Mathematical Foundations of Computing

doi:10.3934/mfc.2023043

DISTRIBUTIONALLY ROBUST SPARSE PORTFOLIO OPTIMIZATION MODEL UNDER SATISFACTION CRITERION

Zhongyan Wang[1,2,3], Xiaodong Zhu*[1,2,3], Shaojian Qu[1,2,3], M. Faisal Nadeem[4] and Beibei Zhang[5]

[1]The Research Institute for Risk Governance and Emergency Decision-Making
Nanjing University of Information Science and Technology, Nanjing 210044, China

[2]School of Management Science and Technology
Nanjing University of Information Science and Technology, Nanjing 210044, China

[3]Collaborative Innovation Center on Forecast
and Evaluation of Meteorological Disasters(CIC-FEMD)
Nanjing University of Information Science and Technology, Nanjing, China

[4]Department of Mathematics, COMSATS University Islamabad
Lahore Campus, Lahore 54000, Pakistan

[5]School of Economics and Management, Anhui Jianzhu University, Hefei, China

Abstract. We propose a distributionally robust portfolio optimization model with cardinality constraints under the satisfaction criterion. We aim to maximize the probability of achieving the target return of the proposed portfolio selection model while the number of assets the investors hold is limited. For practical significance, we cite a measure of shortfall-aware aspiration level to the portfolio optimization problem and convert it into a CVaR measure. In our model, we consider a worst-case and assume the distribution of returns of assets is ambiguous. We reformulate the CVaR-based measure equivalently to semi-definite programming for its tractability. A Benders' decomposition algorithm is designed to solve the proposed model efficiently. Numerical tests are utilized through actual market data to validate the proposed method. The results indicate that our algorithm can effectively solve the proposed model, and the sparse portfolio selection model under the satisfaction criterion achieves high robustness and perform better than classical models. Furthermore, we prove that taking the number of assets as the decision variable is a much more efficient method.

1. **Introduction.** The satisfaction criterion has played a significant role in many fields, especially in the area of decision making, see Zopounidis et al. [37] and Yager et al. [33]. The classical portfolio selection model aim to maximize the expected return to attain optimality criterion, which can be unrealistic. Faced with various complicated, uncertain situations, the limitation of information restricts the scope of choice and the specific grasp of the state of investors. Therefore, it is more practical for investors to choose satisfaction criteria to make decisions. In fact, in the investment territory, the decisions that investors make are often

2020 *Mathematics Subject Classification.* Primary: 90C17, 90C15; Secondary: 90C29.

Key words and phrases. Satisfaction criterion, sparse portfolio selection, distributionally robust optimization, Benders' decomposition.

*Corresponding author: Xiaodong Zhu.

simplified as constructing a satisfactory portfolio. The classical portfolio selection model aims to maximize the expected return to attain optimality criterion, which can be unrealistic. Faced with complicated and uncertain situations, the limitation of information restricts the scope of choice and the specific grasp of the state of investors. Therefore, it is more practical for investors to choose satisfaction criteria to make decisions. Lanzillotti [21] proposed that large companies usually have a pricing goal related to a long-range profit horizon, and their goals drive their decisions. The satisfaction criterion reflects the certainty preference in the decision-making process through the certainty that satisfies the requirements of investors. Decision makers prefer to set up a prespecified profit target, and their portfolio selection strategies are driven by their goals, (e.g. the work of [2] and [19]).

Bordley [5] proposed selecting an action that maximizes the probability of meeting an uncertain target. In daily decision-making, decision-makers generally focus on the level of aspiration, or expected level in other words. We further extend the satisfaction criterion, and the emphasis is placed on maximizing the probability of reaching the aspiration level. Hence aspiration-level plays a vital role. However, achieving an aspiration level is generally not a computationally tractable model. As such, a target-based method is usually applied to simple problems such as newsvendor problem, (e.g. the work of Parlar [26] and Weng , Chen and Sim [7] , and Chen, Long and Perakis [6]). We aim to bring a new point of view to portfolio selection, see Chen [7]. And we will offer a more practical approach in studying portfolio optimization problems that can attract investors who take the satisfaction criterion as investment criteria.

In the practical investment process, in addition to setting an investment target, investors usually prefer a lower transaction cost and adjustment frequency when they allocate their assets. The optimal solution of the classical mean-variance model is prone to produce too many small weights in the process of portfolio optimization, and these small weights need more maneuverability and may cause extra management costs. At the same time, there is also the problem of excessive weight, so the risk must be more effectively diversified. Sparse portfolio selection is gradually developed by using sparse optimization theory under such background.

The theory of sparse portfolio selection has been widely studied in recent decades. According to sparse portfolio theory, investors can avoid extreme investment weight and effectively reduce the difficulty of portfolio management and transaction costs by controlling the number of risk assets selected in the portfolio. A sparse portfolio can help investors choose the optimal investment strategy in the risk-return framework and better cope with various changes in the financial market. In Ledoit and Wolf's study [22], norm-constrained portfolios often perform better on the Sharpe ratio than other classical portfolio strategies. DeMiguel et al. [12] discussed several construction methods of sparse combination and tested the validity out of sample. Giuzio [14] proposed a new criterion based on q-entropy and improved the sparsity and robustness of the model.

In general, sparse optimization problems usually involve a cardinality constraint which can be difficult to deal with. Many researchers have applied L_p regularization term to handle this problem. Ning et al. [4] provide a necessary condition and a sufficient condition such that any k-sparse vector x can be recovered from Ax via $L_1 - L_2$ This approach is also widely used in the field of sparse portfolios. (e.g. the work of Lai et al. [20], Fastrich et al. [13], Zhao, Xu, Wang and Zhang [35] and Corsaro et al. [10]). Considering the characteristic of decision variables in portfolio optimization problems, some researchers proposed that the cardinality constraint in sparse portfolio problems can be dealt with by introducing a binary variable. Huang et al. [17] introduced a binary variable to convert the cardinality constraint into a tractable condition. Kobayashi et al. [18] introduced the feasible set for a vector composed of binary decision variables for selecting assets, corresponding to the cardinality constraint. Motivated by existing investigation, we bring in a new binary variable to deal with sparsity and improve the tractability. We also include a transaction cost in our proposed model.

The robust optimization (RO) method overcomes some obstacles under uncertainties. In many complicated situations, we can only use historical data to obtain partial information on the return distribution. Under such circumstances, we introduce the distributionally robust optimization (DRO). In distributionally robust optimization, optimal solutions are evaluated under the worst-case expectation with respect to a set of probability distributions of uncertain parameters. The ambiguity set is used to describe the uncertainty of the parameter and is often moment-based, see Zymler et al. [38], Hanasusanto et al. [16] and Nie et al. [25]. Moment-based ambiguity sets, which contain probability distributions of asset returns whose moments satisfy certain conditions, are commonly adopted in distributionally robust portfolio optimization. Ling et al. [24] constructed a robust active portfolio model with a downside risk measure under ellipsoidal and polytopic fuzzy sets, assuming that the mean value and the covariance matrix are known. The work of Xu, Wang and Dai [32] provided two kinds of safe approximation under different moment-based distribution sets via distributionally robust optimization. Our work focus on the distributionally robust portfolio optimization model that is based on the moment-based ambiguity set developed by Delage and Ye [11] with a cardinality constraint for limiting the number of invested assets.

Based on the above literature, we innovatively apply the measure of shortfall-aware aspiration to the sparse portfolio selection model. A shortfall-aware aspiration criterion overcomes the deficiencies of the proposed aspiration-level criterion, see Chen [7]. We aim to maximize the proposed criterion, which is equivalent to a CVaR measure (conditional value-at-risk). CVaR was put forward by Rockafellar and Uryasev [28] to evaluate the risk of fluctuations of an uncertain cash flow. Considering minimizing a CVaR measure under stochastic programming has already been studied by many scholars. Zhu and Fukushima [36] considered minimizing the worst-case CVaR under various ambiguity sets and utilized it to the robust portfolio optimization. Sarin, Sherali, and Liao [29] aim to minimize CVaR for stochastic scheduling problems and transform it into a mixed-integer programming model, demonstrating the effectiveness of minimizing CVaR. Therefore, the proposed sparse portfolio optimization problem under the satisfaction criterion is converted to a worst-case CVaR measure with cardinality constraints. Due to the moment-based ambiguity set and cardinality constraint, our portfolio optimization model is formulated as a mixed 0-1 integer semidefinite optimization problem, which is a computationally challenging optimization problem. In this case, scholars usually try to transform the problem into a convex optimization problem. Shen et.al [31] considered a modified proximal symmetric ADMMs for multi-block separable convex optimization with linear constraints. To tackle with the mixed 0-1 integer semidefinite program, the existing solver cannot provide a bounded solution, so the Benders' decomposition algorithm of primal-dual problem is pointedly designed in this paper. In order to improve the effectiveness of the algorithm and find the optimal sparsity, this paper also innovatively takes sparsity as a decision variable and compares the running time of the improved algorithm.

The paper is organized as follows. In Section 2, the measure of shortfall-aware aspiration level is introduced, and we show that the satisfaction criterion optimization problem can be transformed into a stochastic optimization problem sequence with CVaR objectives. In Section 3, we construct a sparse portfolio selection model with cardinality constraint and convert it to a mixed 0-1 integer model. Using techniques in distributionally robust optimization, we consider the reformulation of sparse portfolio selection with moment uncertainties. In Section 4, we estimate the covariance matrix in the ambiguity we construct in Section 3. Section 5 proposes an effective algorithm, and Section 6 demonstrates some computational results. Finally, Section 7 concludes our work.

Notations. A symbol with the tilde sigh, i.e.$\widetilde{\xi}$, denotes the random variable. Upper case bold letters such as $\mathbf{M}$ denotes a matrix and $\mathbf{M}^T$ denotes the transpose of the matrix. Lower case bold letters such as $\mathbf{x}$ represent vectors and $\mathbf{x}^T$ denote the transpose of the vector. $\mathbf{1}$ denotes a vector for which all of its elements are 1 and $\mathbf{0}$ in a similar way. In addition, $x^+ = \max\{x, 0\}$ and $x^- = \min\{x, 0\}$. $\langle A, B\rangle_F$ denotes the Frobenius inner product.

2. **The sparse portfolio selection model under satisfaction criterion.**

This section proposes a sparse portfolio selection model under the satisfaction criterion. The classical portfolio selection model aims to attain a return maximation or risk minimization. In such a model, the decision maker has to select the utility function or determine the suitable parameter for the mean risk functional. This largely depends on the decision-maker's subjective choices and can be hard to determine in practice. In most scenarios, ensuring that the investors gain the required excess return with high reliability is of great significance. In this case, it satisfies investors. Additionally, to restrict transaction costs, we augment a carnality constraint on the number of assets. Because of these, we establish a sparse portfolio selection model under investor satisfaction criteria with the cardinality constraint.

2.1. **Preliminaries.**

Before establishing the model, we introduce the following notations. We assume that an investor plans to invest in n risky assets and $\mathbf{x} = (x_1, \ldots, x_n)^T \in \mathcal{X} \subseteq \mathbb{R}^n$ denotes the n dimensional weight vector of each asset. As the return of risky assets is unpredictable in real markets, we denote a random vector $\widetilde{\xi} = \left(\widetilde{\xi}_1, \ldots, \widetilde{\xi}_n\right)^T \subseteq \mathbb{R}^n$ defined on a probability space $(\Xi,\ \mathfrak{F},\ \mathbb{P})$ and $\widetilde{\xi}_i$ represents the individual return rate of asset i in the target portfolio. Ξ represents a state-space, and σ-algebra F indicates the events in Ξ. P represents the exact distribution of $\widetilde{\xi}_i$. $\mathbb{P}$ represents the exact distribution of $\widetilde{\xi}$. $f\left(\mathbf{x}, \widetilde{\xi}\right)$ stands for the overall return depending on asset weights and random variables $\widetilde{\xi}$. We denote $\kappa\left(\widetilde{\xi}\right)$ as the investment target that the investors aim to realize. In most cases, we don't exactly know the target, or aspiration level. Therefore, we relax the assumption of a known target and we denote a target level function $\kappa\left(\widetilde{\xi}\right)$ by a random variable $\widetilde{\xi}$. $P(\bullet)$ denotes the probability of achieving the target.

$$\begin{aligned}
&\max_{\mathbf{x}} \; P\left(f\left(\mathbf{x},\widetilde{\xi}\right)-\kappa\left(\widetilde{\xi}\right)\geq 0\right)\\
s.t.&\sum_{i=1}^{n} x_i = R_0,\\
&x_i \geq 0,\; i=1,\ldots,n,\\
&\|\mathbf{x}\|_0 \leq K,\quad \mathbf{x}\in\mathbb{R}^n.
\end{aligned}\tag{1}$$

$\sum_{i=1}^{n} x_i = R_0$ denotes that the total wealth invested on the assets are equal to R_0. We put a short-selling restriction on risky assets, i.e. the weight of each asset x_i is no less than 0. $\|\mathbf{x}\|_0$ is a representation of a norm, denoting the number of non-zero elements in $\mathbf{x}$. $\|\mathbf{x}\|_0 \leq K$ implies that the number of risky assets to be invested is no more than K. Model (1) evaluates the chance of achieving the overall return target under investment constraints.

The investors aim to achieve the requirement with high probability. However, when setting a goal, we should not only consider the excess value above the target level ,but also the probability of shortfall from the target. Also, the possibilities of large losses and small disturbances differs. Based on these motivations, we consider a shortage perception expectation level criteria used in Chen et al. [7].

Definition 2.1. Let $\widetilde{\ddagger}$ be the difference between the actual return and the target which satisfies the following properties:

$$\begin{aligned}
&E\left(\widetilde{\ddagger}\right) > 0,\\
&P\left(\widetilde{\ddagger} < 0\right) > 0,
\end{aligned}\tag{2}$$

the shortfall-aware aspiration-level criterion is defined as:

$$\alpha\left(\tilde{z}\right) \triangleq \sup_{a>0}\left(E\left(\min\left\{\tilde{z}/a, 1\right\}\right)\right)\tag{3}$$

Here, $E\left(\widetilde{\ddagger}\right)$ is the expected value of $\widetilde{\ddagger}$ and $P\left(\widetilde{\ddagger} < 0\right)$ denotes the probability of not achieving the target. Under these two assumptions, we denote the shortfall-aware aspiration-level criterion by $\alpha\left(\widetilde{\ddagger}\right)$ in formula (3).

We now apply the shortfall-aware aspiration-level criterion into our portfolio selection model. As defined in model (1), the difference between the actual return and the investment target can be expressed as $f\left(\mathbf{x},\widetilde{\xi}\right)-\kappa\left(\widetilde{\xi}\right)$.Next, we present the properties of $\alpha\left(\left(f\left(\mathbf{x},\widetilde{\xi}\right)-\kappa\left(\widetilde{\xi}\right)\right)\right)$ in Theorem 2.1.

Theorem 2.2. *Given an investment target* $\kappa\left(\widetilde{\xi}\right)$, $f\left(\mathbf{x},\widetilde{\xi}\right)$ *stands for the overall return, the shortfall-aware aspiration-level criterion* $\alpha\left(\left(f\left(\mathbf{x},\widetilde{\xi}\right)-\kappa\left(\widetilde{\xi}\right)\right)\right)$ *satisfies the following properties:*

(i)$\alpha\left(\left(f\left(\mathbf{x},\widetilde{\xi}\right)-\kappa\left(\widetilde{\xi}\right)\right)\right)\leq P\left(\left(f\left(\mathbf{x},\widetilde{\xi}\right)-\kappa\left(\widetilde{\xi}\right)\right)\geq 0\right)$.

(ii)$\alpha\left(\left(f\left(\mathbf{x},\widetilde{\xi}\right)-\kappa\left(\widetilde{\xi}\right)\right)\right)\in(0,1)$.

(iii)$\alpha\left(\left(f\left(\mathbf{x},\widetilde{\xi}\right)-\kappa\left(\widetilde{\xi}\right)\right)\right) = E\left(\min\left\{\frac{1}{a^*}\left(f\left(\mathbf{x},\widetilde{\xi}\right)-\kappa\left(\widetilde{\xi}\right)\right),1\right\}\right)$ *with a finite* $a^*>0$.

(iv)$\alpha\left(\widetilde{\ddagger}\right) = \sup_\gamma\left\{1-\gamma : CVaR_{1-\gamma}\left(\kappa\left(\widetilde{\xi}\right)-f\left(\mathbf{x},\widetilde{\xi}\right)\right)\leq 0,\gamma\in(0,1)\right\}$, *where*

$$CVaR_{1-\gamma}\left(\widetilde{v}\right)\triangleq\min_\theta\left(\theta+\frac{E\left(\left(\widetilde{v}-\theta\right)^+\right)}{\gamma}\right)\tag{4}$$

Proof of Theorem 2.2. (i) We first consider a Heaviside utility function $\mathcal{H}(\bullet)$ defined as:

$$\mathcal{H}(x)=\begin{cases}1, & x\geq 0,\\ 0, & otherwise.\end{cases}$$

For all $a>0$, $\min\left\{\frac{x}{a},1\right\}\leq\mathcal{H}(x)$. Hence, we have:

$$\begin{aligned}&P\left(\left(f\left(\mathbf{x},\widetilde{\xi}\right)-\kappa\left(\widetilde{\xi}\right)\right)\geq 0\right)\\ =& E\left(\mathcal{H}\left(\left(f\left(\mathbf{x},\widetilde{\xi}\right)-\kappa\left(\widetilde{\xi}\right)\right)\right)\right)\\ \geq& \sup_{a>0}\left(E\left(\min\left\{\frac{1}{a}\left(f\left(\mathbf{x},\widetilde{\xi}\right)-\kappa\left(\widetilde{\xi}\right)\right),1\right\}\right)\right)\\ =& \alpha\left(\left(f\left(\mathbf{x},\widetilde{\xi}\right)-\kappa\left(\widetilde{\xi}\right)\right)\right).\end{aligned}$$

(ii) In Definition 2.1, we have $P\left(\left(f\left(\mathbf{x},\widetilde{\xi}\right)-\kappa\left(\widetilde{\xi}\right)\right)<0\right)>0$, It's easy to figure out:

$$\begin{aligned}&\alpha\left(\left(f\left(\mathbf{x},\widetilde{\xi}\right)-\kappa\left(\widetilde{\xi}\right)\right)\right)\\ \leq& P\left(\left(f\left(\mathbf{x},\widetilde{\xi}\right)-\kappa\left(\widetilde{\xi}\right)\right)\geq 0\right)\\ =& 1-P\left(\left(f\left(\mathbf{x},\widetilde{\xi}\right)-\kappa\left(\widetilde{\xi}\right)\right)<0\right)<1.\end{aligned}$$

According to the definition of $\alpha\left(\widetilde{\ddagger}\right)$, it is obvious that $\alpha\left(\left(f\left(\mathbf{x},\widetilde{\xi}\right)-\kappa\left(\widetilde{\xi}\right)\right)\right)>0$. To sum up, we have $\alpha\left(\left(f\left(\mathbf{x},\widetilde{\xi}\right)-\kappa\left(\widetilde{\xi}\right)\right)\right)\in(0,1)$.

(iii) Observe that for all $a>0$, $E\left(\min\left\{\frac{1}{a}\left(f\left(\mathbf{x},\widetilde{\xi}\right)-\kappa\left(\widetilde{\xi}\right)\right),1\right\}\right)$
$=\frac{1}{a}E\left(\min\left\{\left(f\left(\mathbf{x},\widetilde{\xi}\right)-\kappa\left(\widetilde{\xi}\right)\right),a\right\}\right)$
$=\ \frac{1}{a}\left(E\left(\left(f\left(\mathbf{x},\widetilde{\xi}\right)-\kappa\left(\widetilde{\xi}\right)\right)\right)+E\left(\min\left\{0,a-\left(f\left(\mathbf{x},\widetilde{\xi}\right)-\kappa\left(\widetilde{\xi}\right)\right)\right\}\right)\right).$
According to the properties of the function $E\left(\min\left\{0,a-\left(f\left(\mathbf{x},\widetilde{\xi}\right)-\kappa\left(\widetilde{\xi}\right)\right)\right\}\right)$.
We have $E\left(\min\left\{0,a-\left(f\left(\mathbf{x},\widetilde{\xi}\right)-\kappa\left(\widetilde{\xi}\right)\right)\right\}\right)\to 0$, if $a\to\infty$, Hence, there exists a constant $b>0$, such that:
$E\left(\left(f\left(\mathbf{x},\widetilde{\xi}\right)-\kappa\left(\widetilde{\xi}\right)\right)\right)+E\left(\min\left\{0,b-\left(f\left(\mathbf{x},\widetilde{\xi}\right)-\kappa\left(\widetilde{\xi}\right)\right)\right\}\right)>0.$

We have: $\alpha\left(\left(f\left(\mathbf{x},\widetilde{\xi}\right)-\kappa\left(\widetilde{\xi}\right)\right)\right)$
$=\sup\limits_{a>0}\left(E\left(\min\left\{\frac{1}{a}\left(f\left(\mathbf{x},\widetilde{\xi}\right)-\kappa\left(\widetilde{\xi}\right)\right),1\right\}\right)\right)$
$\geq E\left(\min\left\{\frac{1}{b}\left(f\left(\mathbf{x},\widetilde{\xi}\right)-\kappa\left(\widetilde{\xi}\right)\right),1\right\}\right)$
$=\frac{1}{b}\left(E\left(\left(f\left(\mathbf{x},\widetilde{\xi}\right)-\kappa\left(\widetilde{\xi}\right)\right)\right)+E\left(\min\left\{0,b-\left(f\left(\mathbf{x},\widetilde{\xi}\right)-\kappa\left(\widetilde{\xi}\right)\right)\right\}\right)\right).$

To further prove that the supremum is achievable, we note that
$P\left(\frac{1}{a}\left(f\left(\mathbf{x},\widetilde{\xi}\right)-\kappa\left(\widetilde{\xi}\right)\right)<0\right)>0$ is equivalent to
$E\left(\left(f\left(\mathbf{x},\widetilde{\xi}\right)-\kappa\left(\widetilde{\xi}\right)\right)^{-}\right)>0.$

Hence:
$\lim\limits_{a\to 0}E\left(\min\left\{\frac{1}{a}\left(f\left(\mathbf{x},\widetilde{\xi}\right)-\kappa\left(\widetilde{\xi}\right)\right),1\right\}\right)$
$=1+\lim\limits_{a\to 0}\frac{1}{a}\left(\min\left\{\left(f\left(\mathbf{x},\widetilde{\xi}\right)-\kappa\left(\widetilde{\xi}\right)\right)-a,0\right\}\right)$
$\leq 1+\lim\limits_{a\to 0}\frac{1}{a}\left(\min\left\{\left(f\left(\mathbf{x},\widetilde{\xi}\right)-\kappa\left(\widetilde{\xi}\right)\right),0\right\}\right)$
$=1-\lim\limits_{a\to 0}\frac{1}{a}E\left(\left(f\left(\mathbf{x},\widetilde{\xi}\right)-\kappa\left(\widetilde{\xi}\right)\right)^{-}\right)=-\infty$

Also, $\lim\limits_{a\to 0}E\left(\min\left\{\frac{1}{a}\left(f\left(\mathbf{x},\widetilde{\xi}\right)-\kappa\left(\widetilde{\xi}\right)\right),1\right\}\right)=0.$

From the proof above, we can observe that the supremum cannot be achieved at $a=0$ or $a=\infty$. Hence, he supremum is achieved at a finite $a>0$. That is to say, there exists a finite $a^*>0$, such that $\alpha\left(\widetilde{\xi}\right)=E\left(\min\left\{\frac{1}{a^*}\left(f\left(\mathbf{x},\widetilde{\xi}\right)-\kappa\left(\widetilde{\xi}\right)\right),1\right\}\right)$.

(iv) Using the observation in (ii), we have:
$\sup\limits_{a>0}E\left(\min\left\{\frac{1}{a}\left(f\left(\mathbf{x},\widetilde{\xi}\right)-\kappa\left(\widetilde{\xi}\right)\right),1\right\}\right)$
$=\sup\limits_{\theta<0}\left(1+\frac{1}{\theta}E\left(\left(\kappa\left(\widetilde{\xi}\right)-f\left(\mathbf{x},\widetilde{\xi}\right)-\theta\right)^{+}\right)\right)$

$$= \sup_{\gamma,\theta} \left\{ 1-\gamma : 1 + \tfrac{1}{\theta} E\left(\left(\kappa\left(\widetilde{\xi}\right) - f\left(\mathbf{x},\widetilde{\xi}\right) - \theta \right)^{+} \right) \leq 0, \theta < 0, \gamma \in (0,1) \right\}$$
$$= \sup_{\gamma,\theta} \left\{ 1-\gamma : \theta + \tfrac{1}{\gamma} E\left(\left(\kappa\left(\widetilde{\xi}\right) - f\left(\mathbf{x},\widetilde{\xi}\right) - \theta \right)^{+} \right) \leq 0, \theta < 0, \gamma \in (0,1) \right\}$$
$$= \sup_{\gamma,\theta} \left\{ 1-\gamma : \theta + \tfrac{1}{\gamma} E\left(\left(\kappa\left(\widetilde{\xi}\right) - f\left(\mathbf{x},\widetilde{\xi}\right) - \theta \right)^{+} \right) \leq 0, \gamma \in (0,1) \right\}$$

with $E\left(\left(f\left(\mathbf{x},\widetilde{\xi}\right) - \kappa\left(\widetilde{\xi}\right) \right)^{-} \right) > 0, \theta < 0$ is implied

$$= \sup_{\gamma} \left\{ 1-\gamma : CVaR_{1-\gamma}\left(\kappa\left(\widetilde{\xi}\right) - f\left(\mathbf{x},\widetilde{\xi}\right) \right) \leq 0, \gamma \in (0,1) \right\}. \qquad \square$$

We learn from Theorem 2.2 (i) that there exists an optimal $\mathbf{x}^*$ achieving the following inequality:

$$P\left(f\left(\mathbf{x}^*,\widetilde{\xi}\right) \geq \kappa\left(\widetilde{\xi}\right) \right) \geq \alpha\left(f\left(\mathbf{x}^*,\widetilde{\xi}\right) - \kappa\left(\widetilde{\xi}\right) \right)$$

In combination with Theorem 2.2 (iii), there is an optimal parameter $\mathbf{a}^*$ that attains the tightest bound in meeting the success probability of achieving the target return. Hence, we achieve an approximation to the original target function. From Theorem 2.2 (iv), we can convert the formula (3) into a CVaR measure, which leads to tractability of the satisfaction criterion optimization model. CVaR measure is often used to evaluate decision alternatives with respect to the downside (upside) risk, popularized by Rockafellar and Uryasev [28].

Next, Definition 2.3 presents the definition of a coherent risk measure, which is derived from the work of Artzner et al. [1] directly. Also, CVaR is a coherent risk measure regardless of whether the variables are normally distributed or not.

Definition 2.3. Let $\mathcal{G}$ the set of all risk functions defined on Ω. $\varphi:\ \mathcal{G} \to \mathbb{R}$ denotes a risk measurement. A risk measure $\varphi$is coherent satisfying the four axioms listed below:

(i) Translation invariance: For all $\theta \in \mathbb{R}$, $\varphi\left(\widetilde{\zeta} + \theta\right) = \varphi\left(\widetilde{\zeta}\right) + \theta$, $\widetilde{\zeta} \in \mathcal{G}$.

(ii) Subadditivity: For all $\widetilde{\zeta}_1, \widetilde{\zeta}_2 \in \mathcal{G}$, $\varphi\left(\widetilde{\zeta}_1 + \widetilde{\zeta}_2\right) \leq \varphi\left(\widetilde{\zeta}_1\right) + \varphi\left(\widetilde{\zeta}_2\right)$.

(iii) Positive homogeneity: For all $\lambda \geq 0$ and $\widetilde{\zeta} \in \mathcal{G}$, $\varphi\left(\lambda\widetilde{\zeta}\right) = \lambda\varphi\left(\widetilde{\zeta}\right)$.

(iv) Monotonicity: For all $\widetilde{\zeta}_1, \widetilde{\zeta}_2 \in \mathcal{G}$, with $\widetilde{\zeta}_1 \leq \widetilde{\zeta}_2$, $\varphi\left(\widetilde{\zeta}_1\right) \leq \varphi\left(\lambda\widetilde{\zeta}\right) = \lambda\varphi\left(\widetilde{\zeta}\right)\left(\widetilde{\zeta}_2\right)$.

Artzner et al. [1] proposed that a measure of risk satisfying the four axioms in Definition 2.3 is called coherent. Translation invariance implies that $\varphi\left(\widetilde{\zeta} - \varphi\left(\widetilde{\zeta}\right)\right) = 0$. That is to say, the risk can be spread out to zero by $\varphi\left(\widetilde{\zeta}\right)$. Subadditivity can be interpreted that consolidation of assets does not increase risk. Positive homogeneity axiom means that when the underlying random variable increases or decreases, the risk measure increases or decreases proportionally. Monotonicity implies that a larger variable is accompanied by a higher risk measure. The risk measure CVaR satisfying the above axioms has a preservation of convexity. Thus function $CVaR_{1-\gamma}\left(\kappa\left(\tilde{\xi}\right)-f\left(\mathbf{x},\tilde{\xi}\right)\right)$ is convex on $\mathbf{x}$. In Section 2.2, we will use the connection to express the sparse portfolio optimization model under satisfaction criterion.

2.2. **Sparse portfolio optimization model under satisfaction criterion.** Under the shortfall-aware aspiration-level criterion, our probability objective function is converted to $\alpha(\bullet)$ defined in model (3).

The satisfaction criterion sparse portfolio optimization model is presented as follows:

$$\begin{aligned} &\max_{\mathbf{x}} \alpha\left(f\left(\mathbf{x},\tilde{\xi}\right)-\kappa\left(\tilde{\xi}\right)\right) \\ s.t. &\sum_{i=1}^{n} x_i = R_0, \\ &x_i \geq 0,\ i=1,\ldots,n, \\ &\|\mathbf{x}\|_0 \leq K, \quad \mathbf{x}\in\mathbb{R}^n. \end{aligned} \tag{5}$$

Utilizing the correlation between $\alpha(\bullet)$ and the CVaR measure in Theorem 2.2 (iv), our proposed satisfaction criterion optimization model can be equivalently expressed as follows:

$$\begin{aligned} &\max_{\gamma,\mathbf{x}} 1-\gamma \\ s.t. &CVaR_{1-\gamma}\left(\kappa\left(\tilde{\xi}\right)-f\left(\mathbf{x},\tilde{\xi}\right)\right)\leq 0 \\ &\sum_{i=1}^{n} x_i = R_0, \\ &x_i \geq 0,\ i=1,\ldots,n, \\ &\|\mathbf{x}\|_0 \leq K, \quad \mathbf{x}\in\mathbb{R}^n \\ &\gamma\in(0,1). \end{aligned} \tag{6}$$

where $CVaR_{1-\gamma}\left(\tilde{\mathbf{v}}\right)=\min_{\theta}\left(\theta+\frac{E\left(\left(\tilde{\mathbf{v}}-\theta\right)^{+}\right)}{\gamma}\right)$.

When γ is fixed the CVaR measure in (6) is convex in the decision variable $\mathbf{x}$. However, convexity does not jointly hold in both γ and $\mathbf{x}$. In this case, we can still obtain optimal solution of the form of stochastic optimization problems with CVaR objectives as follows:

$$\begin{aligned}
&\min_{\mathbf{x}} CVaR_{1-\gamma}\left(\kappa\left(\widetilde{\xi}\right) - f\left(\mathbf{x}, \widetilde{\xi}\right)\right) \\
s.t.&\sum_{i=1}^{n} x_i = R_0, \\
&x_i \geq 0,\ i = 1, \ldots, n, \\
&\|\mathbf{x}\|_0 \leq K, \quad \mathbf{x} \in \mathbb{R}^n \\
&\gamma \in (0, 1).
\end{aligned} \tag{7}$$

where $CVaR_{1-\gamma}\left(\widetilde{\mathbf{v}}\right) = \min_{\theta}\left(\theta + \frac{E\left(\left(\widetilde{\mathbf{v}}-\theta\right)^{+}\right)}{\gamma}\right), \gamma \in (0, 1).$

We will discuss the optimal solution of model (7) in later sections. Next, we discuss the connection between model (6) and model (7) via Algorithm 1. In Algorithm 1, we simplify problem (7) to $CVaR(\gamma)$.

Algorithm 1 Binary search

Input: A routine that solves model (7) optimally and $\varepsilon > 0$.
Output: $\mathbf{x}$ and $CVaR(\gamma)$.
1. Set $\gamma_1 := 0$ and $\gamma_2 := 0$.
2. If $\gamma_2 - \gamma_1 < \varepsilon$, stop. Output x.
3. Let $\gamma := (\gamma_1 + \gamma_2)/2$.
Compute $CVaR(\gamma)$in model (7) and obtain the corresponding optimal solution $\mathbf{x}$.
4. If $CVaR(\gamma) \leq 0$, update $\gamma_2 := \gamma$, otherwise update $\gamma_1 := \gamma$.
5. Go to Step 2.

Proposition 2.4. *Suppose that model (6) is feasible. We obtain a solution* $\mathbf{x}$ *through Algorithm 1 with objective* $1-\gamma^{\dagger}$ *satisfying* $\left|\gamma^{\dagger} - \gamma^{*}\right| < \varepsilon$ *in most* $\lceil \log_2(1/\varepsilon) \rceil$ *computations of the subproblem (7), where* $1 - \gamma^{*}$ *is the optimal objective of model (6).*

Proof. We use binary method in Algorithm 1 to reduce the gap between γ_2 and γ_1 by half in each looping. Suppose that $CVaR(\gamma) \leq 0$, then (6) is satisfied and γ is feasible in model (6), hence $\gamma^{*} \leq \gamma$. Otherwise, γ would be infeasible in model (6). Because the function $CVaR(\bullet)$ is nonincreasing in γ, we have $\gamma^{*} > \gamma$. □

The above model combines satisfaction criteria and sparse portfolio optimization problem. According to above analysis, we consider a shortfall-aware aspiration-level criterion and transform the probabilistic objective function of (1) into a minimizing CVaR problem (7). Model (7) simultaneously includes a sparse constraint on asset weights. Next, we deal with the sparse constraint.

2.3. **A 0-1 mixed integer portfolio selection model under satisfaction criterion.** As we consider the sparse constraint $\|\mathbf{x}\|_0 \leq K$, the problem becomes non-convex and the complexity of the problem has increased dramatically. In fact, we can prove a sparse linear problem is NP-hard. In this paper, we introduce a binary variable a make the problem more tractable. We define a binary variable:

$$y_i := \begin{cases} 1, x_i > 0; \\ 0, x_i = 0, \end{cases} \quad i = 1, \ldots, n,$$

where $y_i = 1$ indicates that asset i is included in the investment strategy and $y_i = 0$ otherwise. By introducing the new binary variable, norm constraint of (7) becomes tractable. The original nonnegative constraint of (7) and norm constraint of (7) are converted into the following programming in Theorem 2.5.

Theorem 2.5. *The mixed-integer 0-1 portfolio problem under satisfaction criterion is listed below:*

$$
\begin{aligned}
&\min_{\mathbf{x},\mathbf{y}}\ CVaR_{1-\gamma}\left(\kappa\left(\tilde{\xi}\right)-f\left(\mathbf{x},\tilde{\xi}\right)\right)\\
s.t.&\sum_{i=1}^{n}x_i=R_0,\\
&l_iy_i\leq x_i\leq u_iy_i,\ \ i=1,\dots n,\\
&\sum_{i=1}^{n}y_i\leq K,\\
&y_i\in\{0,1\}\,,\ \ i=1,\dots,n.
\end{aligned}
\tag{8}
$$

The constraint in (8) simultaneously enforces both minimum and maximum position constraints. To be more specific, if a risky asset i is included in the portfolio, then the holding position must surpass a minimum threshold $l_i \geq 0, i = 1, \dots, n$ as well as being lower than a maximum threshold $u_i \geq 0, i = 1, \dots, n$. The constraint $\sum\limits_{i=1}^{n} y_i \leq K$ indicates that the non-zero quantity of x is less than or equal to K, that is to say, the quantity of assets an investor chooses to invest cannot exceed K. As y is introduced, our problem becomes a mixed 0-1 integer programming with a CVaR objective.

2.4. **A sparse portfolio selection model under satisfaction criterion with transaction cost.** In the practice of portfolio trading, a high turnover rate may lead to a large amount of cost. Transaction costs could significantly affect the amount of assets an investor holds. One of the reasons that we investigate a portfolio selection model is that transaction cost plays a vital role in the portfolio trading process. The results in the work of Chen and Wang [9] indicate as the transaction cost increases, the portfolio diversification often decreases.

The transaction cost function can be expressed as follows:

$$
c\left(\mathbf{x}\right)=\sum_{i=1}^{n}c_i\left(x_i\right)=\sum_{i=1}^{n}{c_i}^T\left|x_i-x_i^0\right|
\tag{9}
$$

We call the above $c(\mathbf{x})$ the proportional transaction cost. $c_i(x_i)$ denotes the transaction cost of the ith asset, determining by the changing weight of the ith asset; c_i denotes the unit transaction cost for ith the asset and we denote by $\left(x_1^0, \dots, x_n^0\right)^T$ the initial holding of the assets at the beginning.

We denote the expected portfolio return by $r_p\left(\mathbf{x}, \tilde{\xi}\right)$ and $f\left(\mathbf{x}, \tilde{\xi}\right)$ indicates the overall return of the portfolio considering transaction costs. We have the following function:

$$
f\left(\mathbf{x},\tilde{\xi}\right)=r_p\left(\mathbf{x},\tilde{\xi}\right)-c\left(\mathbf{x}\right)\ =\mathbf{x}^T\xi-{c_i}^T\left|x_i-x_i^0\right|
\tag{10}
$$

Our proposed sparse portfolio selection model under satisfaction criterion with transaction cost is demonstrated as follows:

$$\begin{aligned}
&\min_{\mathbf{x},\mathbf{y}} CVaR_{1-\gamma}\left(\kappa\left(\tilde{\xi}\right)-\mathbf{x}^T\xi-{c_i}^T\left|x_i-x_i^0\right|\right)\\
s.t.&\sum_{i=1}^{n} x_i=R_0,\\
&l_iy_i\leq x_i\leq u_iy_i,\ i=1,\dots n,\\
&\sum_{i=1}^{n} y_i\leq K,\\
&y_i\in\{0,1\},\ i=1,\dots,n.
\end{aligned}\tag{11}$$

In the next section, we evaluate a worst-case situation where the probability distribution of the random return is not specifically known. Although the worst-case scenario doesn't always appear and may lead to conservative solutions, we consider that most investors are risk averse and prefer to mitigate risks during the whole process of investment.

3. **Distributionally robust counterpart.** In many complex situations, complete probability descriptions of ξ are almost never available. In this case, we don't assume the full knowledge of distribution of the random variable. In an efficient way, we apply the distributionally robust approach to handle with the CVaR objective.

We obtain the following worst-case minimizing CVaR objectives sparse portfolio optimization model:

$$\begin{aligned}
&\min_{\mathbf{x}}\sup_{\mathbb{P}\in\mathcal{P}}\left(\theta+\frac{E_{\mathbb{P}}\left(\left(\kappa\left(\tilde{\xi}\right)-f\left(\mathbf{x},\tilde{\xi}\right)-\theta\right)^+\right)}{\gamma}\right)\\
s.t.&\sum_{i=1}^{n} x_i=R_0,\\
&l_iy_i\leq x_i\leq u_iy_i,\ i=1,\dots n,\\
&\sum_{i=1}^{n} y_i\leq K,\\
&\mathbf{x}\in\mathbb{R}^n,\ \theta\in\mathbb{R}\ ,\ y_i\in\{0,1\},\ i=1,\dots n,
\end{aligned}\tag{12}$$

where $\mathbb{P}$ denotes an ambiguity set, satisfying the true distribution $\mathbb{P}\in\mathcal{P}$.

The choice of set $\mathbb{P}$ greatly affects the complexity of computation. Following the recent development of distributionally robust optimization such as Delage and Ye [11], we construct a moment-based ambiguity set.

We assume that $\Xi=\mathbb{R}^n$. In order to construct a moment-based ambiguity set, we assume the first and second order of the random vector $\tilde{\xi}\in\Xi$ be known. Here, we represent the first moment in terms of the mean vector $\mu=\mathbb{E}_{\mathbb{P}}\left[\tilde{\xi}\right]\in\mathbb{R}^n$.

The second order moment is represented as $\Sigma = \mathbb{E}_{\mathbb{P}}\left[\left(\tilde{\xi} - \mu\right)\left(\tilde{\xi} - \mu\right)^T\right] \in \mathbb{S}^n$. Σ denotes the covariance matrix of the random vector $\tilde{\xi}$. $\mathbb{E}_{\mathbb{P}}[\bullet]$ denotes the expectation under the probability distribution $\mathbb{P}$. We assume $\mu \in \mathbb{R}^n$ and $\Sigma \in \mathbb{S}^n$ be obtained in advance.

Based on the information we discussed before, we construct the following ambiguity set:

$$\mathcal{P} := \left\{\mathbb{P} :\ \mathbb{P}\left(\xi \in \Xi\right) = 1, \mathbb{E}_{\mathbb{P}}\left[\tilde{\xi}\right] = \mu, \mathbb{E}_{\mathbb{P}}\left[\left(\tilde{\xi} - \mu\right)\left(\tilde{\xi} - \mu\right)^T\right] = \Sigma\right\}. \tag{13}$$

We obtain the worst-case satisfaction criterion optimization model under moment estimation and cardinality constraints as follows:

CSCO-ME

$$\begin{aligned} &\min_{\mathbf{x},\theta} \theta + \frac{1}{\gamma} \sup_{\mathbb{P}\in\mathcal{P}} \left\{E_{\mathbb{P}}\left(\left(\kappa\left(\tilde{\xi}\right) - f\left(\mathbf{x}, \tilde{\xi}\right) - \theta\right)^+\right)\right\} \\ s.t.&\sum_{i=1}^n x_i = R_0, \\ &l_i y_i \le x_i \le u_i y_i,\ i = 1, \dots n, \\ &\sum_{i=1}^n y_i \le K, \\ &\mathbf{x} \in \mathbb{R}^n,\ \theta \in \mathbb{R}\ ,\ y_i \in \{0, 1\}\ ,\ i = 1, \dots n, \end{aligned} \tag{14}$$

It is obvious that the CSCO-ME cannot be computationally solved directly, especially the expectation $E\left((\bullet)^+\right)$. We can use the moment information in the ambiguity $\mathcal{P}$ we construct above to develop tractable approximation of the CVaR measure. According to Zymler et al. [38], the worst-case expectation can be transformed into a tractable approximation as proposed in Lemma 3.1.

Lemma 3.1. *Let $g : \mathbb{R}^k \to \mathbb{R}$ be a measurable function, and define the worst-case expectation θ_{WC} as:*

$$\theta_{WC} = \max_{\mathbb{P}\in\mathcal{P}} E_{\mathbb{P}}\left(\left(g\left(\tilde{\xi}\right)\right)^+\right), \tag{15}$$

where $\mathcal{P}$ represents the usual set of all probability distributions $\mathbb{R}^k$ with given mean vector μ and covariance matrix $\mathbf{\Sigma} \succ 0$. Then,

$$\theta_{WC} = \inf_{\mathbf{M}\in\mathbb{S}^{k+1}} \left\{\langle \mathbf{\Omega}, \mathbf{M}\rangle : \mathbf{M} \succ \mathbf{0}, \begin{bmatrix}\xi^T & 1\end{bmatrix} \mathbf{M} \begin{bmatrix}\xi^T & 1\end{bmatrix}^T \ge g(\xi)\right\}, \forall \xi \in \mathbb{R}^k, \tag{16}$$

where $\mathbf{\Omega}$ is the second-order moment matrix of $\tilde{\xi}$, $\mathbf{M} \in \mathbb{S}^{n+1}$. mathbfM is a matrix composed of Lagrange multipliers.

On the basis of the above ambiguity set, we are able to utilize Lemma 3.1 to tackle with model (14). Here, $f\left(\mathbf{x}, \tilde{\xi}\right) = \mathbf{x}^T\xi - {c_i}^T\left|x_i - x_i^0\right|$ represents the overall return of the portfolio including transaction costs. For simplicity of calculation, we assume that the initial holding of each asset $x_i^0 = 0$, thus, ${c_i}^T\left|x_i - x_i^0\right|,\ i = 1, \dots, n$ can be expressed as $\mathbf{c}^T\mathbf{x}$, $\mathbf{c} = (c_1, \dots, c_n)$.

Next, we set up a mixed 0-1 integer semi-definite programming by Theorem 3.2. We will abbreviate semi-definite programming as SDP in subsequent contents.

Theorem 3.2. *The satisfaction criterion sparse portfolio selection model under the set $\mathcal{P}$ is equivalent to the mixed 0-1 integer SDP listed below: WCSCO-ME*

$$\begin{aligned}
&\min_{\mathbf{x},\mathbf{y},\mathbf{z},\mathbf{Z},\theta,z_0} \theta + \frac{1}{\gamma}\left\langle \mathbf{M}, \mathbf{\Omega} \right\rangle_F \\
&s.t.\mathbf{M} + \begin{bmatrix} 0 & \mathbf{x}/2 \\ \mathbf{x}^T/2 & \theta - \kappa - \mathbf{c}^T\mathbf{x} \end{bmatrix} \succ \mathbf{0} \\
&\sum_{i=1}^{n} x_i = R_0,\ \sum_{i=1}^{n} y_i \le K, \\
&l_i y_i \le x_i \le u_i y_i,\ i = 1, \dots n, \\
&\mathbf{x} \in \mathbb{R}^n,\ \theta \in \mathbb{R}\ , M \in \mathbb{S}^{n+1},\ y_i \in \{0,1\}\,,\ i = 1, \dots, n.
\end{aligned} \tag{17}$$

Proof. Similar to the work of Delage and Ye [11], the inner part of model (14) can be converted into an SDP problem by dual theory. Consider the following optimization problem with respect to $\mathbb{P}$,

$$\begin{aligned}
&\sup_{\mathbb{P}} \int_{\xi\in\Xi} \left(\kappa\left(\tilde{\xi}\right) - f\left(\mathbf{x}, \tilde{\xi}\right) - \theta\right)^{+} d\mathbb{P}\left(\xi\right) \\
&s.t. \int_{\xi\in\Xi} d\mathbb{P}\left(\xi\right) = 1, \\
&\int_{\xi\in\Xi} \xi d\mathbb{P}\left(\xi\right) = \mu, \\
&\int_{\xi\in\Xi} \xi\xi^{\mathbf{T}} d\mathbb{P}\left(\xi\right) = \Sigma + \mu\mu^{\mathbf{T}}
\end{aligned} \tag{18}$$

We denote the Lagrangian multipliers by $z_0 \in \mathbb{R}$, $\mathbf{z} \in \mathbb{R}^n$ and $\mathbf{Z} \in \mathbb{S}^n$ corresponding to the separate constraints. The dual form of (18) is formulated as follows:

$$\begin{aligned}
&\inf\ z_0 + \mathbf{z}^{\mathbf{T}}\mu + \left\langle \mathbf{Z}, \mathbf{\Sigma} + \mu\mu^{\mathbf{T}} \right\rangle_F \\
&s.t. z_0 + \mathbf{z}^T\xi + \left\langle \mathbf{Z}, \xi\xi^{\mathbf{T}} \right\rangle_F \ge \left(\kappa\left(\tilde{\xi}\right) - f\left(\mathbf{x}, \tilde{\xi}\right) - \theta\right)^{+}, \forall \xi \in \mathbf{\Xi}, \\
&z_0 \in \mathbb{R}, \mathbf{z} \in \mathbb{R}^n, \mathbf{Z} \in \mathbb{S}^n.
\end{aligned} \tag{19}$$

Let $\mathbf{M} := \begin{bmatrix} \mathbf{Z} & \frac{1}{2}\mathbf{z} \\ \frac{1}{2}\mathbf{z}^T & z_0 \end{bmatrix}$ and $\Omega := \begin{bmatrix} \mathbf{\Sigma} + \mu\mu^{\mathbf{T}} & \mu \\ \mu^{\mathbf{T}} & 1 \end{bmatrix}$.
The problem (19) can be represented to the following SDP problem:

$$\begin{aligned}
&\inf_{\mathbf{M}\in\mathbb{S}^{n+1}} \left\langle \mathbf{M}, \mathbf{\Omega} \right\rangle_F \\
&s.t.\mathbf{M} \succ \mathbf{0} \\
&\begin{bmatrix} \xi^T & 1 \end{bmatrix} \mathbf{M} \begin{bmatrix} \xi^T & 1 \end{bmatrix}^T \ge \kappa\left(\tilde{\xi}\right) + \mathbf{c}^T\mathbf{x} - \mathbf{x}^T\xi - \theta, \forall \xi \in \mathbf{\Xi}
\end{aligned} \tag{20}$$

The constraint of (17) can be rewritten equivalently as:

$\begin{bmatrix} \xi^T & 1 \end{bmatrix} \left(\mathbf{M} + \begin{bmatrix} 0 & \frac{1}{2}\mathbf{x} \\ \frac{1}{2}\mathbf{x}^T & \theta - \kappa\left(\widetilde{\xi}\right) - \mathbf{c}^T\mathbf{x} \end{bmatrix} \right) \begin{bmatrix} \xi^T & 1 \end{bmatrix}^T \geq 0, \forall \xi \in \mathbf{\Xi}$ Since $\mathbf{\Sigma} \succ 0$, the strong duality holds. Thus, $\underset{p\in\mathcal{P}}{sup}\ E_p \left[\left(\kappa\left(\widetilde{\xi}\right) - f\left(\mathbf{x}, \widetilde{\xi}\right) - \theta \right)^+ \right]$ can be expressed equivalently as follows:

$$\begin{aligned}
&\inf_{\mathbf{z},\mathbf{Z},z_0} \langle \mathbf{M}, \mathbf{\Omega} \rangle_F \\
&s.t. \mathbf{M} \succ \mathbf{0}, \begin{bmatrix} \mathbf{Z} & \frac{1}{2}(\mathbf{z}+\mathbf{x}) \\ \frac{1}{2}(\mathbf{z}+\mathbf{x})^T & z_0 + \theta - \kappa\left(\widetilde{\xi}\right) - \mathbf{c}^T\mathbf{x} \end{bmatrix} \succcurlyeq 0, \\
&\begin{bmatrix} \xi^T & 1 \end{bmatrix} \mathbf{M} \begin{bmatrix} \xi^T & 1 \end{bmatrix}^T \geq \kappa\left(\widetilde{\xi}\right) + \mathbf{c}^T\mathbf{x} - \mathbf{x}^T\xi - \theta, \forall \xi \in \mathbf{\Xi} \\
&z_0 \in \mathbb{R}\ , \mathbf{z} \in \mathbb{R}\ ^n, \mathbf{Z} \in \mathbb{S}^n
\end{aligned} \tag{21}$$

If we replace the inner part by problem (14), the CSCO-ME model can be converted to the WCSCO-ME model. □

Clearly, (17) is an SDP program with cardinality constraints. The combination of semi-definite positive programming and cardinality constraints increases the difficulty of directly solving the proposed problems. The degree of difficulty also depends on the quantity of assets. Because the order of the covariance matrix and the other coefficient matrixes will increase with the amount of assets increasing. In Section 5, a Benders' decomposition algorithm is pointedly devised to work out the proposed models. In the following numerical tests, we will prove that our algorithm can efficiently solve both small-scale and large-scale problems.

4. **Contraction estimation of covariance matrix.** In Section 4, we are going to construct the ambiguity set $\mathcal{P}$ by data-driven approach. The estimation of the covariance matrix Σ plays a vital role in constructing $\mathcal{P}$. We can derive an estimation of Σ from historical data by minimizing mean-squared error(MSE).

In mathematical statistics, the mean square error (MSE) is the expected value of the square of the difference between the parameter estimate and the parameter truth value. MSE is a convenient method to measure the mean error. MSE can evaluate the degree of change of data. The smaller the value of MSE, the better the accuracy of the prediction model to describe the experimental data.

Let $X_i \in \mathbb{R}^n, i = 1, \dots, N$ be a sample of independent identical distributed random vectors with zero mean and covariance Σ. In order to minimize the MSE, we need to find a corresponding estimator $\hat{\Sigma}$:

$$E\left\{ \left\| \hat{\Sigma} - \Sigma \right\|_F^2 \right\}. \tag{22}$$

Due to the difficulties in calculation, we consider estimators that employ shrinkage, see Ledoit and Wolf [22].

The unstructured classical estimator of Σ is the sample covariance $\hat{S}$ defined as:

$$\hat{S} = \frac{1}{N}\sum_{i=1}^{N} x_i x_i^T. \tag{23}$$

This estimator is unbiased, satisfying $E\left\{\hat{S}\right\} = \Sigma$, and is also the maximum likelihood solution if $N \geq n$. However, it does not necessarily achieve low MSE due to its high variance and is usually ill-posed when n is large.

Therefore, $\hat{F}$ is proposed as a naive but most well-conditioned estimate for Σ, see Chen et al [8].

$$\hat{F} = \frac{Tr\left(\hat{S}\right)}{n} I, \tag{24}$$

where I denotes an identity matrix. This estimate will lead to reduced variance but increased bias. Chen et al. [8] constructed a convex combination with $\hat{S}$ and $\hat{F}$, achieving both low bias and low variance.

$$\hat{\Sigma} = (1 - \hat{\rho})\,\hat{S} + \hat{\rho}\hat{F}. \tag{25}$$

The estimator $\hat{\Sigma}$ is characterized by the shrinkage coefficient $\hat{\rho}$, $\hat{\rho} \in [0, 1]$. The matrix $\hat{F}$is referred to as the shrinkage target.

We aim to find a shrinkage coefficient $\hat{\rho}$ that minimizes the MSE (1).

Ledoit and Wolf propose the Rao-Blackwell Ledoit-Wolf (RBLW) [23] estimator, which

$$\hat{\rho}_{LW} = \frac{\sum\limits_{i=1}^{N} \left\| x_i x_i^T - \hat{S} \right\|_F^2}{N^2 \left[Tr\left(\hat{S}^2\right) - \frac{Tr^2\left(\hat{S}\right)}{n} \right]}. \tag{26}$$

It can be proved that when $N \to \infty$, $n \to \infty$ and $n/N \to c$, $0 < c < \infty$. $\hat{\rho}_{LW} \to \hat{\rho}^*$, under the situation that the sample distribution is unknown, see Ledoit and Wolf [23]. $\hat{\rho}^*$ denotes the optimal shrinkage coefficient that minimize the MSE.

The following steps present a process for estimating the covariance matrix σ.

Step 1. Input N historical returns of n risky assets.
Step 2. Compute the mean return μ and the covariance matrix σ.
Step 3. Compute the estimators $\hat{F}$ and $\hat{\rho}_{LW}$.
Step 4. Compute the estimation $\hat{\sigma}$ of σ.

In this way, we estimate the covariance matrix using the actual market data through our proposed method. At this point, we already obtain the first and second moments information and construct our ambiguity. Next, we make use of information available so far to design the computationally solvable algorithm.

5. **The Benders' decomposition algorithm.** According to the analysis of the previous sections, we obtain a mixed 0-1 integer semi-definite programming, see WCSCO-ME. The existing solvers for semi-positive definite programming, such as SDPT3, can only solve part of our problem, and may take too much time. To be more specific, when we use Yalmip to solve our problem, there exists the following problems. First, according to the running mode of software, it is difficult for software to match a suitable solver by itself. Second, When we try to specify the solver manually, we always get unbounded solutions. Or the software could not obtain the optimal solution in a limited number of steps. Thus, to dispose our problem more effectively, we apply a dual transformation to the model and propose a Benders' decomposition algorithm.

Section 5 will present an improved Benders' decomposition technique, pointedly designed to figure out the WCSCO-ME model. Benders [3] first proposed a decomposition method to solve mixed-integer programming problem. Based on Bender's research, many scholars modified the Benders' decomposition and proposed methods suitable for different models. More recently, the work of Rahmaniani et al. [27] combined the complementary of both Benders' decomposition method and Lagrangian dual decomposition method and put forward a new and high-performance approach.

The basic principle of our proposed Benders decomposition is to divide the mixed integer semi positive definite programming problem into an integer programming problem and a semi positive definite programming, respectively. In this way, we can effectively get the feasible solutions separately. We set y as a fixed feasible 0-1 integer vector, the problem (27) turns into a semi-definite linear programming. Problems in such a form can be easily solved by existing solvers. According to the duality theory in Shapiro's work [30], we propose the dual form of WCSCO-ME as:

WCSCO-ME -Dual

$$\begin{aligned} & \sup \varpi R_0 + (\mathbf{L}z_2 - \mathbf{U}z_1)^T \tilde{\mathbf{y}} \\ s.t. & \frac{1}{\gamma}\Omega - \mathbf{S} - \Pi \mathbf{0}, \\ & q_0 = 1,\ z_1 - z_2 - \mathbf{q} + \mathbf{c} \succcurlyeq \pi_0 1, \\ & \varpi \in \mathbb{R}\ ,\ \mathbf{S} \in \mathbb{S}_+^{n+1},\ z_1 \in \mathbb{R}_+^n, \\ & z_2 \in \mathbb{R}_+^n,\ \Pi = \begin{pmatrix} \mathbf{Q} & \mathbf{q} \\ \mathbf{q}^T & q_0 \end{pmatrix} \in \mathbb{S}_+^{n+1} \end{aligned} \tag{27}$$

Where Π and S are dual matrix variables related to the constraints of (17), ϖ, z_1 and z_2 are dual variables and dual vectors relevant to with the constraints of (17). $\mathbf{L} := diag\,(l_i)$ and $\mathbf{U} := diag\,(u_i)$, $i = 1, \ldots n$, representing upper and lower bound diagonal matrixes, respectively.

Next, we utilize the partitioning procedures in the work of Delage and Ye [11], the WCSCO-ME model is equivalent to the following model:

$$\begin{aligned} & \min \eta \\ s.t. & \eta \geq \varpi R_0 + (\mathbf{L}z_2 - \mathbf{U}z_1)^T \mathbf{y}, \\ & \sum_{i=1}^{n} y_i \leq K, \\ & \varpi \in \mathbb{R}\ ,\ \mathbf{S} \in \mathbb{S}_+^{n+1},\ z_1 \in \mathbb{R}_+^n, \\ & y_i \in \{0,1\}\ ,\ i = 1, \ldots, n, \eta \in \mathbb{R} \end{aligned} \tag{28}$$

Meanwhile, $\mathbf{y}$ becomes a decision variable and the problem (27) turns into the above mixed-integer 0-1 programming. Next, we demonstrate the pseudo code of our improved Benders' decomposition algorithm as shown in Algorithm 2.

Algorithm 2 Benders' decomposition algorithm for WCSCO-ME

1: Initialization: $k \leftarrow 1, \mathcal{V}_U \leftarrow \infty, \mathcal{V}_L \leftarrow -\infty$, predetermined tolerance $\varepsilon, \beta \in (0, 1]$, W_0. Initial feasible solution $y^1 \in \{0,1\}^n$.
2: Solve semi-definite linear programming problem (27) and derive the optimal solution $\left(\varpi^k, z_1^k, z_2^k, S^k, \Pi^k\right)$ and the optimal value D^k.
3: $\mathcal{V}_U \leftarrow D^k$.
4: while $\mathcal{V}_U - \mathcal{V}_L \leq \varepsilon$ do
5: Solve 0-1 integer programming problem (28) with $\varpi = \varpi^k, z_1 = z_1^k, z_2 = z_2^k$ and derive the optimal value η^k and the optimal solution y^k. $\mathcal{V}_L \leftarrow \max\left\{\eta^k,\ \mathcal{V}_L\right\}$.
6: Solve semi-definite linear programming problem (27) with $y = y^k$ and derive the optimal solution $\left(\varpi^{k+1}, z_1^{k+1}, z_2^{k+1}, S^{k+1}, \Pi^{k+1}\right)$ and the optimal value $D = D^{k+1}$. 7: $\mathcal{V}_U \leftarrow \min\left\{D^{k+1},\ \mathcal{V}_U\right\}$.
8: $k \leftarrow k + 1$.
9: end while
10: return Solve WCSCO-ME with $y = y^k$, and derive the optimal solution x^* and the optimal value v^* .

For further investigation, the sparsity K is taken as a decision variable in step 5. We will prove the improvement of the algorithm effectiveness in the numerical tests. We now prove that at each iteration, the solution of problem (27) and problem (28) achieves a tighter upper bound and lower bound, respectively.

Theorem 5.1. *Assume that the proposed model WCSCO-ME-Dual is feasible for any given $\mathbf{y} \in \{0,1\}^n$. Given $\varepsilon > 0$, Algorithm 2 must stop iterating in a finite step.*

Proof. We first use reduction to absurdity to prove Theorem 5.1. It is assumed that step $k$is not the termination of the algorithm. It is obvious that an optimal solution $\mathbf{y}^{k+1}$ of the problem (28) is found and satisfies the condition that $\mathbf{y}^{k+1}$ is superior to $\mathbf{y}^i$, $i \in \{1, \dots, k\}$. According to the supposed conditions, $\mathbf{y}^t$ is feasible to the mixed-integer programming (11), $t \in \{1, \dots, k\}$, and we can obtain the objective value D^t of problem (27). We can observe that:

$$D^t \geq D^k \geq {\mathcal{V}_U}^k > {\mathcal{V}_L}^k \geq \eta^k. \tag{29}$$

If the optimal solution $\mathbf{y}^{k+1}$ repeats the previous solutions $\mathbf{y}^t, t \in \{1, \dots, k\}$, then we have,

$$\eta^{k+1} \geq D^t,\ t \in \{1, \dots, k\}\,. \tag{30}$$

The above conclusion is contradicted with (29). The algorithm will stop the iteration in a finite step as the primal problem has restricted solutions.

When $k \geq 2$, we observe as follows:

$$\eta^{k-1} \leq \eta^k \leq v^* \leq D^k \leq D^{k+1}. \tag{31}$$

Assuming the algorithm stop the iteration at step k, we have $D^k - \eta^k \leq \varepsilon$, such that $\eta^k = v^* = D^k$. Correspondingly, $(\mathbf{x}^*, \mathbf{y}^*)$ are the optimal solutions selected by the algorithm to the primal problems. $\square$

Table 1. Description of data sets

Data set	n	N	Index	Sample Horizon	Frequency
SSE50	32	581	ShangZheng 50Index	01/01/2017-12/31/2019	Daily
CSI300	125	399	HuShen 300 Index	01/08/2010-12/27-2019	Weekly
DJ30	30	235	Dow Jones Industrial Average Index	01/04/2021-12/31/2021	Daily

6. Numerical tests.

In this section, the satisfaction criterion sparse portfolio selection model will be numerically tested on real market data sets. We select two data sets from Chinese stock market and one data set from American stock market. All of our work is operated on a notebook (Intel i7 CPU) with Matlab R2021b, YALMIP.

6.1. Preparations.

To keep the experiment representative, we conduct the numerical tests are on two indices from Chinese stock market and one indice from American stock market, see Table 1 for details. The SSE50 and the CSI30 are obtained from RESSET and the DJ30 is obtained from Yahoo Finance. We present some fundamental information of the three data sets in Table 1.

In Table 1, n denotes the number of the sample stocks, and N denotes the length of historical data. The SSE50 contains 581 days of daily returns for 32 stocks. The CSI300 contains 399 weeks of weekly return for 125 stocks. And DJ30 is consists of 235 days of daily returns for 30 stocks. Our data set consists of daily and weekly data, which can be more representative.

In order to better carry out the subsequent numerical experiment, we choose the following parameters. Without loss of generality, we set the initial wealth $R_0 = 1$. Considering the intrinsic property of CVaR, we carry out the sensitivity investigation of the confidence level $1-\gamma$, which indicates that the probability of not exceeding specific loss. We set 99% 0.95% and 0.90% as the confidence level $1-\gamma$. We estimate the covariance matrix σ according to the method in Section 4. $c_i = 0.01, i = 1, \ldots, n$, representing the unit transaction cost for the ith stock and $x_0 = 0, i = 1, \ldots, n$.

In the following sections, we will utilize annualized portfolio return and Sharpe ratio to compare the performance of different models under variable parameters. Sharpe ratio is an effective standard to evaluate portfolio performance. Its purpose is to calculate how much excess return a portfolio will generate for each unit of total risk it takes. Sharpe ratio is defined as follows:

$$SR = \frac{R_p - R_f}{\sigma_p}, \tag{32}$$

where R_p denotes the portfolio return, R_f denotes the risk-free rate which is set as 0.03 in this section. σ_p denotes the standard deviation of the portfolio. The higher the Sharpe ratio is, the higher the portfolio's risk-return per unit of risk.

6.2. Data analysis and comparison.

In this subsection, Algorithm 2 is used to solve the proposed satisfaction criterion sparse portfolio selection model for the three data sets we demonstrates above. We first report the Sharpe ratio under different confidence level of the resulting sparse portfolio for CSI300 data set in Table 2. Specifically, we use the parameter K to impose restrictions on the number of the portfolio.

Table 2. Sharpe ratio for the CSI300 data set

K	$\gamma = 0.01$	$\gamma = 0.05$	$\gamma = 0.1$
6	1.7169	1.7201	1.7276
7	1.9015	1.9062	1.9100
8	1.7044	1.9194	1.9232
9	1.7319	1.9174	1.9209
10	1.6939	1.7032	1.7720

Table 3. Selected stocks in CSI300 data set with $\gamma = 0.05$

K	Composition	APR
6	(35,43,91,123,124,125)	0.6575
7	(35,43,70,91,123,124,125)	0.6474
8	(35,40,43,70,91,123,124,125)	0.6558
9	(35,40,43,70,88,91,123,124,125)	0.6558
10	(35,40,43,47,70,88,91,123,124,125)	0.5809

Table 4. Sharpe ratio for the SSE50 data set

K	$\gamma = 0.01$	$\gamma = 0.05$	$\gamma = 0.1$
6	1.2937	1.6610	1.6847
7	2.1216	2.1486	2.1701
8	1.5964	2.2323	2.2559
9	1.6124	2.2378	2.2617
10	2.2739	2.3042	2.3281

We can observe from Table 2 that when the number of the assets K is fixed, the Sharpe ratio increases as γ increases. In addition, Sharpe ratios under different situations are all greater than 1.5, which means that each portfolio performed well with a higher return than volatility risk. To better analyze the character of the portfolio, we investigate the annualized portfolio return and the asset composition of the portfolio of the SSE50 in Table 3. APR denotes the Annualized Portfolio Return. For simplicity, we only list the consequences conducted under $\gamma = 0.05$.

From the results in Table 3, we can observe that as K changes, the 35th(China petroleum & chemical corporation), the 43rd(ICBC: Industrial and Commercial Bank of China), the 123rd(Anyuan Coal Industry Group Corporation), 124th(Better Life Commercial Chain Share Corporation), 125th(Yinchuan Xinhua Commercial Group Corporation) stock are always selected. These five stocks cover various industries, including bank, state-owned enterprise and retail enterprise, reflecting good market trends in this period. In Table 4, we carry out the same work as Table 3 for SSE50 data set.

Observing the results in Table 4, it also follows the rule that when the number of the assets K is fixed, the Sharpe ratio increases as γ increases. When $\gamma = 0.05$ and $\gamma = 0.1$, the Sharpe ratio increases when the number of stocks increase. However, when $\gamma = 0.01$, the Sharpe ratio changes with no law.

In the same way, we investigate the annualized portfolio return of the SSE50 under $\gamma = 0.05$ in Table 5. We can observe that as K changes, the 7th (Shanghai Automotive Corporation), 8th (Shanghai Fosun Pharmaceutical Corporation), 9th

Table 5. Selected stocks in SSE50 data set with $\gamma = 0.05$

K	Composition	APR
6	(7,8,9,10,15,17)	0.3897
7	(5,7,8,10,13,15,17)	0.4112
8	(5,7,8,9,10,13,15,17)	0.4242
9	(5,7,8,9,10,12,13,15,17)	0.4250
10	(4,5,7,8,9,10,12,13,15,17)	0.4268

Table 6. Sharpe ratio for the DJ30 data set

K	$\gamma = 0.01$	$\gamma = 0.05$	$\gamma = 0.1$
6	0.9271	1.3221	1.8462
7	1.9282	1.9897	2.0380
8	1.7532	1.8362	2.12
9	1.1812	1.9292	1.9778
10	2.0337	2.0678	2.0958

Table 7. Selected stocks in DJ30 data set with $\gamma = 0.05$

K	Composition	APR
6	(3,7,13,20,21,24)	0.2872
7	(3,7,9,11,13,20,21)	0.4347
8	(3,6,7,9,11,13,20,21)	0.4012
9	(3,6,7,9,11,13,17,20,21)	0.4088
10	(3,6,7,9,11,13,14,17,18,20)	0.4252

(Jiangsu Hengrui Medicine Corporation), 10th (Kweichow Moutai Corporation), 15th (Inner Mongolia Yili Industrial Group Corporation), 17th (Air China Limited Corporation) are always selected. These six stocks cover various industries, which shows that in this period the Shanghai stock market was generally performing well.

In the following, we will test the performance of the stocks in DJ30 data set. We list the results in Table 6 of the Sharpe ratio of different confidence level with $K = 6, 7, 8, 9, 10$, respectively.

We can see from Table 6, we have the same conclusion the magnitude of γ has a positive effect on the Sharpe ratio. What's more, the Sharp ratio makes a bigger change when γ changes for DJ30 data set. When γ varies, the Sharpe ratio changes with no apparent regularity.

As the same work in Table 5 and Table 6, we find out that the 3rd (The Boeing Company), 7th (General Electric Company) and 13th (Intel Corporation) stocks are selected in all the five scenarios in Table 7. The selected stocks include aerospace industry, electrical appliance industry and semiconductor industry. It suggests that the performance of high-tech industry is better than other stocks in DJIA.

Next, we compare the performance of our proposed model with equal-weight model [12], mean-variance model [15] and sparse mean-variance model [34] in Table 8. Without loss of generality, K is set as 8 and $1 - \gamma$ is set as 0.95.

We compare the proposed model under satisfaction criteria with the traditional portfolio optimization model. Equal-weight model, see [12], ignores the differences in the performance of different assets. When the number of assets is large, it

TABLE 8. Selected stocks in CSI300 data set with $\gamma = 0.05$

Data set	Modle	SR
SSE50	WCSCO-ME	2.2323
	Equal-Weight Model	1.4695
	Mean-Variance Model	1.6936
	Sparse Mean-Variance Model	2.1563
CSI300	WCSCO-ME	1.8940
	Equal-Weight Model	1.4828
	Mean-Variance Model	1.6501
	Sparse Mean-Variance Model	1.8438
DJ30	WCSCO-ME	1.8362
	Equal-Weight Model	1.1085
	Mean-Variance Model	1.6109
	Sparse Mean-Variance Model	1.7737

TABLE 9. Comparison of a fixed K and an optimal K for DJ30 data set

K	SR	Time
6	1.3221	3.003247s
7	1.9897	3.047106s
8	1.8362	2.994480s
9	1.9292	3.074431s
10	2.0678	3.070506s
4^*	2.5327^*	$2.989623s^*$

will produce tiny weights of assets, which is not conducive to the actual investment operation. Mean-variance model in [15] is the classic model of a portfolio. Compared with the equal-weight model, it takes into account the differentiation of assets. However, it usually regards the mean and variance of assets as fixed, which is not in line with the actual investment environment. Sparse mean-variance model [34] took asset weights and asset quantities into account. Data experiments show that it performs better than the traditional model. Compared with above models, our proposed model incorporated the satisfaction criterion into the model and the asset weight is also considered. Our model is more adaptable to complicated investment environments and performs better on numerical indicators. We can conclude that the proposed satisfaction criterion sparse portfolio selection model is superior than the classical models, proving the validity of the WCSCO-ME model.

6.3. **Further investigation.** In the previous section, our research is based on a fixed number of assets. We find that when the confidence level $1-\gamma$ is fixed, the return of the portfolio will change with the change of the number of assets. We have reason to believe there exists an optimal parameter K^* that achieves an optimal Sharpe ratio SR^*.

Therefore, we carry out further study in this subsection. In Algorithm 2, we no longer assume a fixed K. Instead, we set K as a decision variable and drive the algorithm to find an optimal solution. And we compare the running time of the algorithm with K as a decision variable and K as a fixed value. Without loss of generality, the parameter $1-\gamma$ is set as 0.05.

Table 10. Comparison of a fixed K and an optimal K for SSE50 data set

K	SR	Time
6	1.6610	3.104740s
7	2.1486	3.099053s
8	2.2323	3.177687s
9	2.2378	3.199388s
10	2.3042	3.191076s
11*	2.3653*	$2.853323s{*}^{*}$

Table 11. Comparison of a fixed K and an optimal K for CSI300 data set

6	1.7201	344.740864s
7	1.9062	360.449267s
8	1.8940	397.408613s
9	1.8940	402.645228s
10	1.7032	367.352081s
18*	1.9581*	$205.572563s^{*}$

The results in Table 9 show that more amount of stocks don't always lead to higher Sharpe ratio. When we choose only four stocks out of 30 stocks in DJ30 data set, the optimal portfolio performs much better than other situations. It implies that it is of great significance for investors to limit the amount of assets they hold.

We can observe from Table 10 that the optimal K is 11, which means that there exists an optimal selection consisting 11 stocks out of 32 stocks in SSE50 data set. When the number of stocks exceeds 11, the Sharpe ratio of the portfolio will decrease and the transaction cost will increase.

From the observations in Table 11, 18 stocks are selected out of 125 stocks in CSI300 data set to form an optimal portfolio. Obviously, if we construct a portfolio consisting 125 stocks, each stock accounted for 0.008 on an average weight. Such a tiny percentage may cause more management fees and extra costs. More assets also represent more uncertainty, which stands for greater risk.

In addition to the conclusions above, we also find out that the running time reduces when we install K as a decision variable. That is to say, it is more efficient to let the algorithm find an optimal solution itself than to set a fixed value in advance.

Furthermore, we discover that the more stocks to be selected, the less running time the algorithm takes. In DJ30 data set, it consists of 235 days of daily returns for 30 stocks to be selected, we simply save about 0.01 0.05 seconds. As for SSE50 data set, our sample size is 32 stocks for 581 days, and our method save about 0.1 0.3 seconds. When the data set is big enough, our operational efficiency has been greatly improved, see Table 11. The CSI300 data set contains 399 weeks of weekly return for 125 stocks. When we conduct the numerical experiments on the CSI300 data set with a fixed K, we need about 350 seconds, or even over 400 seconds. When it comes to automatically selecting the optimal solution, the running time is reduced to nearly 205 seconds, saving over 50% of the primal time.

In summary, the numerical results on the three data sets demonstrate that the proposed satisfaction criterion sparse portfolio selection model can be solved efficiently by our primal-dual Benders' decomposition algorithm. From the results above, there are not much difference about the Sharpe ratio between the case of different confidence level $1-\gamma$ for Chinese data sets. The performance of the model between different sparsity K differ noticeably. In particular, we find out that there exists an optimal K that achieves an optimal Sharpe ratio. We can improve the effectiveness of our algorithm by setting K as decision variable.

7. **Conclusion.** We develop a new sparse portfolio optimization model under the satisfaction criterion considering transaction costs. We consider a measure of shortfall-aware aspiration level and transform the objective function into a CVaR form. We estimate the covariance matrix using a data-driven approach to construct an appropriate ambiguity set based on the first and second moments. The robust counterpart is a mixed 0-1 integer semidefinite programming problem via distributionally robust optimization. Our proposed Benders' decomposition algorithm can efficiently give the optimal solution to the problems. We also find that taking the number of assets as a decision variable will significantly improve the algorithm's efficiency, leading to our further work.

Our study considered a sparse portfolio optimization model under satisfaction criterion and achieved robust outcomes. However, limited to the length and workload, we fail to consider more forms of satisfaction criterion. We would like to figure out more efficient method in the future. Moving forward, our work can be extended to a two-stage or multi-stage problem. Furthermore, we can consider other types of ambiguity sets, e.g. Wasserstein distance of return distributions or deterministic approximation via decision rules.

REFERENCES

[1] P. Artzner, F. Delbaen, J. Eber and D. Heath, Coherent measures of risk, *Mathematical Finance*, **3** (1999), 203-228.

[2] M. Barbati, S. Greco, M. Kadziński and R. Słowiński, Optimization of multiple satisfaction levels in portfolio decision analysis, *Omega*, **78** (2018), 192-204.

[3] J. F. Benders, Partitioning procedures for solving mixed-variables programming problems, *Numerical Mathematic*, **4** (1962), 238-252.

[4] N. Bi, J. Tan and W.-S. Tang, A new sufficient condition for sparse vector recovery via $l1$ - $l2$ local minimization, *Analysis and Applications*, **19** (2021), 1019-1031.

[5] R. Bordley and M. LiCalzi, Decision analysis using targets instead of utility functions, *Decisions in Economics and Finance*, **1** (2000), 53-74.

[6] L. G. Chen, D. Z. Long and G. Perakis, The impact of a target on newsvendor decisions, *Manufacturing & Service Operations Management*, **1** (2015), 78-86.

[7] W. Chen and M. Sim, Goal-driven optimization, *Operations Research*, **2** (2009), 342-357.

[8] Y. L. Chen, A. Wiesel, Y. C. Eldar and A. O. Hero, Shrinkage algorithms for MMSE covariance estimation, *IEEE Transactions on Signal Processing*, **10** (2010), 5016-5029.

[9] Z. P. Chen, S. Peng and A. Lisser, A sparse chance constrained portfolio selection model with multiple constraints, *Journal of Global Optimization*, **4** (2020), 825-852.

[10] S. Corsaro, V. De Simone, Z. Marino and F. Perla, l1-Regularization for multi-period portfolio selection, *Annals of Operations Research,*, **294** (2020), 75-86.

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

[12] V. DeMiguel, L. Garlappi, F. J. Nogales and R. Uppal, A generalized approach to portfolio optimization: Improving performance by constraining portfolio norms, *Management Science*, **5** (2009), 798-812.

[13] B. Fastrich, S. Paterlini and P. Winker, Constructing optimal sparse portfolios using regularization methods, *Computational Management Science*, **3** (2014), 417-434.

[14] M. Giuzio, D. Ferrari and S. Paterlini, Sparse and robust normal and t- portfolios by penalized Lq-likelihood minimization, *European Journal of Operational Researche*, **250** (2016), 251-261.

[15] M. H, Portfolio selection, *The Journal of Finance*, **1** (1952), 77-91.

[16] G. A. Hanasusanto, V. Roitch, D. Kuhn and W. Wiesemann, A distributionally robust perspective on uncertainty quantification and chance constrained programming, *Mathematical Programming*, **151** (2015), 35-62.

[17] R. Huang, S. Qu, X. Yang, F. Xu, Z. Xu and W. Zhou, Sparse portfolio selection with uncertain probability distribution, *Applied Intelligence*, **10** (2021), 6665-6684.

[18] K. Ken, T. Yuichi and N. Kazuhide, Cardinality-constrained distributionally robust portfolio optimization, *European Journal of Operational Research*, **3** (2023).

[19] J. H. Kim, Y. Lee, W. C. Kim and F. J. Fabozzi, Goal-based investing based on multi-stage robust portfolio optimization, *Annals of Operations Research*, **1** (2022), 1141-1158.

[20] Z. R. Lai, P. Y. Yang, L. D. Fang and X. T. Wu, Short-term sparse portfolio optimization based on alternating direction method of multipliers, *Journal of Machine Learning Research*, **19** (2018), Paper No. 63, 28 pp.

[21] R. F. Lanzillotti, Pricing objectives in large companies, *The American Economic Review*, **5** (1958), 921-940.

[22] O. Ledoit and M. Wolf, Improved estimation of the covariance matrix of stock returns with an application to portfolio selection, *Journal of Empirical Finance*, **5** (2003), 603-621.

[23] O. Ledoit and M. Wolf, A well-conditioned estimator for large-dimensional covariance matrices, *Journal of Multivariate Analysis*, **88** (2004), 365-411.

[24] A. F. Ling and L. Tang, A numerical study for robust active portfolio management with worst-case downside risk measure, *Mathematical Problems in Engineering*, (2014), 912389, 13 pp.

[25] J. Nie, L. Yang, S. Zhong and G. Zhou, Distributionally robust optimization with moment ambiguity sets, *Journal of Scientific Computing*, **94** (2023), Paper No. 12, 27 pp.

[26] M. Parlar and Z. Kevin Weng, Balancing desirable but conflicting objectives in the newsvendor problem, *IIE Transactions*, **2** (2003), 131-142.

[27] R. Rahmaniani, S. Ahmed, T. G. Crainic, M. Gendreau and W. Rei, The Benders dual decomposition method, *Operations Research*, **68** (2020), 878-895.

[28] R. T. Rockafellar and S. Uryasev, Conditional value-at-risk for general loss distributions, *Journal of Banking & Finance*, **7** (2002), 1443-1471.

[29] S. C. Sarin, H. D. Sherali and L. R. Liao, Minimizing conditional-value-at-risk for stochastic scheduling problems, *Journal of Scheduling*, **17** (2014), 5-15.

[30] A. Shapiro, On duality theory of convex semi-infinite programming, *Optimization*, **54** (2005), 535-543.

[31] Y. Shen, Y. Zuo, L. Sun and X. Zhang, Modified proximal symmetric ADMMs for multi-block separable convex optimization with linear constraints, *Analysis and Applications*, **20** (2022), 401-428.

[32] F. M. Xu, M. H. Wang, Y. H. Dai and D. C. Xu, A sparse enhanced indexation model with chance and cardinality constraints, *Journal of Global Optimization*, **70** (2018), 5-25.

[33] R. R. Yager, Multi-criteria decision making with interval criteria satisfactions using the golden rule representative value, *IEEE Transactions on Fuzzy Systems*, **2** (2018).

[34] Y. Yen, Sparse weighted-norm minimum variance portfolios, *Review of Finance*, **3** (2016), 1259-1287.

[35] Z. H. Zhao, F. M. Xu, M. H. Wang and C. Y. Zhang, A sparse enhanced indexation model with norm and its alternating quadratic penalty method, *Journal of the Operational Research Society*, **3** (2019), 433-445.

[36] S. S. Zhu and M. Fukushima, Worst-case conditional value-at-risk with application to robust portfolio management, *Operations Research*, **5** (2009), 1155-1168.

[37] C. Zopounidis, E. Galariotis, M. Doumpos, S. Sarri and K. Andriosopoulos, Multiple criteria decision aiding for finance: An updated bibliographic survey, *European Journal of Operational Research*, **2** (2015), 339-348.

[38] S. Zymler, D. Kuhn and B. Rustem, Distributionally robust joint chance constraints with second-order moment information, *Mathematical Programming*, **137** (2013), 167-198.

Received August 2023; revised October 2023; early access November 2023.