] Breiman L. Optimal gambling systems for favorable games. In: Fourth Berkeley Symposium on Mathematical Statistics and Probability. Oakland, CA, USA: University of California Press, **1961**: 65−78.
[5] Ziemba W. The symmetric downside-risk Sharpe ratio. *Journal of Portfolio Management*, **2005**, *32* (1): 108–122.
[6] Luenberger D. Investment Science. Oxford, UK: Oxford University Press, **1998**.
[7] Mossin J. Optimal multiperiod portfolio policies. *The Journal of Business*, **1968**, *41* (2): 215–229.
[8] Hakansson N. On optimal myopic portfolio policies, with and without serial correlation of yields. *The Journal of Business*, **1971**, *44* (3): 324–334.
[9] Boyd S, Vandenberghe L. Convex Optimization. Cambridge, UK: Cambridge University Press, **2004**.
[10] MacLean L, Thorp E, Ziemba W. Good and bad properties of the Kelly criterion. In: The Kelly Capital Growth Investment Criterion: Theory and Practice. Singapore: World Scientific, **2010**: 563−574.
[11] Chopra V, Ziemba W. The effect of errors in means, variances, and covariances on optimal portfolio choice. *Journal of Portfolio Management*, **1993**, *19* (2): 6–11.
[12] Kantorovich L V, Rubinshtein G S. On a space of totally additive functions. *Vestnik of the St. Petersburg University*, **1958**, *13* (4): 52–59.
[13] Esfahani P M, Kuhn D. Data-driven distributionally robust optimization using the Wasserstein metric: Performance guarantees and tractable reformulations. *Mathematical Programming*, **2018**, *171* (1): 115–166.
[14] Gao R. Finite-sample guarantees for Wasserstein distributionally robust optimization: Breaking the curse of dimensionality. *Operations Research*, **2022**, *71* (6): 2291–2306.
[15] Gao R, Kleywegt A. Distributionally robust stochastic optimization with Wasserstein distance. *Mathematics of Operations Research*, **2022**, *48* (2): 603–655.
[16] Gao R, Chen X, Kleywegt A. Wasserstein distributionally robust optimization and variation regularization. *Operations Research*, **2022**, *72* (3): 1177–1191.
[17] Wu Q, Li J Y M, Mao T. On generalization and regularization via Wasserstein distributionally robust optimization. arXiv: 2212.05716, **2022**.
[18] Villani C. Optimal Transport: Old and New. Berlin, Heidelberg: Springer-Verlag, **2009**.
[19] Kuhn D, Esfahani P M, Nguyen V A, et al. Wasserstein distributionally robust optimization: Theory and applications in machine learning. In: Operations Research & Management Science in the Age of Analytics. Catonsville, MD, USA: Institute for Operations Research and the Management Sciences, **2019**: 130−166.
[20] Li J Y M, Mao T. A general Wasserstein framework for data-driven distributionally robust optimization: Tractability and applications. arXiv: 2207.09403, **2022**.
[21] Rujeerapaiboon N, Kuhn D, Wiesemann W. Robust growth-optimal portfolios. *Management Science*, **2016**, *62* (7): 2090–2109.
[22] Sun Q, Boyd S. Distributional robust Kelly gambling. arXiv: 1812.10371, **2018**.
[23] Li J Y M. Wasserstein–Kelly portfolios: A robust data-driven solution to optimize portfolio growth. arXiv: 2302.13979, **2023**.
[24] Hsieh C-H. On solving robust log-optimal portfolio: A supporting hyperplane approximation approach. *European Journal of Operational Research*, **2024**, *313* (3): 1129–1139.
[25] Fama E F. Random walks in stock market prices. *Financial Analysts Journal*, **1965**, *21*: 55–59.
[26] Fama E F. Efficient capital markets: A review of theory and empirical work. *Journal of Finance*, **1970**, *25* (2): 383–417.
[27] Evstigneev I V, Schenk-Hoppé K R. From rags to riches: On constant proportions investment strategies. *International Journal of Theoretical and Applied Finance*, **2002**, *5*: 563–573.
[28] Föllmer H, Schied A. Stochastic Finance: An Introduction in Discrete Time. Third edition. Berlin: Walter de Gruyter, **2011**.
[29] Bellini F, Klar B, Müller A, et al. Generalized quantiles as risk measures. *Insurance: Mathematics and Economics*, **2014**, *54*: 41–48.

CSTR: 32290.14.JUSTC-2024-0030
DOI: 10.52396/JUSTC-2024-0030
JUSTC, **2025**, 55(8): 0805