Journal of Industrial and Management Optimization

https://www.aimspress.com/journal/jimo

JIMO, 22(1): 612–641.
DOI:10.3934/jimo.2026023
Received: 19 September 2025
Revised: 11 November 2025
Accepted: 01 December 2025
Published: 17 December 2025

Research article

An exact algorithm for a cardinality-constrained index tracking model considering investment preferences in portfolio optimization

Chun Wang, Zhongming Wu*, Wei Xu and Yu Yuan

School of Management Science and Engineering, Nanjing University of Information Science and Technology, 219 Ningliu Road, Nanjing, Jiangsu Province, China

* **Correspondence:** Email: wuzm@nuist.edu.cn.

Abstract: Sparse index tracking has emerged as a prominent passive portfolio management strategy, leveraging sparse portfolios to replicate the performance of financial benchmarks. However, while investors often exhibit behavioral biases that influence portfolio selection, existing research has yet to incorporate investor preference information. This paper investigated the cardinality-constrained index tracking problem by integrating investor preferences and enforcing sparsity through a minimal subset of selected assets. Specifically, we proposed a novel approach that incorporates fuzzy preference relations into the sparse tracking model via an additive consistency under priority vectors. This framework captures investor preference heterogeneity while harmonizing subjective inputs with objective tracking performance. An effective matrix decomposition method was employed to address the cardinality-constrained optimization, yielding a precise lower bound and substantially decreasing computational demands. By combining the supergradient method with a branch-and-bound algorithm, our model enhances the efficiency of optimal portfolio selection. Extensive numerical experiments demonstrated that the proposed method outperforms existing lower-bound techniques and mixed-integer programming solvers in tracking efficiency. Furthermore, the preference-aware model balances customization and risk-adjusted returns, achieving superior risk-return performance across multiple tracking metrics. These findings provide practical insights for investor decision-making in portfolio management.

Keywords: index tracking; cardinality constraint; fuzzy preference relation; branch-and-bound method; lower bound analysis
Mathematics Subject Classification: 91G10, 90C10, 65K05

1. Introduction

The mission of an index fund is to mirror the performance of a designated financial benchmark. In recent years, investor demand for index-based products has surged, evidenced by massive cumulative

capital inflows. This enormous scale of investment funding introduces significant practical application challenges. For large-scale funds managing billions, perfectly replicating a broad benchmark becomes operationally complex and prohibitively expensive due to transaction costs, liquidity constraints, and continuous rebalancing. Furthermore, in practical application, fund managers must also increasingly cater to specific investor mandates, such as adhering to environmental, social, and governance (ESG) criteria or excluding certain sectors. Therefore, the central challenge for fund managers is not simply to track an index, but how to do so efficiently while minimizing these costs—a problem known as the "index tracking problem." Given this context, our study focuses on providing efficient solutions to the index tracking problem, specifically addressing the practical needs of large-scale investment funds. We develop models that construct sparse portfolios while simultaneously integrating specific investor preferences.

Recent research has demonstrated significant advancements in index tracking problems, emphasizing algorithmic efficiency and the integration of multi-dimensional constraints. In the algorithmic domain, reinforcement learning frameworks [1] have been introduced to optimize the multi-period tracking control problem of large-scale systems. For ultra-large-scale indices, distributed alternating direction method of multipliers (ADMM) algorithms [2] integrate quantile regression and sparsity constraints, enabling real-time processing in high-dimensional scenarios. For model optimization, existing research has shifted toward multi-objective synergy. For instance, Ling et al. [3] developed the ESG-index tracking hierarchical model, which employs fuzzy logic to quantify sustainability preferences while utilizing robust optimization to control errors. One line of research explores evolutionary algorithms under carbon footprint constraints [4], while another study by Yuen et al. [5] proposes a metaheuristic-based framework to handle practical constraints. Additionally, Wasserstein distance-based distributionally robust frameworks [6] have been developed, all balancing environmental goals with tracking precision. All in all, the current trend gives priority to the computational lightweighting of the tracking portfolio and the customized constraint expansion of the model.

In the real-world portfolio selection process, the implementation of index tracking faces challenges: Large asset pools incur high transaction and management costs [7]. A natural solution is to approximate the index using sparse portfolios [8], which reduces complexity while maintaining tracking accuracy. Sparsity induction has emerged as a pivotal strategy in portfolio optimization, where regularization methods like the ℓ_1 norm regularization (LASSO) provide computational efficiency but suffer from over-shrinkage bias by disproportionately penalizing large weights [9]. Li et al. [10] formulated two constrained optimization problems, solved by their NNOMP-PGD and ADMM-ℓ_0 algorithms, to explicitly control either the tracking inaccuracy or the quantity of chosen assets. Yamagata et al. [11] proposed an ℓ_0-norm constrained method, solved via primal-dual splitting, to control both asset limits and turnover sparsity. A novel sparse fused lasso classification model was introduced by Wu et al. [12], who also formulated a unified multi-block linearized ADMM algorithm for its solution. While non-convex alternatives, such as the minimax concave penalty [13] and smoothly clipped absolute deviation [14], mitigate this bias through adaptive shrinkage, where their non-convexity introduces instability in high-dimensional settings. Mixed norms like the elastic net further balance sparsity with group diversification, yet their performance hinges on delicate parameter tuning [15, 16]. Most importantly, these methods lack explicit control over portfolio composition, a limitation addressed by cardinality constraints—an approach that directly restricts the number of nonzero positions to align with institutional mandates such as exchange-traded fund (ETF) holding limits [17]. However, enforcing cardinality transforms portfolio optimization into an NP-hard

combinatorial problem [18], necessitating advanced computational frameworks. Existing approaches to solving cardinality constraint problems can be broadly categorized into heuristic methods and exact methods. Heuristic and metaheuristic methods prioritize computational speed over guaranteed optimality. This category includes approaches like genetic algorithms [19], particle swarm optimization [20], quantum algorithms [21], as well as other metaheuristic-based frameworks [22]. While often fast, these methods frequently yield subpar tracking accuracy in practice. In contrast, exact methods, particularly mixed-integer linear programming (MILP), guarantee global optimality for moderate-scale problems [23,24]. Benidis et al. [25] introduced a majorization-minimization method to jointly minimize tracking error and cardinality penalties. Zheng et al. [26] proposed a novel reparameterization method and then used stochastic neural networks to solve the optimization problem with cardinality constraints. Among these, the branch-and-bound algorithm stands out for its theoretical rigor, systematically pruning suboptimal solutions via lower-bound comparisons. Innovations such as semidefinite relaxation [27] have enhanced its efficiency by tightening lower bounds. A superior relaxation lower bound for the branch-and-bound technique was derived by Zheng et al. [28], utilizing perspective reformulation and diagonal decomposition.

Beyond technical challenges, investor preferences play a pivotal role in portfolio customization, while the fuzzy information is effectively utilized in decision-making and portfolio selection [29, 30]. We notice that fuzzy preference relations (FPRs) provide an effective mathematical framework to model uncertain preferences, where decision-makers express pairwise comparisons between alternatives on a 0-1 scale [31]. Consistency is a critical attribute of an FPR, as it dictates the logical coherence of an individual's ranking of different options [32]. The conceptualization of consistency in preference relations has evolved along two distinct yet complementary dimensions. The first strand, termed judgment-based consistency, emphasizes the internal coherence of pairwise comparisons within FPRs. Tanino and Tetsuzo [33] defined additive and multiplicative consistency for FPRs, while Li et al. [34] provided expressions for calculating FPR consistency indices. In contrast, priority-based consistency, as developed in [35] and [36], redirected attention to the congruence between judgment elements and an associated priority vector. While judgment-based consistency prioritizes the self-consistency of the comparison matrix itself, priority-based consistency bridges judgments with latent priority structures, offering workable insights for decision-making. Consequently, priority vectors derived from FPRs meeting an additive consistency criterion are suitable for integration with the investment vector. Nowadays, FPRs are extensively utilized across diverse domains, including comprehensive assessment, solution rating, and decision analysis [37,38]. Summarizing the above literature, the existing research has the following two deficiencies and gaps. On the one hand, the tightness and solvability of branch-and-bound's lower bounds are crucial to its efficacy; loose estimates can cause excessive node exploration and long runtimes. Recent advances in cutting-plane generation [39] and quantum-inspired search [40] partially alleviate this bottleneck, but a systematic solution remains elusive. In the latest research of the branch-and-bound algorithm, Xu et al. [41] bridged this gap by proposing a novel lower-bound formulation that takes advantage of asset correlation structures, allowing faster convergence in cardinality-constrained index tracking without sacrificing optimality guarantees. In order to enhance the branch-and-bound approach for solving the cardinality-constrained index tracking optimization, this study mainly makes use of this precise new lower bound. On the other hand, in reality, cognitive and psychological differences lead to behavioral biases among investors, making it crucial to integrate personalized preference information into portfolio selection for effective implementation. Guo et al. [42]

offered a novel definition for additively consistent FPRs by connecting FPRs with the mean-variance framework via priority vectors. Upon summarizing the theoretical development and application of FPRs, we find that research on investors' preference information has not yet been considered in the field of index tracking. Inspired by this, our research pioneers the integration of FPRs into the index tracking model between priority vectors and investment vectors. This approach captures preference variations and integrates effectively with portfolio weights during tracking. Finally, this paper forms a hybrid framework of behavior-aware sparse portfolios, ensuring that sparse portfolios not only replicate market indices but also align with investor-specific preferences.

The contributions of this study can be summed up as follows:

(1) This study represents the first attempt to use priority vectors to include FPR into the index tracking model, potentially bridging the gap between qualitative decision-making and quantitative optimization.
(2) We elucidate the technical and theoretical foundations for a new relational expression linking judgment elements and the priority vector. This expression adheres to the additive consistency criterion and is compatible with the investment vector.
(3) We develop an efficient and exact branch-and-bound algorithm specifically for this new NP-hard optimization problem, which utilizes a novel γ-based lower bound to significantly prune the search space.
(4) An extensive series of numerical experiments is conducted using four real-world datasets. The findings demonstrate that the proposed model is computationally efficient, achieves superior tracking performance, and also delivers higher returns when satisfying customized preferences.

This paper is organized as follows. Section 2 reviews the fundamentals of FPRs, additive consistency, and various sparse index tracking models. Section 3 gives the theoretical analysis of constructing the index tracking model considering the FPRs and remodels index tracking as a combinatorial optimization problem under cardinality constraints. Section 4 describes the global optimal solution method for the cardinality-constrained problem. Section 5 details the solving algorithm, including the supergradient and branch-and-bound methods. Section 6 validates the model and algorithm efficiency. Finally, Section 7 provides conclusions and suggests future research directions.

2. Fuzzy preference relations and sparse index tracking

This section introduces the basic knowledge of FPRs and sparse index tracking models. We assume an index comprises n assets. Let w_i for $i = 1, 2, \ldots, n$ represent the weight of the i-th asset, forming the portfolio weight vector $\mathbf{w} = [w_1, w_2, \ldots, w_n]^\top \in \mathbb{R}^n$. This vector satisfies the budget constraint $\mathbf{e}^\top \mathbf{w} = 1$, where $\mathbf{e} = [1, 1, \ldots, 1]^\top \in \mathbb{R}^n$ is a unit vector.

2.1. Fuzzy preference relations

Following Orlovsky [31], the FPR judgment matrix supplied by the investor is defined as $P = (p_{ij})_{n\times n}$. In this matrix, each element $pij \in [0, 1]$ indicates the preference intensity of asset i relative to asset j, fulfilling

$$p_{ij} + p_{ji} = 1, p_{ii} = 0.5, p_{ij} \geq 0, i, j \in \mathbb{N}^+.$$

When $p_{ij} = 0.5$, it signifies indifference between assets i and j. A value of $0.5 < p_{ij} \leq 1$ implies a preference for asset i over asset j, whereas $0 \leq p_{ij} < 0.5$ indicates a preference for asset j over asset i. The following definition is presented to describe an FPR with additive consistency.

Definition 1. *(from [33]) states that An FPR $P = (p_{ij})_{n\times n}$ possesses additive consistency if the relation*

$$p_{ij} + 0.5 = p_{ik} + p_{kj}, i, j, k \in \mathbb{N}^+.$$

The importance of additive consistency in FPRs is to guarantee coherent and dependable decision-making by upholding consistent preference judgments across various options; failing to do so may lead to contradictory rankings.

Based on Definition 1, the consistency of p_{ij} relies on the value of $p_{ik} + p_{kj} - 0.5$. The smaller the deviation of p_{ij} from $p_{ik} + p_{kj} - 0.5$, the greater the consistency degree of p_{ij}. In this context, Li et al. [34] subsequently formulated the following consistency index for an FPR.

Definition 2. *The consistency index of an FPR P is*

$$CI(P) = 1 - \frac{2}{3n(n-1)(n-2)} \sum_{i,j,k=1, i\neq j\neq k}^{n} \left| p_{ik} + p_{kj} - p_{ij} - 0.5 \right|, \tag{2.1}$$

where the range of $CI(P)$ is the interval $[0, 1]$. The consistency level of the FPR P increases with the value of $CI(P)$, whereas $CI(P) = 1$ signifies perfect consistency.

Formulation (2.1) provides the criterion for additive consistency in an FPR, which relies on the correlations between the judgment elements. It is important to note that this method, while functional, does not fully encompass the vital link to investment strategies. Consequently, Xu [36] proposed a more sophisticated additive consistency framework for FPRs based on the linkage between judgment elements and the priority vector, enabling a connection between FPRs and portfolio selection.

Definition 3. *An FPR $P = (p_{ij})_{n\times n}$ is referred to as additively consistent when there exists a non-negative vector* $\mathbf{w}$ *fulfilling the condition*

$$p_{ij} = 0.5\left(w_i - w_j + 1\right), \quad i, j \in \mathbb{N}^+. \tag{2.2}$$

When p_{ij} satisfies (2.2), Definition 2 must also be satisfied. However, even if the FPR P demonstrates additive consistency (2.1), Definition 3 does not guarantee the existence of a corresponding priority vector in every instance. This means that although (2.2) is correct, it cannot encompass all FPRs that satisfy additive consistency. This necessitates a revised formulation of the connection between w_i, w_j, and p_{ij} such that a compatible priority vector can be derived from any FPR satisfying additive consistency. Guo et al. [42] introduced a reformulation of the relationships linking judgment elements to priority vectors in FPRs. Accordingly, their updated definition of an additively consistent FPR is provided below.

Definition 4. *states that An FPR $P = (p_{ij})_{n\times n}$ fulfills additive consistency and the complementarity of symmetric elements if*

$$p_{ij} = \frac{n-1}{2}\left(w_i - w_j\right) + 0.5, \quad i, j \in \mathbb{N}^+, \tag{2.3}$$

where $w = [w_1, w_2, ..., w_n]^\top \in \mathbb{R}^n$ represents the priority vector of P, which must satisfy $e^\top w = 1$ and $w_i \geq 0$ for all $i \in \mathbb{N}^+$. When $p_{ij} = 1$, $j \neq i$, and $p_{ii} = 0.5$, w_i achieves its maximum value 1. Given that w_i reaches its minimum of 0 under the specific judgments $p_{ij} = 0$ ($j \neq i$) and $p_{ii} = 0.5$, it follows that w_i is bounded to the interval [0, 1]. This bounded range precludes short-selling in the constructed securities market model.

Definition 4 converges perfectly with Definition 1. In other words, if an FPR is additively consistent per Definition 1, a priority vector satisfying Definition 4 is guaranteed to exist. This demonstrates that fully consistent FPRs possess a unique priority vector, which can be seamlessly integrated into an investment strategy.

As a practical example of its economic implication, consider an investor comparing three assets: a technology stock, a utility stock, and a pharmaceutical stock. If the investor strongly prefers the technology stock over the utility stock, they might assign $p_{12} = 0.8$. By definition, this implies $p_{21} = 1 - 0.8 = 0.2$. If they have no preference between the technology and pharmaceutical stocks, they would set $p_{13} = 0.5$. If they slightly prefer the pharmaceutical stock over the utility stock, they might set $p_{32} = 0.6$. This $P = (p_{ij})_{n\times n}$ matrix quantifies the investor's subjective and fuzzy judgments, which our model can then translate into economic constraints on the portfolio's composition.

2.2. Sparse index tracking models

As a widely used tracking statistic, the empirical tracking error (ETE) serves to assess the accuracy with which the tracking portfolio replicates the index [43,44]. Suppose that the return of each asset spans T days, so the i-th index return vector is denoted by $\mathbf{r}^i = [r_1^i, \ldots, r_T^i]^\top \in \mathbb{R}^\top$. Let $\mathbf{X} = [\mathbf{r}^1, \ldots, \mathbf{r}^n]^\top \in \mathbb{R}^{T\times n}$ be the matrix of historical asset returns. The standard formulation is defined as

$$\mathrm{ETE}(\mathbf{w}) = \frac{1}{T}\|\mathbf{X}\mathbf{w} - \mathbf{r}^b\|_2^2,$$

where $\mathbf{r}^b = [r_1^b, \ldots, r_T^b]^\top \in \mathbb{R}^\top$ represents the benchmark index returns over T days. Current approaches for sparse portfolio construction can be categorized into three paradigms.

The conventional sparse index tracking methodology entails two discrete phases: the selection of assets and the subsequent allocation of capital. In the first stage, an asset subset is chosen, which is followed by the second stage, where capital is distributed among the selected assets. Allocating capital suitably is the subsequent challenge after choosing a subset of k assets. This, however, demands precise knowledge of the benchmark portfolio weight vector and its changes, which can be expensive due to frequent rebalancing by index providers. In many instances, the benchmark weight vector may not be available. To overcome this, an optimized allocation strategy can be employed that functions independently of the benchmark weight vector [45]:

$$\begin{aligned} \min_{\mathbf{w}\in\mathbb{R}^n} \quad & \frac{1}{T}\left\|\mathbf{X}(\mathbf{w}\odot\mathbf{z}) - \mathbf{r}^b\right\|_2^2 \\ \text{s.t.} \quad & (\mathbf{w}\odot\mathbf{z})^\top\mathbf{1} = 1, \\ & \mathbf{w} \geq \mathbf{0}, \end{aligned} \tag{2.4}$$

where $\mathbf{z}_i$ is 1 if the i-th stock is chosen and 0 otherwise. Specifically, $(\mathbf{w}\odot\mathbf{z})$ is a vector where the i-th element is $w_i z_i$. Note that in this specific two-step formulation, the binary vector $\mathbf{z}$ is considered a fixed

parameter, not a decision variable in the optimization problem (2.4). This approach minimizes tracking error using selected assets and enhances robustness by iteratively removing assets to achieve sparsity. However, the optimality of the resulting tracking portfolio is uncertain.

The second method integrates these two steps by directly constraining the cardinality of the tracking portfolio:

$$\begin{aligned} \min_{\mathbf{w}\in\mathbb{R}^n} \quad & \frac{1}{T}\left\|\mathbf{X}\mathbf{w}-\mathbf{r}^b\right\|_2^2+\lambda\|\mathbf{w}\|_0 \\ \text{s.t.} \quad & \mathbf{e}^\top\mathbf{w}=1, \\ & \mathbf{w}\geq\mathbf{0}, \end{aligned}$$

where $\lambda \geq 0$ is a parameter controlling the portfolio's sparsity. The non-convex ℓ_0 norm necessitates specialized solvers, with mixed-integer programming being prevalent [46]. However, computational complexity escalates exponentially with problem dimension.

The third approach is to write the constraints considered in the index tracking problem into a non-convex form, which is because the fund manager imposes some position constraints in order to avoid extreme positions or minimal order brokerage costs. In this context, the optimization challenge is structured as follows:

$$\begin{aligned} \min_{\mathbf{w}\in\mathbb{R}^n} \quad & \frac{1}{T}\left\|\mathbf{X}\mathbf{w}-\mathbf{r}^b\right\|_2^2+\lambda\mathbf{z}^\top\mathbf{1} \\ \text{s.t.} \quad & \mathbf{e}^\top\mathbf{w}=1, \\ & \mathbf{l}\odot\mathbf{z}\leq\mathbf{w}\leq\mathbf{u}\odot\mathbf{z}, \\ & \mathbf{z}\in\{0,1\}^N, \end{aligned}$$

where $\mathbf{z}$ serves as the indicator function; however, unlike in (2.4), $\mathbf{z}$ is a variable in this case. Let $\mathbf{l},\mathbf{u}\in\mathbb{R}_+^N$ denote the lower and upper holding limits for the selected stocks, respectively, under the constraint $\mathbf{0}\leq\mathbf{l}\leq\mathbf{u}$. While heuristic methods such as threshold-based reweighting and evolutionary algorithms [47, 48] have been proposed, they often fail to provide optimality guarantees.

In summary, all problem formulations discussed in this section lack explicit control of portfolio sparsity and have their disadvantages in robustness and computational efficiency. This limitation of sparsity can be addressed through the cardinality constraint by using k.

3. Sparse index tracking model considering FPRs

This section first analyzes the theoretical basis of considering investors' preferences through FPRs, then adds cardinality constraints and revenue constraints, and finally integrates them into a sparse index tracking model that can consider the subjective preferences of investors.

3.1. Incorporating fuzzy preference relations

Investors frequently exhibit varying degrees of preference variances during the actual investment process due to differences in cognition, experience, and ways of thinking. Definition 4 indicates that an investor's preferences may be considered when designing a portfolio, therefore prompting the subsequent theoretical examination.

Notice that $p_{ij} \geq 0$, $i, j \in \mathbb{N}^+$, and $j \neq i$ in (2.3), so $|w_i - w_j| \geq -\frac{1}{n-1}$ must hold if $i \in (0, n]$ and $j \in (i, n]$. Next, we use some matrix techniques to rewrite the weight subtraction term and extend the preference difference $|w_i - w_j|$ between individual stocks to the preference difference $|v_a - v_b|$ between stocks in industry a and industry b. Suppose that industry a has h assets and industry b has l assets; then $v_a = \sum_{i=1}^{h} w_{ai}$ and $v_b = \sum_{j=1}^{l} w_{bj}$.

Suppose that n stocks may be categorized into m groups based on industry. In the current study, we assume that the preference information is completely unknown. Then, create a matrix $\mathbf{A}$, with $\frac{1}{2}m(m-1)$ multiplied by n as its dimension. $A_{ai} = 1$ if the asset i is a member of industry a, $A_{bj} = -1$ if the asset j is a member of industry b, and both are 0 otherwise. Based on the above techniques, (2.3) can be equivalently rewritten into the form of industry preference by the matrix expression:

$$v_{ab} = \frac{m-1}{2}(v_a - v_b) + \frac{1}{2}, \tag{3.1}$$

$$\Rightarrow \mathbf{v} = \frac{m-1}{2}\mathbf{A}\mathbf{w} + \frac{1}{2}\mathbf{e}_m, \quad a, b \in \mathbb{N}^+, \tag{3.2}$$

where $\mathbf{e}_m = [1, 1, \ldots, 1]^\top \in \mathbb{R}^m$ is a vector of ones. This formulation ensures that the portfolio weights align with the investor's industry preferences and can be well integrated with the investment vector. We summarize some important properties of (3.2) in the following theorem.

Theorem 1. *The elements* $\mathbf{v}$*, as defined in* (3.2)*, adhere to the complementarity criterion for symmetric elements in FPRs and also fulfill the requirements for additive consistency. Each element in* $\mathbf{v}$ *is bounded within the range* $\mathbf{v}$ *is [0, 1].*

Proof: The complementarity of symmetric elements is proven as:

$$v_{ab} + v_{ba} = \frac{m-1}{2}(v_a - v_b) + \frac{1}{2} + \frac{m-1}{2}(v_b - v_a) + \frac{1}{2} = 1,$$

$$v_{aa} = \frac{m-1}{2}(v_a - v_a) + \frac{1}{2}\mathbf{e}_m = \frac{1}{2}\mathbf{e}_m.$$

The additive consistency definition of FPRs is proven as:

$$\begin{aligned} v_{ab} + \frac{1}{2} &= \frac{m-1}{2}(v_a - v_b) + \frac{1}{2} + \frac{1}{2} \\ &= \frac{m-1}{2}(v_a - v_c) + \frac{1}{2} + \frac{m-1}{2}(v_c - v_b) + \frac{1}{2} \\ &= v_{ac} + v_{cb}. \end{aligned}$$

For the range of v_a and $\mathbf{e}^\top \mathbf{v} = 1$ for FPRs, (3.1) is first rewritten into a summation form:

$$\begin{aligned} \sum_{b=1}^{m} v_{ab} &= \frac{m-1}{2}\left(m v_a - \sum_{b=1}^{m} v_b\right) + \frac{m}{2}, \\ \Rightarrow v_a &= \frac{1}{m}\left[1 + \frac{2}{m-1}\left(\sum_{b=1}^{m} v_{ab} - \frac{m}{2}\right)\right]. \end{aligned}$$

When $v_{ab} = 0, a \neq b$, v_a reaches its minimum value:

$$\min\ v_a = \frac{1}{m}\left[1 + \frac{2}{m-1}(\frac{1}{2} - \frac{m}{2})\right] = 0.$$

When $v_{ab} = 1, a \neq b$, v_a reaches its maximum value:

$$\max\ v_a = \frac{1}{m}\left[1 + \frac{2}{m-1}(\frac{1}{2} + (m-1) - \frac{m}{2})\right] = \frac{2}{m}.$$

Given that the FPR must be a second-order square matrix or larger ($m \geq 2$), it follows that $v_a \leq 1$. When there is no difference between the alternatives, $v_{ab} = \frac{1}{2}$ shows that each alternative's weight is $\frac{1}{m}$, which is consistent with the real circumstances.

So, the range of each element in $\mathbf{v}$ is [0, 1], and this completes the proof. □

Due to $\mathbf{v} \geq \mathbf{0}$, we can construct the subsequent formula that incorporates the investment vector and investor preference information from (3.2):

$$\mathbf{Aw} + \frac{1}{m-1}\mathbf{e}_m \geq \mathbf{0}. \tag{3.3}$$

To sum up, we obtain the constraints of investors on industry preference information. According to Theorem 1, the constraint (3.3) ensures that the solution w_i from (3.2) possesses the properties of complementary symmetry, additive consistency, and nonnegativity.

A clarification regarding the "unknown preference" constraint is as follows:

(1) The model is designed to operate without any a priori FPR provided by the decision-maker. The objective, therefore, is to generate a portfolio $\mathbf{w}$ that is simultaneously quantitatively optimal and behaviorally interpretable.
(2) To achieve this dual objective, we introduce constraint (3.3). This constraint does not represent a specific, known preference; rather, it functions as a structural rationality constraint that embeds the mathematical structure of the additive consistent FPR model from [42]. Its inclusion is derived from the fundamental requirement that the implied preference degrees $\mathbf{v}$, calculated from the final portfolio $\mathbf{w}$, must be valid.
(3) By confining the optimal solution $\mathbf{w}$ to a space that can be mapped back to a consistent FPR, the constraint prevents behaviorally nonsensical solutions. This inverts the traditional paradigm: rather than requiring an FPR as an input, the model provides both the optimal portfolio $\mathbf{w}$ and its implied consistent FPR as an output, serving as a decision-support tool for the investor to review the preferences implicitly revealed by the optimal solution.

This approach establishes a crucial foundation for the integration of qualitative and quantitative analysis. It paves the way for future research where this framework could be extended to integrate known or partially known FPRs, allowing them to serve as explicit guides for portfolio construction.

3.2. Integrated index tracking model

We now integrate the two core concepts discussed. As established in Section 2.2, enforcing sparsity via a cardinality constraint is essential for reducing the transaction and management costs associated

with large portfolios. As derived in Section 3.1, the FPR constraint is our novel mechanism for encoding investor preferences. Combining the primary objective of tracking error minimization with two key constraints—a return constraint and a normalization constraint, we formulate the complete cardinality-constrained tracking error with fuzzy preference relations (CCTE-FPR) model:

$$\begin{aligned} \min_{\mathbf{w}\in\mathbb{R}^n} \quad & \frac{1}{T}\|\mathbf{X}\mathbf{w}-\mathbf{r}^b\|_2^2 \\ \text{s.t.} \quad & \mathbf{r}^\top\mathbf{w} \geq s, \\ & \mathbf{e}^\top\mathbf{w} = 1, \\ & \|\mathbf{w}\|_0 \leq k, \\ & \mathbf{A}\mathbf{w} + \frac{1}{m-1}\mathbf{e}_m \geq \mathbf{0}, \end{aligned} \tag{3.4}$$

where only $k \in \mathbb{N}^+$ ($k \ll n$) assets are selected with nonzero weights. $\mathbf{r} = [\bar{r_1}, \ldots, \bar{r_n}]^\top$ is the vector of mean returns for each asset over the entire training window of T days, and $\mathbf{w}$ represents the portfolio we wish to construct. The level of the portfolio's expected return over this window is required to be no less than s.

The branch-and-bound method is well-suited for solving cardinality-constrained portfolio optimization, a recognized advantage in the literature. We begin by reformulating CCTE-FPR model (3.4) using the Lagrangian relaxation technique. This is achieved by introducing Lagrangian multipliers $\alpha \in \mathbb{R}^+, \beta \in \mathbb{R}$, and $\boldsymbol{\gamma} = [\gamma_1^b, \ldots, \gamma_m^b]^\top \in \mathbb{R}_+^m$, leading to the following relaxed problem:

$$\begin{aligned} \min_{\mathbf{w}\in\mathbb{R}^n} \quad & L(\mathbf{w},\alpha,\beta,\boldsymbol{\gamma}) = \frac{1}{T}\Big(\mathbf{w}^\top\mathbf{X}^\top\mathbf{X}\mathbf{w} - 2(\mathbf{r}^b)^\top\mathbf{X}\mathbf{w} + (\mathbf{r}^b)^\top\mathbf{r}^b\Big) \\ & \qquad + (\beta\mathbf{e}^\top - \alpha\mathbf{r} - \boldsymbol{\gamma}\mathbf{A})\mathbf{w} + (\alpha s - \beta - \frac{1}{m-1}\boldsymbol{\gamma}\mathbf{e}_m) \\ \text{s.t.} \quad & \|\mathbf{w}\|_0 \leq k. \end{aligned} \tag{3.5}$$

In problem (3.5), the discrete cardinality constraint is the sole one that is challenging to handle directly. While exhaustive enumeration could theoretically find the global optimum, combinatorial explosion renders it computationally infeasible for large asset pools. Accordingly, we present a novel and precise lower bound in the subsequent sections to strengthen the branch-and-bound algorithm for this problem.

4. A global optimal method for cardinality-constrained index tracking model

This section employs a specialized matrix processing technique to approximate the asset return covariance matrix as a diagonal one. Leveraging dual theory, we reformulate the problem's lower bound as a max-min optimization. Consequently, the lower bound problem reduces to a concave optimization over the Lagrangian multipliers.

4.1. A matrix processing approach

For addressing the sparse discrete constraints in problem (3.5), the branch-and-bound method is an effective instrument. Its efficiency hinges on the tightness of the lower bounds used in the search. Consequently, to enhance computational performance, we employ a straightforward and precise

construction method to determine the lower bound for the tracking portfolio problem. Specifically, we derive a reduced diagonal matrix from the original covariance matrix via a specialized processing technique, thereby obtaining a lower-bounding formulation for the initial problem.

We begin by defining $\mathbf{Q} = \mathbf{X}^\top \mathbf{X}$ as the return covariance matrix. It is clear that $\mathbf{Q}$ is an n-dimensional positive definite matrix. Due to this, the original Lagrangian formula (3.5) is as follows:

$$\begin{aligned} \min_{\mathbf{w}\in\mathbb{R}^n} \quad & L(\mathbf{w},\alpha,\beta,\boldsymbol{\gamma}) = \frac{1}{T}\Big(\mathbf{w}^\top\mathbf{Q}\mathbf{w} - 2(\mathbf{r}^b)^\top(\mathbf{X}\mathbf{w}) + (\mathbf{r}^b)^\top\mathbf{r}^b\Big) \\ & \qquad + (\beta\mathbf{e}^\top - \alpha\mathbf{r} - \boldsymbol{\gamma}\mathbf{A})\mathbf{w} + (\alpha s - \beta - \frac{1}{m-1}\boldsymbol{\gamma}\mathbf{e}_m) \\ \text{s.t.} \quad & \|\mathbf{w}\|_0 \le k. \end{aligned} \tag{4.1}$$

We now focus on the objective function $L(\mathbf{w},\alpha,\beta,\boldsymbol{\gamma})$ and aim to develop a diagonal positive definite matrix smaller than $\mathbf{Q}$, leveraging its properties of symmetry and positive definiteness.

Adopting the method from Xu et al. [41], $\boldsymbol{\Phi} \in \mathbb{R}^{n\times n}$ is defined as the inverse of the symmetric positive definite matrix $\mathbf{Q}$. We then proceed to construct a diagonal matrix $\boldsymbol{\Lambda} \in \mathbb{R}^{n\times n}$ such that:

$$\boldsymbol{\Lambda} = \begin{bmatrix} \frac{1}{\sum_{j=1}^n |\Phi_{1j}|} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \frac{1}{\sum_{j=1}^n |\Phi_{nj}|} \end{bmatrix}.$$

Given that $\mathbf{Q}$ is positive definite, its inverse $\boldsymbol{\Phi}$ shares this property. Consequently, for any $i \in 1,2,\ldots,n$, the corresponding diagonal element meets the condition $\sum_{j=1}^n |\Phi_{ij}| > 0$. This result directly entails that $\boldsymbol{\Lambda} > \mathbf{0}$, confirming $\boldsymbol{\Lambda}$ as a diagonal positive definite matrix. Let $\mathbf{I}_n$ be the n-dimensional identity matrix, and define $\mathbf{M} = \boldsymbol{\Lambda}\boldsymbol{\Phi} - \mathbf{I}_n$. For any $i \in \{1,2,\ldots,n\}$:

$$\begin{aligned} \sum_{j_1=1}^n |M_{ij_1}| &= \sum_{j_1=1, j_1\neq i}^n \frac{|\Phi_{ij_1}|}{\sum_{j=1}^n |\Phi_{ij}|} + |M_{ii}| \\ &= 1 - \frac{|\Phi_{ii}|}{\sum_{j=1}^n |\Phi_{ij}|} + |M_{ii}| \\ &\le -M_{ii} + |M_{ii}|. \end{aligned}$$

This leads to the conclusion that matrix $\mathbf{M}$ is negative semidefinite, a result derived from the proofs in [41, Proof of Lemma1]. Specifically, $\mathbf{M} = \boldsymbol{\Lambda}\boldsymbol{\Phi} - \mathbf{I}_n \preceq \mathbf{0}$, which implies $\boldsymbol{\Lambda} \preceq \boldsymbol{\Phi}^{-1} = \mathbf{Q}$, thus establishing the relation $\mathbf{0} \prec \boldsymbol{\Lambda} \preceq \mathbf{Q}$.

Utilizing the aforementioned methods in conjunction with problem (4.1), we can construct a diagonal positive definite matrix $\boldsymbol{\Lambda}$ possessing the property that $\mathbf{w}^\top(\boldsymbol{\Lambda} - \mathbf{Q})\mathbf{w} \le 0$ for all nonzero $\mathbf{w} \in \mathbb{R}_n^+$. In other words, by performing straightforward numerical computations on the inverse of $\mathbf{Q}$, we can obtain a smaller diagonal matrix $\boldsymbol{\Lambda}$. Consequently, a tighter lower bound in problem (4.1) can be derived by solving:

$$\begin{aligned} \min_{\mathbf{w}\in\mathbb{R}^n} \quad & L(\mathbf{w},\alpha,\beta,\boldsymbol{\gamma}) = \frac{1}{T}\mathbf{w}^\top\boldsymbol{\Lambda}\mathbf{w} + \Big(\beta\mathbf{e}^\top - \alpha\mathbf{r} - \boldsymbol{\gamma}\mathbf{A} - \frac{2}{T}(\mathbf{r}^b)^\top\mathbf{X}\Big)\mathbf{w} \\ & \qquad + \frac{1}{T}(\mathbf{r}^b)^\top\mathbf{r}^b + (\alpha s - \beta - \frac{1}{m-1}\boldsymbol{\gamma}\mathbf{e}_m) \\ \text{s.t.} \quad & \|\mathbf{w}\|_0 \le k. \end{aligned} \tag{4.2}$$

Note that, for the sake of simplicity of the formula, the constant part $\frac{1}{T}(\mathbf{r}^b)^\top \mathbf{r}^b + (\alpha s - \beta - \frac{1}{m-1}\boldsymbol{\gamma}\mathbf{e}_m)$ as a whole is denoted by C. We can see that the optimal values in (4.2) are less than or equal to the optimal values in (3.4) for any $\alpha \in \mathbb{R}^+$, $\beta \in \mathbb{R}$, and $\boldsymbol{\gamma} \in \mathbb{R}^m_+$. The best lower bound is therefore obtained by maximizing the lower bound (4.2).

4.2. Reconstructing lower bound in ellipsoid form

We examine the dual model of the quadratic objective function in order to derive the analytical solution of the inner minimization issue, as cardinality constraints may be met by comparing the diagonal elements. The ellipsoid-like structure implies that the optimization problem has a well-defined geometric interpretation, which provides stronger bounds and guarantees more reliable and precise solutions.

By reformulating the Lagrangian function $L(\mathbf{w}, \alpha, \beta, \boldsymbol{\gamma})$ into an ellipsoid-like equation, we leverage a specific structural property. This construction guarantees that the new dual lower bound is consistently tighter than the continuous relaxation bound, as formally established in the subsequent theorems:

$$L(\mathbf{w}, \alpha, \beta, \boldsymbol{\gamma}) = \frac{1}{T}(\mathbf{w} + \mathbf{m})^\top \mathbf{Q}(\mathbf{w} + \mathbf{m}) - T\mathbf{m}^\top \mathbf{Q}\mathbf{m} + C, \tag{4.3}$$

where $\mathbf{m} = \frac{T}{2}[(\beta\mathbf{Q}^{-1}\mathbf{e} - \mathbf{Q}^{-1}(\alpha\mathbf{r})^\top - \mathbf{Q}^{-1}(\boldsymbol{\gamma}\mathbf{A})^\top] - \mathbf{Q}^{-1}\mathbf{X}^\top\mathbf{r}^\mathbf{b}$. We also substitute matrix $\mathbf{Q}$ with $\boldsymbol{\Lambda}$, formulating a new lower bound optimization problem:

$$\begin{aligned} \min_{\mathbf{w}\in\mathbb{R}^n} \quad & L(\mathbf{w}, \alpha, \beta, \boldsymbol{\gamma}) = \frac{1}{T}(\mathbf{w} + \mathbf{m})^\top \boldsymbol{\Lambda}(\mathbf{w} + \mathbf{m}) - T\mathbf{m}^\top \mathbf{Q}\mathbf{m} + C \\ \text{s.t.} \quad & \|\mathbf{w}\|_0 \le k. \end{aligned} \tag{4.4}$$

By exploiting the diagonal structure of $\boldsymbol{\Lambda}$, the objective function in (4.4) can be equivalently expressed as the summation:

$$L(\mathbf{w}, \alpha, \beta, \boldsymbol{\gamma}) = \frac{1}{T}\sum_{j=1}^{n} \Lambda_{jj}(w_j + m_j)^2 - T\mathbf{m}^\top \mathbf{Q}\mathbf{m} + C.$$

Given that the cardinality constraint applies exclusively to $\mathbf{w}$, we address this complex constraint by deriving the dual of (4.4):

$$\begin{aligned} \max_{\alpha,\boldsymbol{\gamma}\in\mathbb{R}^+,\beta\in\mathbb{R}} \min_{\mathbf{w}\in\mathbb{R}^n} \quad & L(\mathbf{w}, \alpha, \beta, \boldsymbol{\gamma}) = \frac{1}{T}\sum_{j=1}^{n} \Lambda_{jj}(w_j + m_j)^2 - T\mathbf{m}^\top \mathbf{Q}\mathbf{m} + C \\ \text{s.t.} \quad & \|\mathbf{w}\|_0 \le k. \end{aligned} \tag{4.5}$$

Next, we analyze the optimal solution associated with the inner minimization problem in (4.5), which is subject to the cardinality constraint.

For any fixed $\bar{\alpha} \in \mathbb{R}^+$, $\bar{\beta} \in \mathbb{R}$, and $\bar{\boldsymbol{\gamma}} \in \mathbb{R}^m_+$, the inner minimization problem is

$$\begin{aligned} \min_{\mathbf{w}\in\mathbb{R}^n} \quad & L(\mathbf{w}, \bar{\alpha}, \bar{\beta}, \bar{\boldsymbol{\gamma}}) = \frac{1}{T}\sum_{j=1}^{n} \Lambda_{jj}(w_j + m_j)^2 - T\mathbf{m}^\top \mathbf{Q}\mathbf{m} + C \\ \text{s.t.} \quad & \|\mathbf{w}\|_0 \le k. \end{aligned} \tag{4.6}$$

In the absence of the cardinality constraint, the minimal value of (4.6) would be achieved by setting $w_j = -m_j$ for all $j = 1, 2, \ldots, n$. The presence of the constraint, however, dictates that at most k elements of $\mathbf{w}$ are nonzero, consequently requiring at least $(n-k)$ elements to be zero. For objective minimization, the j-th summation term admits one of two values:

$$\Lambda_{jj}(w_j + m_j)^2 = \begin{cases} 0, & w_j = -m_j, \\ \Lambda_{jj} m_j^2, & w_j = 0. \end{cases}$$

Since $\Lambda_{jj} m_j^2 \geq 0$ for any $j \in \{1, 2, \ldots, n\}$, it suffices to set the k largest instances of $\Lambda_{jj} m_j^2$ to zero. The optimal value of (4.6) is then obtained by

$$L^*(\bar{\alpha}, \bar{\beta}, \bar{\boldsymbol{\gamma}}) = \frac{1}{T} \sum_{j \in \mathcal{I}} \Lambda_{jj} m_j^2 - T \mathbf{m}^\top \mathbf{Q} \mathbf{m} + C,$$

with the optimal solution $w_j^* \in \mathbb{R}^n$ given as:

$$w_j^* = \begin{cases} 0, & j \in \mathcal{I}, \\ -m_j, & j \notin \mathcal{I}, \end{cases}$$

where $\mathcal{I}$ corresponds to the indices of the first $(n - k)$ smallest values in the sequence $\Lambda_{jj} m_j^2 : j = 1, 2, \ldots, n$.

Lastly, (4.5) can be reduced to the following optimization problem:

$$\max_{\alpha, \boldsymbol{\gamma} \in \mathbb{R}^+, \beta \in \mathbb{R}} \quad \Gamma(\alpha, \beta, \boldsymbol{\gamma}) = \frac{1}{T} \sum_{j \in \mathcal{I}} \Lambda_{jj} m_j^2 - T \mathbf{m}^\top \mathbf{Q} \mathbf{m} + C. \tag{4.7}$$

For the index sets $\mathcal{I}$, we define a diagonal matrix $\boldsymbol{\Omega} \in \mathbb{R}^{n \times n}$ where the diagonal elements are:

$$\boldsymbol{\Omega}_{jj} = \begin{cases} \boldsymbol{\Lambda}_{jj}, & j \in \mathcal{I}, \\ \mathbf{0}, & j \notin \mathcal{I}. \end{cases}$$

4.3. Analysis of dual problems

The strength of the boundaries, indicating how closely they approximate the original optimal value, directly impacts the efficiency of the algorithm. Evaluating the relationship of different boundaries reveals the quality of the relaxation: A dual model producing bounds closer to the primal optimum indicates a stronger formulation, better aligned with the problem structure. This analysis aids in selecting appropriate strategies, balancing computational effort, and bounding quality. A key relationship between the optimal values of certain dual problems is presented in Theorem 2.

Theorem 2. *Let $\mathcal{L}_c^*$ and $\mathcal{L}_d^*$ represent the optimal values of* (3.4) *and* (4.5)*, respectively. Meanwhile, $\mathcal{L}^*$ refers to the optimal value of the unconstrained Lagrangian dual of* (3.4)*. If $\mathbf{Q}^{-1} \boldsymbol{\Omega} \mathbf{Q}^{-1} \geq \mathbf{0}$, the relationship follows the inequality $\mathcal{L}^* \leq \mathcal{L}_d^* \leq \mathcal{L}_c^*$.*

Proof: As discussed earlier, (4.5) represents the lower bound optimization problem of (3.4). Hence, it is straightforward to conclude that $\mathcal{L}_d^* \leq \mathcal{L}_c^*$. For the model without the cardinality constraint, it is observed that $\mathbf{w}^* = -\mathbf{m}$ in (4.3). Consequently, the expression for $\mathcal{L}$ is rewritten as follows:

$$\max_{\alpha, \boldsymbol{\gamma} \in \mathbb{R}^+, \beta \in \mathbb{R}} \quad L(\alpha, \beta, \boldsymbol{\gamma}) = -T \mathbf{m}^\top \mathbf{Q} \mathbf{m} + C.$$

Clearly, for any $\alpha \in \mathbb{R}^+, \beta \in \mathbb{R}$, and $\boldsymbol{\gamma} \in \mathbb{R}^m_+$, $L(\alpha, \beta, \boldsymbol{\gamma}) \leq \Gamma(\alpha, \beta, \boldsymbol{\gamma})$. Given that $\Gamma(\alpha, \beta, \boldsymbol{\gamma})$ in (4.7) serves as an equivalent representation of $\mathcal{L}_d$, it follows directly that $\mathcal{L}^* \leq \mathcal{L}^*_d$.

This completes the proof. □

Therefore, in comparison to these lower bound problems, the value of $\mathcal{L}_d$ obtained by solving (4.7) has the potential to eliminate more branches, making it more stable than other lower bounds and enhancing computational efficiency.

5. Solution methods

In addressing the lower bound problem in (4.7), the key challenge lies in determining the index set $i \in \mathcal{I}$. In this section, we employ the supergradient method [49] to resolve this issue. Leveraging the established lower bound properties, the depth-first branch-and-bound approach enables exact determination of the global optimal solution for the CCTE-FPR model.

5.1. Supergradient method

The conventional interval splitting method [50] in portfolio selection processes has a significant drawback: The method discretizes the interval into $n(n-1)+1$ segments, each requiring a subproblem solution. As the asset universe expands, the breakpoint count grows rapidly, leading to exponential increases in computation time. In contrast, by utilizing gradient-based information, the supergradient method [49] avoids the need for exhaustive interval partitions, significantly reducing the number of subproblems that need to be solved. This leads to faster and more scalable computations, particularly in large-scale problems, as it effectively captures the necessary information with fewer computational resources.

Building upon this foundation, we leverage the gradient information to construct a supergradient-based algorithm. The gradient functions of (4.7) are:

$$\begin{aligned}
\nabla_\alpha \Gamma(\alpha, \beta, \boldsymbol{\gamma}) &= -\frac{1}{T} \sum_{j \in \mathcal{I}} \Lambda_{jj} m_j (\mathbf{Q}^{-1}\mathbf{r})_j + \frac{1}{T}\mathbf{m}^\top \mathbf{r} + s, \\
\nabla_\beta \Gamma(\alpha, \beta, \boldsymbol{\gamma}) &= \frac{1}{T} \sum_{j \in \mathcal{I}} \Lambda_{jj} m_j (\mathbf{Q}^{-1}\mathbf{e})_j - \frac{1}{T}\mathbf{m}^\top \mathbf{e} - 1, \\
\nabla_{\boldsymbol{\gamma}} \Gamma(\alpha, \beta, \boldsymbol{\gamma}) &= -\frac{1}{T} \sum_{j \in \mathcal{I}} \Lambda_{jj} m_j (\mathbf{A}\mathbf{Q}^{-1})_j + \frac{1}{T}\mathbf{m}^\top \mathbf{A} - \frac{1}{m-1}\mathbf{e}_m.
\end{aligned} \tag{5.1}$$

Next, in Theorem 3, we establish the concavity of the lower bound function $\Gamma(\alpha, \beta, \boldsymbol{\gamma})$. Consequently, the optimal value can be located using a gradient-based approach.

Theorem 3. *Assume a fixed pair* $(\alpha_0, \beta_0, \boldsymbol{\gamma}_0)$. *Every Lagrangian multiplier* $(\alpha, \beta, \boldsymbol{\gamma}) \in \mathbb{R}^+ \times \mathbb{R} \times \mathbb{R}^m_+$ *obeys the subsequent inequality:*

$$\Gamma(\alpha, \beta, \boldsymbol{\gamma}) \leq \Gamma(\alpha_0, \beta_0, \boldsymbol{\gamma}_0) + \nabla_\alpha \Gamma(\alpha_0, \beta_0, \boldsymbol{\gamma}_0)(\alpha - \alpha_0) + \nabla_\beta \Gamma(\alpha_0, \beta_0, \boldsymbol{\gamma}_0)(\beta - \beta_0) + \nabla_\gamma \Gamma(\alpha_0, \beta_0, \boldsymbol{\gamma}_0)(\boldsymbol{\gamma} - \boldsymbol{\gamma}_0).$$

Proof: For the fixed pair $(\alpha_0, \beta_0, \boldsymbol{\gamma}_0)$, the index set $\mathcal{I}$ and the vector $\mathbf{m}_0 = \frac{T}{2}[\beta_0 \mathbf{Q}^{-1}\mathbf{e} - \alpha_0 \mathbf{Q}^{-1}\mathbf{r} - \mathbf{Q}^{-1}(\boldsymbol{\gamma}_0 \mathbf{A})^\top] - \mathbf{Q}^{-1}\mathbf{X}^\top \mathbf{r}^{\mathbf{b}}$ are determined. The constant part $\frac{1}{T}(\mathbf{r}^b)^\top \mathbf{r}^b + (\alpha_0 s - \beta_0 - \frac{1}{m-1}\boldsymbol{\gamma}_0 \mathbf{e}_m)$ as a whole

is denoted by C_0. The corresponding objective value is:

$$\begin{aligned}\Gamma(\alpha_0,\beta_0,\boldsymbol{\gamma}_0) &= \frac{1}{T}\sum_{j\in\mathcal{I}}\Lambda_{jj}(m_0)_j^2 - \frac{1}{T}\mathbf{m_0}^\top\mathbf{Q}\mathbf{m_0} + C_0\\ &= \frac{1}{T}\mathbf{m}_0^\top(\boldsymbol{\Omega}-\mathbf{Q})\mathbf{m}_0 + C_0.\end{aligned}$$

Next, we demonstrate that (4.7) is a concave function regarding $(\alpha,\beta,\boldsymbol{\gamma})$.

Denote $\mathbf{H} = \mathbf{Q}^{-1}(\boldsymbol{\Omega}-\mathbf{Q})\mathbf{Q}^{-1}$. For any $(\alpha,\beta,\boldsymbol{\gamma}) \in \mathbb{R}^+\times\mathbb{R}\times\mathbb{R}_m^+$, the Hessian matrix of (4.7) is a constant symmetric matrix, given by:

$$\mathbf{I} = \begin{bmatrix}\mathbf{r}^\top\mathbf{H}\mathbf{r} & -\mathbf{r}^\top\mathbf{H}\mathbf{e} & \mathbf{r}^\top\mathbf{H}\mathbf{A}^\top\\ -\mathbf{r}^\top\mathbf{H}\mathbf{e} & \mathbf{e}^\top\mathbf{H}\mathbf{e} & -\mathbf{e}^\top\mathbf{H}\mathbf{A}^\top\\ \mathbf{A}\mathbf{H}\mathbf{r} & -\mathbf{A}\mathbf{H}\mathbf{e} & \mathbf{A}\mathbf{H}\mathbf{A}^\top\end{bmatrix}.$$

Define the $\mathbf{M}$ matrix as

$$\mathbf{M} = \begin{bmatrix}1 & 0 & 0\\ \frac{\mathbf{r}^\top\mathbf{H}\mathbf{e}}{\mathbf{r}^\top\mathbf{H}\mathbf{r}} & 1 & 0\\ \frac{\mathbf{r}^\top\mathbf{H}\mathbf{A}^\top}{\mathbf{r}^\top\mathbf{H}\mathbf{r}} & 0 & 1\end{bmatrix}.$$

Using the basic transformation, we get

$$\mathbf{M}\begin{bmatrix}\mathbf{r}^\top\mathbf{H}\mathbf{r} & \mathbf{r}^\top\mathbf{H}\mathbf{e} & \mathbf{r}^\top\mathbf{H}\mathbf{A}^\top\\ \mathbf{r}^\top\mathbf{H}\mathbf{e} & \mathbf{e}^\top\mathbf{H}\mathbf{e} & \mathbf{e}^\top\mathbf{H}\mathbf{A}^\top\\ \mathbf{A}\mathbf{H}\mathbf{r} & \mathbf{A}\mathbf{H}\mathbf{e} & \mathbf{A}\mathbf{H}\mathbf{A}^\top\end{bmatrix}\mathbf{M}^\top.$$

Then, we get the following matrix:

$$\begin{bmatrix}\mathbf{r}^\top\mathbf{H}\mathbf{r} & 0 & 0\\ 0 & \frac{\mathbf{r}^\top\mathbf{H}\mathbf{r}\mathbf{e}^\top\mathbf{H}\mathbf{e}-(\mathbf{r}^\top\mathbf{H}\mathbf{e})^2}{\mathbf{r}^\top\mathbf{H}\mathbf{r}} & 0\\ 0 & 0 & \frac{\mathbf{r}^\top\mathbf{H}\mathbf{r}\mathbf{A}^\top\mathbf{H}\mathbf{A}-(\mathbf{r}^\top\mathbf{H}\mathbf{A}^\top)^2}{\mathbf{r}^\top\mathbf{H}\mathbf{r}}\end{bmatrix}.$$

$\boldsymbol{\Omega}-\mathbf{Q}$ is a negative semidefinite matrix since $\boldsymbol{\Omega}\preceq\boldsymbol{\Lambda}\preceq\mathbf{Q}$. Consequently, $\mathbf{H}$ is likewise a negative semidefinite matrix. As a result, $\mathbf{r}^\top\mathbf{H}\mathbf{r}\le 0$.

Performing the Cholesky decomposition $\mathbf{H} = \mathbf{U}^\top\mathbf{U}$ yields the following derivation:

$$\begin{aligned}(\mathbf{r}^\top\mathbf{H}\mathbf{e})^2 &= ((\mathbf{U}\mathbf{r})^\top(\mathbf{U}\mathbf{e}))^2\\ &= \mathbf{r}^\top\mathbf{H}\mathbf{r}\mathbf{e}^\top\mathbf{H}\mathbf{e},\end{aligned}$$

where the inequality is established by the application of the Cauchy-Schwarz inequality. This implies:

$$\frac{\mathbf{r}^\top\mathbf{H}\mathbf{r}\mathbf{e}^\top\mathbf{H}\mathbf{e}-(\mathbf{r}^\top\mathbf{H}\mathbf{e})^2}{\mathbf{r}^\top\mathbf{H}\mathbf{r}}\le 0.$$

Similarly,

$$\frac{\mathbf{r}^\top\mathbf{H}\mathbf{r}\mathbf{A}^\top\mathbf{H}\mathbf{A}-(\mathbf{r}^\top\mathbf{H}\mathbf{A}^\top)^2}{\mathbf{r}^\top\mathbf{H}\mathbf{r}}\le 0.$$

Thus, the Hessian matrix $\mathbf{H}$ is negative semidefinite. This means (4.7) is a concave function with regard to $(\alpha, \beta, \boldsymbol{\gamma})$. Next, for any $(\alpha, \beta, \boldsymbol{\gamma}) \in \mathbb{R}^+ \times \mathbb{R} \times \mathbb{R}_+^m$, it holds that

$$\begin{aligned}\Gamma(\alpha, \beta, \boldsymbol{\gamma}) &= \Gamma(\alpha_0, \beta_0, \boldsymbol{\gamma}_0) + \nabla_\alpha \Gamma(\alpha_0, \beta_0, \boldsymbol{\gamma}_0)(\alpha - \alpha_0) + \nabla_\beta \Gamma(\alpha_0, \beta_0, \boldsymbol{\gamma}_0)(\beta - \beta_0) \\ &\quad + \nabla_\gamma \Gamma(\alpha_0, \beta_0, \boldsymbol{\gamma}_0)(\boldsymbol{\gamma} - \boldsymbol{\gamma}_0) + \frac{1}{T}[\alpha - \alpha_0\beta - \beta_0\boldsymbol{\gamma} - \boldsymbol{\gamma}_0]\mathbf{H}[\alpha - \alpha_0\beta - \beta_0\boldsymbol{\gamma} - \boldsymbol{\gamma}_0]^\top \\ &\leq \Gamma(\alpha_0, \beta_0, \boldsymbol{\gamma}_0) + \nabla_\alpha \Gamma(\alpha_0, \beta_0, \boldsymbol{\gamma}_0)(\alpha - \alpha_0) + \nabla_\beta \Gamma(\alpha_0, \beta_0, \boldsymbol{\gamma}_0)(\beta - \beta_0) \\ &\quad + \nabla_\gamma \Gamma(\alpha_0, \beta_0, \boldsymbol{\gamma}_0)(\boldsymbol{\gamma} - \boldsymbol{\gamma}_0).\end{aligned}$$

This completes the proof. □

According to Theorem 3, the extreme point of the concave function (4.7) for the variable group $(\alpha, \beta, \boldsymbol{\gamma})$ reaches the optimum when its supergradient is zero. Based on the search principle of the supergradient method, this paper designs the following iterative process: The current objective function value is denoted as Obj, and the current optimal lower bound is Cur_LB. In the initialization stage of the algorithm, the initial lower bound Cur_LB is obtained by solving the dual relaxation problem, and the Lagrange multipliers are initialized based on this solution. We establish a convergence tolerance, ϵ, using $\left|\text{Obj} - \text{Cur\_LB}\right| \leq 0.01$ as the stopping criterion for Algorithm 1.

Algorithm 1: Supergradient method

Input: Return rate $\mathbf{r} \in \mathbb{R}^n$ and its matrix $\mathbf{X} \in \mathbb{R}^{T \times n}$, returns of normalized benchmark index $\mathbf{r}^b$, parameters $\alpha, \beta, \boldsymbol{\gamma}$, minimum expected return s, maximum iteration k, and error tolerance ϵ.

Output: The lower bound Cur_LB of (4.7)

```
while ϵ > 0.01 do
    Obj ← Obtain the objective value of (4.7) using α, β, and γ;
    ϵ ← |Obj − Cur_LB|;
    if Cur_LB < Obj then
        Cur_LB ← Obj;
    else
        Step ← Set the searching step size to |β|;
        G1, G2, G3 ← Obtain the supergradient via Function (5.1);
        α ← α + Step ∗ G1;
        β ← β + Step ∗ G2;
        γ ← γ + Step ∗ G3;
    end
end
```

Ultimately, the Cur_LB generated by the algorithm serves as a tight lower bound for the initial problem (3.4). This design ensures effective convergence through supergradient direction correction and an adaptive step size strategy.

5.2. Branch-and-bound method

The branch-and-bound framework, built upon the preceding lower bound analysis, is employed to address the cardinality-constrained index tracking problem, with its complete procedure detailed in Algorithm 2.

Algorithm 2: Depth-first branch-and-bound framework

```
GB_UB ← Initialize upper bound;
List ← Initialize with root problem {P_0};
while List ≠ ∅ do
    P ← Extract problem from List;
    List = List-P;
    Bounding Phase ;
    if ||w||_0 ≤ k then
        Cur_UB ← Solve P for updated upper bound;
        if Cur_UB < GB_UB then
            GB_UB = Cur_UB;
        end
        Continue;
    end
    Cur_LB ← Compute lower bound via Algorithm 1;
    Pruning Phase ;
    if Cur_LB > GB_UB then
        Prune current branch;
        Continue;
    end
    Branching Phase;
    Select undetermined variable w_j from P;
    Generate subproblems P' (w_j = 1) and P" (w_j = 0);
    Append {P', P"} to List;
    return GB_UB
end
```

The proposed depth-first branch-and-bound algorithm systematically navigates the solution space represented as a tree structure, employing bounding techniques for the pruning of suboptimal paths to efficiently solve combinatorial optimization problems. To be specific, we employ a tight initial upper bound construction strategy. The initialization step involves solving the original unconstrained problem, which yields the initial multiplier pair $(\alpha_0, \beta_0, \boldsymbol{\gamma}_0)$. These multipliers are then subsequently utilized in Algorithm 1 to create a feasible solution $\mathbf{w}_0$, representing a candidate portfolio configuration. The initial upper bound GB_UB is then initialized as the objective value $f(\mathbf{w}_0)$. The algorithm maintains a dynamically updated list (List) to manage subproblems generated during the branching process. Each subproblem P $\in$ List represents a node in the search tree, initialized with the root node P_0 that lacks preassigned investment decisions. For every selected subproblem P, its lower bound is rigorously computed via Algorithm 1.

A critical pruning criterion is applied: If Cur_LB > GB_UB, all descendant branches of P are discarded. Conversely, if Cur_LB ≤ GB_UB, branching proceeds by selecting an undetermined variable w_j and generating two child subproblems $P'(w_j = 1)$ and $P''(w_j = 0)$, which are appended to the List mentioned above for depth-first exploration. Notably, the root node P_0 undergoes an initial bounding check: If its lower bound exceeds GB_UB, branching is triggered to explore potential improvements; otherwise, $\mathbf{w}_0$ is confirmed as optimal. By integrating adaptive bounding with depth-first node prioritization, the algorithm, as a modified tree searching method, systematically prunes non-promising regions while exhaustively evaluating viable solutions, thereby ensuring convergence to the global optimum.

The proposed algorithm offers distinct advantages over conventional methods:

(1) It directly enforces the cardinality constraint without resorting to continuous relaxation, thereby avoiding approximation errors and enabling the derivation of tighter lower bounds.
(2) For lower bound computation, our approach focuses on solving a concave optimization problem, which is computationally simpler and more scalable compared to the semidefinite programming techniques widely adopted in existing literature.
(3) The computational complexity stems from repeated inversions of high-dimensional covariance matrices, which we address by implementing the warm-start strategy from [17]. This approach stores and adaptively updates inverse matrices across parent-child subproblems using rank-one modifications, leveraging the branching structure to significantly reduce inversion time in empirical evaluations.

In conclusion, this approach balances exploration efficiency and solution quality, adhering to the theoretical guarantees established in Section 4. These innovations collectively enhance the algorithm's scalability for large-scale combinatorial optimization problems.

6. Numerical experiments

This section first confirms the feasibility of the algorithm, and then we compute the optimal portfolios for several sizes to compare its efficiency on different lower bounds. Then, the solution strategy is compared with the CPLEX solver. We also compare the sparse method with a cardinality constraint with other sparse methods. Finally, we compare the CCTE_FPR model that incorporates investor preferences with the original CCTE model.

6.1. Dataset and settings

To evaluate the lower bound's performance, we use the S&P 500 index datasets, which are highly representative and span a variety of businesses. The remaining three datasets, DJIA 30, NASDAQ 100, and NYSE, are used for out-of-sample experiments. Data are sourced from Yahoo Finance's historical database, covering daily adjusted closing prices of constituent stocks from different lengths from January 2014 to October 2024. To maintain data quality, assets not covering the full data period were removed to ensure consistency.

Throughout all tables, n denotes the total number of stocks and k the cardinality constraint, respectively. All stocks are divided into five industries according to information technology, health care, finance, consumer goods, and infrastructure manufacturing. Analyzing the number of nodes in

experimental studies provides insight into the efficiency of the branching strategy, the tightness of the bounds, and the overall scalability of the algorithm. Time refers to the CPU running time in seconds. We set the minimum expected portfolio return $s = 0.1$. It is important to note that, as defined in Section 3.2, **r** represents the vector of mean returns per asset over the entire training window, not a daily return. Therefore, $s = 0.1$ represents the minimum required portfolio return over the full training period, which is a reasonable and conventional constraint. For the solution algorithms, we set the convergence tolerance $\epsilon = 0.01$ and a maximum iteration limit of 1000 for the supergradient method. The cardinality k is varied as an experimental parameter, as shown in the tables. The model in Equation (3.4) does not include transaction costs in its optimization, but they are included in the out-of-sample performance evaluation. The specific numerical settings for all parameters are based on the work of Xu et al. [41], where the reasonableness of these settings was validated.

To assess index tracking performance, we employ a rolling window in which the optimal tracking portfolio is derived during a training period T_{tr} and the index movement is approximated using the derived portfolio during a testing period T_{tst}. The process is repeated multiple times to use the entire dataset, with the total testing days being $T - T_{tr}$, where T is the total data period. If the training and testing cycle is repeated τ times, it results in a total of $T_{tr} + \tau T_{tst}$ time frames for each experiment. Table 1 details each index, including window sizes T_{tr}, T_{tst}, and the total data period T.

Table 1. The information of the tested stock datasets.

Dataset	n	Period	T_{tr}	T_{tst}
DJIA 30	28	01/03/2020—-10/25/2024	70	28
S&P 500	50	01/03/2020—-05/31/2024	65	29
NASDAQ 100	83	01/03/2014—-08/16/2024	100	27
NYSE	96	01/11/2019—-10/25/2024	100	27

6.2. Algorithm performance

To validate the global optimization capability of our algorithm, we conduct an initial experiment selecting 5 stocks from a pool of 50 candidate assets and compare the results against an exhaustive enumeration approach. The comparative numerical outcomes are documented in Table 2.

Table 2. Verification of global optimality.

Ours		Enumeration	
Stock	Weight	Stock	Weight
DE	0.1364	DE	0.1364
ADI	0.2349	ADI	0.2349
NOW	0.1368	NOW	0.1368
MMC	0.1464	MMC	0.1464
MDLZ	0.2379	MDLZ	0.2379
Calculating time: 0.11 s		Calculating time: 216.19 s	

As seen in Table 2, our algorithm produces the same portfolio allocations and weights as the enumeration method. This result validates the algorithm's effectiveness in attaining the global optimum for the index tracking portfolio optimization model under cardinality constraints.

Given the pivotal importance of the lower bound within the branch-and-bound framework, we evaluate the computational performance of our proposed method by benchmarking it against several established bounds. Employing the supergradient method to compute all these bounds, we summarize the results in Table 3.

Table 3. Comparison among different lower bounds (S&P 500).

		Ours		Ball		Box		Continuous	
n	k	Nodes	Time(s)	Nodes	Time(s)	Nodes	Time(s)	Nodes	Time(s)
50	5	343	0.11	399	0.16	1085	0.35	1105	0.24
	6	803	0.21	961	0.30	2867	0.67	2969	0.49
	7	1865	0.33	3963	1.06	10,227	2.32	6971	1.20
	8	3249	0.52	4675	1.26	12,629	2.83	14,549	2.40

As shown in Table 3, our algorithm outperforms others in terms of the lower bound due to its tighter and more computationally efficient formulation. This advantage arises from the use of a tighter diagonal matrix lower bound, as presented in Section 4, which significantly enhances the pruning process in the branch-and-bound procedure. By providing a more accurate lower bound, our algorithm enables the earlier elimination of suboptimal branches, thereby reducing the search space and improving overall computational efficiency. This tighter bound not only accelerates the solution process but also ensures that the algorithm converges more quickly to an optimal or near-optimal solution, making it particularly effective in large-scale and complex optimization problems.

Table 4. Comparison among different lower bounds (common datasets).

			Ours		Ball		Box		Continuous	
	n	k	Nodes	Time(s)	Nodes	Time(s)	Nodes	Time(s)	Nodes	Time(s)
DJIA 30	28	5	13,855	1.49	17,105	2.79	11,121	1.85	21,287	2.81
		8	16,799	1.86	30,791	5.27	23,331	8.85	50,351	7.22
		10	14,317	1.84	27,153	5.01	11,721	4.05	47,715	6.95
NASDAQ 100	83	5	55,701	11.98	81,463	35.46	117,913	115.62	133,327	63.11
NYSE	96	5	81,083	45.01	274,141	357.07	134,635	103.58	161,379	133.24

To further validate the extensive practicability of our method, we evaluate it on additional datasets, as presented in Table 4. Consistent with the findings from the S&P 500 dataset, our lower bound consistently demonstrates superior performance in enhancing search efficiency across most cases. Moreover, for large-scale problems such as the NYSE dataset with $n = 96$, our approach reduces the computation time to only 45.01 seconds, which is 45.5% faster than the Box bound. This highlights the exceptional scalability of our algorithm, especially in high-dimensional portfolio optimization scenarios. The improved searching efficiency achieved by our tighter lower bound not only reinforces

its effectiveness in a wide range of problem sizes but also emphasizes its robustness and versatility in real-world applications.

In addition to deriving a tight lower bound, the selection of an appropriate solution algorithm is equally vital for enhancing the overall computational efficiency. Our empirical evaluation includes two primary components: a comparison of computational performance and an out-of-sample investment simulation.

We first conduct a comparative analysis between our proposed supergradient method and commercial mixed-integer quadratic programming solvers, specifically CPLEX. It is important to note that while both methods solve the same optimization problem (3.4), they employ different strategies: the CPLEX column refers to applying the default solver, whereas the supergradient column utilizes our custom branch-and-bound algorithm equipped with a much tighter Λ-based lower bound solved via the supergradient method. This difference in lower bound quality is the primary driver of our method's superior performance. The empirical findings, displayed in Table 5, demonstrate this computational advantage. The supergradient method leverages the dual problem's concave structure to update Lagrangian multipliers with linear complexity, whereas CPLEX experiences exponential growth in the number of explored nodes. These findings confirm that our methodology significantly alleviates the computational burden and offers enhanced scalability, presenting an efficient alternative to conventional solvers for large-scale problems.

To evaluate the practical efficacy and out-of-sample performance of the proposed portfolio allocation models, we conduct a comprehensive investment simulation. We replicate an actual investment strategy over the entire testing horizon: at the beginning of each designated test period, the portfolio is rebalanced based on the newly computed optimal weights, and all acquired returns from the previous period are reinvested. To evaluate net returns under realistic conditions, a standard U.S. market transaction cost model is applied, featuring a rate of \$0.005 per share with a \$1 minimum per trade. Assuming a \$1 million investment budget, this model underscores the transaction cost advantages inherent to sparse portfolio structures. This experimental setup closely adheres to the methodology described by Yamagata [11]. The primary metric for evaluating performance is the normalized accumulated return, denoted as *Ret*., calculated at the end of the entire testing period. This metric in Table 5 reflects the net profit-and-loss of the strategy after accounting for all periodic rebalancing and the associated transaction costs. This difference in lower bound quality is the primary reason for our method's superior performance, which consequently leads to better long-term returns.

Table 5. Comparison among different solving strategies (S&P 500).

		Supergradient			CPLEX		
n	k	Nodes	Time(s)	*Ret*.	Nodes	Time(s)	*Ret*.
50	5	343	0.11	1.1500	648,300	16.70	1.0017
	6	803	0.21	1.3780	3,685,801	99.67	0.9989
	7	1865	0.33	1.2981	14,040,303	945.94	1.0073

6.3. Tracking performance of different methods

In this subsection, we provide a comparative assessment of the proposed cardinality-constrained sparse method (TE_CC) against two other methods: The sparse model of 1.5 norm (TE_$\ell_{1.5}$) and the

original index tracking (TE) model without a sparsity constraint. The TE_$\ell_{1.5}$ model replaces the cardinality constraint with the $\ell_{1.5}$ norm, offering benefits such as reduced overfitting and enhanced interpretability while still promoting sparsity under the given weight and constraint conditions.

To assess the effectiveness of these models, we use the rolling window method introduced above for training and testing. We next introduce the following five evaluation criteria to evaluate the performance of portfolios generated by the TE_CC method. These criteria are used to evaluate model effectiveness in the following analysis. Experimental results for four datasets (S&P 500, NASDAQ 100, NYSE, and DJIA 30) are shown in Table 6.

E1: Mean Tracking Error (*MTE*). This metric quantifies the replication accuracy of the index-tracking portfolio. For a specified sparsity level, the *MTE* is calculated as:

$$MTE = \frac{1}{T - T_{tr}} \left\| \mathrm{diag}(\mathbf{XW}) - \mathbf{r}^b \right\|_2,$$

where the operator diag(·) extracts the diagonal elements of a matrix. Here, $\mathbf{X} \in \mathbb{R}^{(T-T_{tr})\times\tau}$ represents the test data, $\mathbf{r}^b \in \mathbb{R}^{(T-T_{tr})}$ is the benchmark return vector, and $\mathbf{W} \in \mathbb{R}_+^{\tau\times(T-T_{tr})}$ aggregates all testing period portfolio weights.

E2: Sharpe Ratio (*SR*). The risk-adjusted return is measured by the Sharpe ratio, defined as:

$$SR = \frac{Mean - R_f/N_Y}{Std},$$

where $R_f = 0.036$ denotes the risk-free rate (proxied by the U.S. 10-Year Treasury Rate), and N_Y is the annual trading frequency.

E3: Risk Reduction (*RR*). This metric assesses the relative risk profile compared to a naive $1/n$ diversification strategy:

$$RR = \frac{y_{naive}^\top \sum y_{naive}}{y^\top \sum y},$$

where y_{naive} represents the equally weighted portfolio. An $RR > 1$ indicates superior risk reduction of the proposed model over the benchmark, suggesting a more favorable risk-return trade-off.

E4: Excess Return (*ER*). The outperformance relative to a predefined benchmark return is calculated as:

$$ER = \frac{Mean - \lambda}{\lambda},$$

with λ being the minimum expected return specified in the optimization constraints.

E5: Information Ratio (*IR*). This ratio evaluates the consistency of excess returns relative to tracking error volatility:

$$IR = \frac{\frac{1}{k}\sum_{i=1}^{k} AER_i}{\sigma(AER)},$$

where $AER = [AER_1, AER_2, \ldots, AER_k]$ denotes the series of average excess returns and k indicates the number of nonzero solutions in the evaluated set.

From the results in Table 6, the cardinality-constrained sparse method (TE_CC) surpasses the other models in *Ret* and *ER* metrics, and performs comparably or better in other indicators. From Figure 1, it can be observed that the cumulative wealth of the solution obtained by the TE_CC is higher than that of the other models. Overall, these results demonstrate that the TE_CC is superior to the other models

while achieving a similar benefit-risk ratio. Additionally, we observe that the performance of *MTE* is not expected. To be specific, we can find that the TE_CC achieves a higher *MTE* on datasets other than NASDAQ 100, indicating that the high yield of sparse tracking is sometimes at the expense of accuracy. Similarly, the higher Sharpe ratio but lower information ratio on S&P 500 and NYSE may be due to the fact that the index tracking portfolio performs better than the risk, but the relative return of the portfolio relative to the benchmark is weak and lacks effective excess returns. A few experimental results that fall short of expectations reflect a risk-return tradeoff: Aggressive allocations boost returns but increase volatility. In conclusion, this experiment provides a comprehensive evaluation of the performance of TE_CC in comparison to the TE_$\ell_{1.5}$ and TE, ensuring the robustness and generalization of the cardinality constraint method.

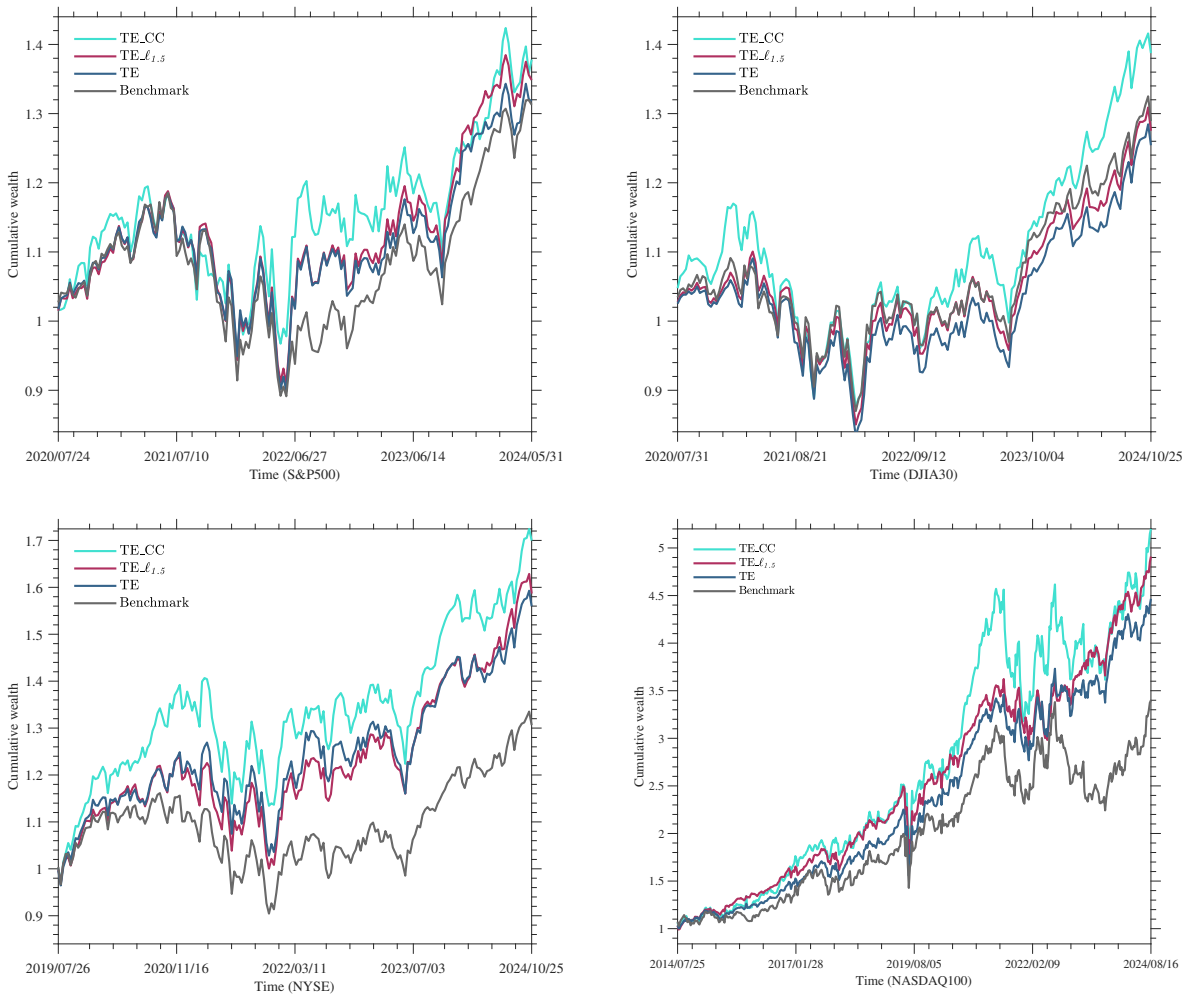

Figure 1. Cumulative wealth of different sparsity methods across four out-of-sample datasets.

To further elucidate the practical utility of our CCTE-FPR model, we conduct a sensitivity analysis on how portfolio sparsity impacts key performance metrics. Since our model allows fund managers to explicitly control the number of selected assets, it is crucial to understand the trade-offs between tracking accuracy, risk-adjusted returns, and potential transaction costs as k varies. We conducted a

series of experiments on the S&P 500 dataset using the CCTE-FPR model, systematically varying k from 5 to 25. The results, presented in Table 7, clearly illustrate these trade-offs.

Table 6. The tracking performance of TE_CC, TE_$\ell_{1.5}$, and TE for four datasets.

	S&P 500				NASDAQ 100			
	TE_CC	TE_$\ell_{1.5}$	TE	Benchmark	TE_CC	TE_$\ell_{1.5}$	TE	Benchmark
MTE(10^{-3})	1.0059	**0.5834**	0.6312	–	**0.6053**	0.9074	0.7647	–
Ret.	**1.3780**	1.3486	1.3180	1.3129	**5.2004**	4.9100	4.4654	3.3989
SR	**0.0732**	0.0674	0.0601	0.0534	0.1167	**0.1401**	0.1251	0.0817
RR	0.8078	1.0498	**1.0657**	–	0.6404	**1.1216**	1.0902	–
ER	**0.2569**	0.1535	0.0938	–	**0.3209**	0.2419	0.1851	–
IR	0.6461	**0.6654**	0.3760	–	**1.8437**	0.9274	0.8417	–
	NYSE				DJIA 30			
	TE_CC	TE_$\ell_{1.5}$	TE	Benchmark	TE_CC	TE_$\ell_{1.5}$	TE	Benchmark
MTE(10^{-3})	0.5773	**0.3416**	0.3797	–	0.6525	**0.3337**	0.3780	–
Ret.	**1.6997**	1.5867	1.5596	1.3064	**1.3875**	1.2752	1.2543	1.2894
SR	**0.1179**	0.1013	0.0963	0.0439	**0.0757**	0.0573	0.0524	0.0495
RR	**1.1499**	1.0073	1.0328	–	**1.0903**	0.9393	0.9137	–
ER	**0.7716**	0.5690	0.5296	–	**0.3565**	0.0749	0.0193	–
IR	2.8848	**3.5950**	3.0105	–	**0.9628**	0.3953	0.0900	–

Table 7. Impact of sparsity level k on CCTE-FPR model performance (S&P 500).

k	*MTE*(10^{-3})	*Ret.*	*SR*	*RR*	*ER*	*IR*
5	1.2217	1.3512	0.0701	0.7850	0.2491	0.6015
10	1.0059	1.3780	0.0732	0.8078	0.2569	0.6461
15	0.8734	1.4015	0.0758	0.8990	0.2788	0.7002
20	0.7810	**1.4122**	**0.0765**	0.9514	**0.2903**	**0.7210**
25	**0.7153**	1.4090	0.0740	**0.9976**	0.2885	0.7104

Analyzing the data in Table 7, we observe several key trends. As expected, as k increases, the tracking error *MTE* monotonically declines; this is because incorporating more assets allows the portfolio to replicate the benchmark index more precisely. In contrast, the return-oriented metrics, such as *Ret.*, *SR*, *ER*, and *IR*, do not increase monotonically with k. Instead, they appear to peak around $k \approx 20$. This suggests that while adding assets initially enhances returns through diversification, the marginal benefit diminishes beyond this point, and performance may slightly decrease, possibly due to overfitting or the inclusion of assets with lower risk-adjusted returns. Concurrently, the *RR* metric steadily increases and approaches 1, indicating that as k grows, the portfolio's risk profile gradually converges with that of a naive diversification strategy. In summary, this analysis provides investors with a clear decision-making basis: for the S&P 500 dataset, a cardinality of $k \approx 20$ appears to yield a near-optimal balance, effectively minimizing *MTE* while maximizing risk-adjusted returns. This finding also highlights the practical

value of our model in enabling customized portfolio construction by controlling k.

6.4. Comparative analysis of different index tracking models

To demonstrate the specific importance and impact of the novel FPR constraint, we conduct a direct comparative analysis. We evaluate two models: a basic cardinality-constrained tracking error model (Model 1) and an extension augmented with investor fuzzy preference relations (Model 2). This comparison directly isolates the influence of adding the FPR constraint to a sparse tracking model. We utilize the same multi-period investment rolling window method as before to demonstrate the numerical performance. The results for both models across the four datasets are presented in Table 8.

Table 8. The numerical results of Model 1 and Model 2 for four datasets.

	S&P 500		NASDAQ 100		NYSE		DJIA 30	
	Model 1	Model 2	Model 1	Model 2	Model 1	Model 2	Model 1	Model 2
$MTE(10^{-3})$	**0.9981**	1.0059	0.6091	**0.6053**	0.5793	**0.5773**	0.6613	**0.6525**
Ret.	1.2870	**1.3780**	4.9471	**5.2004**	**1.7029**	1.6997	1.3767	**1.3875**
SR	0.0548	**0.0732**	0.1122	**0.1167**	**0.1186**	0.1179	0.0744	**0.0757**
RR	0.8078	0.8078	**0.6433**	0.6404	1.1286	**1.1499**	1.0903	1.0903
ER	0.0856	**0.2569**	0.2887	**0.3209**	**0.7751**	0.7716	**0.3616**	0.3565
IR	0.2170	**0.6461**	1.6486	**1.8437**	**2.8880**	2.8848	**0.9633**	0.9628

Crucially, Model 2 consistently delivers higher cumulative returns, superior Sharpe ratios, and better information ratios across most datasets. This improvement can be explained by the fact that the FPR constraint acts as a form of expert-informed regularization. Model 1 optimizes only on historical data, which can lead it to select assets or sector allocations that, while mathematically optimal for past tracking, are intuitively risky or misaligned with forward-looking market sentiment. The FPR constraint, by encoding the investor's subjective expertise, guides the optimization away from these historically optimized but potentially suboptimal solutions. In essence, the FPR constraint injects valuable qualitative information into the quantitative model, helping it prioritize assets with stronger perceived fundamentals or better alignment with the investor's strategy. This leads to an improvement in overall risk-adjusted returns, as the portfolio is not just tracking an index, but doing so in a way that is filtered through an additional layer of human expertise.

7. Conclusions

This study pioneered the integration of cardinality constraints to address sparsity requirements in index-tracking models while establishing a groundbreaking framework that unifies investors' FPRs with investment vectors, thereby constructing an optimal portfolio model that explicitly account for investor-specific characteristics. For cardinality constraint implementation, we developed an efficient matrix decomposition technique that significantly reduced computational complexity while achieving sparse solutions that satisfied cardinality requirements. Building upon this foundation, the proposed sparse model achieved computational efficiency gains through a depth-first branch-and-bound algorithm. The

flexible constraint framework can ensure practical applicability for large-scale portfolio optimization.

The key theoretical innovation of this study lies in the first integration of additively consistent FPRs into cardinality-constrained index tracking models, which enabled joint optimization of objective tracking performance and subjective behavioral rationality. Regarding FPRs integration, we established the theoretical connection between FPRs and cardinality-constrained index tracking models, which enabled joint optimization of tracking performance and investor preferences. As far as the practical significance of investment is concerned, investors could compare FPRs with inherent preferences to stabilize rational investment behaviors and identify and modify irrational preferences through portfolio feedback. This highlights the dual value of our research. The unknown preference establishes the crucial foundation for a unified qualitative-quantitative framework. The value of our contribution is thus twofold: (1) In the immediate, practical case of unknown preferences, it provides a "rationality check" and a vital decision-support tool. (2) For future research, this same framework is robust and extensible, paving the way for the integration of known or partially known FPRs, allowing them to serve as explicit, prescriptive guides for portfolio construction.

The shortcoming of the existing research is that the framework's reliance on historical covariance matrices may prove to be inadequate during structural market shifts, and its integration of inaccurate cognitive biases will lead to subjective risk enhancement. Future work will explore online preference learning and multi-period dynamic portfolio extensions. Subsequent research will prioritize adaptive covariance estimation via machine learning, coupled with data-driven preference optimization through behavioral analytics to reconcile investor priorities. This work will bridge algorithmic innovation with behavioral finance and offer a rigorous yet practical tool for personalized index tracking in real-world investment scenarios.

Author contributions

Chun Wang: Data curation, Formal analysis, Validation, Writing – Original Draft;
Zhongming Wu: Conceptualization, Methodology, Supervision, Writing – Review and Editing;
Yu Yuan: Conceptualization, Methodology, Supervision, Writing – Review and Editing;
Wei Xu: Conceptualization, Methodology, Writing – Review and Editing.

Use of Generative-AI tools declaration

The authors declare they have not used Artificial Intelligence (AI) tools in the creation of this article.

Acknowledgments

This research was partially supported by the National Natural Science Foundation of China (Nos. 12471291, 12001286), the Natural Science Foundation of Jiangsu Province (No. BK20241899), and the Ministry of Education of China Humanities and Social Sciences Research Project (No. 24YJC630237).

Conflict of interest

All authors declare no conflicts of interest in this paper.

References

1. J. Li, H. Nie, T. Chai, F. L. Lewis, Reinforcement learning for optimal tracking of large-scale systems with multitime scales, *Sci. China Inf. Sci.*, **66** (2023), 1–25. https://doi.org/10.1007/s11432-022-3796-2
2. A. Granzer-Guay, R. H. Kwon, Risk-return adaptive receding Horizon Index Tracking Strategy, *Eng. Econ.*, **69** (2024), 189–212. https://doi.org/10.1080/0013791X.2024.2402688
3. A. Ling, J. Li, L. Wen, Y. Zhang, When trackers are aware of ESG: Do ESG ratings matter to tracking error portfolio performance, *Econ. Model.*, **125** (2023), 106346. https://doi.org/10.1016/j.econmod.2023.106346
4. I. Rais, S. Alam, C. Kumar, S. S. Meghwani, Bi-objective Enhanced Index Tracking: Performance Analysis of Meta-heuristic Algorithms with Real-World Constraints, *Proceedings of Third International Conference on Computing and Communication Networks*, (2023), 453–469. https://doi.org/10.1007/978-981-97-2671-4_35
5. M. C. Yuen, S. C. Ng, M. F. Leung, H. Che, A metaheuristic-based framework for index tracking with practical constraints, *Complex Intell. Syst.*, **9** (2022), 3469–3469. https://doi.org/10.1007/s40747-022-00918-z
6. L. Zhao, G. Li, S. Penev, Regularized distributionally robust optimization with application to the index tracking problem, *Ann. Oper. Res.*, **337** (2024), 397–424. https://doi.org/10.1007/s10479-023-05726-3
7. T. Zhang, S. Lai, Portfolio selection balancing concentration and diversification, *J. Ind. Manag. Optim.*, **10** (2025), 3618–3647. https://doi.org/10.3934/jimo.2025025
8. J. E. Beasley, N. Meade, T. J. Chang, An evolutionary heuristic for the index tracking problem, *Eur. J. Oper. Res.*, **148** (2003), 621–643. http://dx.doi.org/10.1016/S0377-2217(02)00425-3
9. J. Brodie, I. Daubechies, C. De Mol, D. Giannone, I. Loris, Sparse and stable Markowitz portfolios, *Proc. Natl. Acad. Sci.*, **106** (2009), 12267–12272. https://doi.org/10.1073/pnas.0904287106
10. X. P. Li, Z. L. Shi, C. S. Leung, H. C. So, Sparse Index Tracking With K-Sparsity or ϵ-Deviation Constraint via ℓ_0-Norm Minimization, *IEEE Trans. Neural. Netw. Learn. Syst.*, **34** (2023), 10930–10943. https://doi.org/10.1109/TNNLS.2022.3171819
11. E. Yamagata, S. Ono, Sparse Index Tracking: Simultaneous Asset Selection and Capital Allocation via ℓ_0-Constrained Portfolio, *IEEE Open. J. Signal. Process.*, **5** (2024), 810–819. https://doi.org/10.1109/OJSP.2024.3389810
12. X. Wu, R. Liang, Z. Zhang, Z. Cui, Multi-block linearized alternating direction method for sparse fused Lasso modeling problems, *Appl. Math. Model.*, **137** (2025), 115694. https://doi.org/10.1016/j.apm.2024.115694
13. H. Jiang, W. Zheng, Y. Dong, Sparse and robust estimation with ridge minimax concave penalty, *Inf. Sci.*, **571** (2021), 154–174. https://doi.org/10.1016/j.ins.2021.04.047
14. F. Khan, S. Muhammadullah, A. Sharif, C. C. Lee, The role of green energy stock market in forecasting China's crude oil market: An application of IIS approach and sparse regression models, *Energy Econ.*, **130** (2024), 107269. https://doi.org/10.1016/j.eneco.2023.107269

15. H. Zhao, L. Kong, H. D. Qi, Optimal portfolio selections via $\ell_{1,2}$-norm regularization, *Comput. Optim. Appl.*, **80** (2021), 853–881. https://doi.org/10.1007/s10589-021-00312-4

16. Z. Wu, K. Sun, Z. Ge, Z. Allen-Zhao, T. Zeng, Sparse portfolio optimization via ℓ_1 over ℓ_2 regularization, *Eur. J. Oper. Res.*, **319** (2024), 820–833. https://doi.org/10.1016/j.ejor.2024.07.017

17. D. Bertsimas, R. Shioda, Algorithm for cardinality-constrained quadratic optimization, *Comput. Optim. Appl.*, **43** (2009), 1–22. https://doi.org/10.1007/s10589-007-9126-9

18. R. Moral-Escudero, R. Ruiz-Torrubiano, A. Suárez, Selection of optimal investment portfolios with cardinality constraints, *2006 IEEE International Conference on Evolutionary Computation*, (2006), 2382–2388. http://dx.doi.org/10.1109/CEC.2006.1688603

19. K. Michell, W. Kristjanpoller, Strongly-typed genetic programming and fuzzy inference system: An embedded approach to model and generate trading rules, *Appl. Soft Comput.*, **90** (2020), 106169. https://doi.org/10.1016/j.asoc.2020.106169

20. X. Zhao, J. Dou, Bi-objective integrated supply chain design with transportation choices: a multi-objective particle swarm optimization, *J Ind. Manag. Optim.*, **15** (2019), 1263–1288. https://doi.org/10.3934/jimo.2018095

21. K. Smith-Miles, Understanding instance hardness for optimisation algorithms: Methodologies, open challenges and post-quantum implications, *Appl. Math. Model.*, **142** (2025), 115965. https://doi.org/10.1016/j.apm.2025.115965

22. M. C. Yuen, S. C. Ng, M. F. Leung, H. Che, Metaheuristics for index-tracking with cardinality constraints, *2021 11th International Conference on Information Science and Technology*, (2021), 646–651. https://doi.org/10.1109/ICIST52614.2021.9440584

23. D. Maringer, O. Oyewumi, Optimal construction and rebalancing of index-tracking portfolios, *Eur. J. Oper. Res.*, **264** (2018), 370–387. https://doi.org/10.1016/j.ejor.2017.06.055

24. S. Ma, Y. Gao, B. Zhang, Output-space branch-and-bound reduction algorithm for solving generalized linear multiplicative programming programs, *J. Appl. Math. Comput.*, **2024**, 1–31. https://doi.org/10.1007/s12190-024-02202-4

25. K. Benidis, Y. Feng, D. P. Palomar, Sparse Portfolios for High-Dimensional Financial Index Tracking, *IEEE Trans. Signal Process.*, **66** (2018), 155–170. https://doi.org/10.1109/TSP.2017.2762286

26. Y. Zheng, B. Chen, T. M. Hospedales, Y. Yang, Index Tracking with Cardinality Constraints: A Stochastic Neural Networks Approach, *Proc. AAAI Conf. Artif. Intell.*, **34** (2020), 1242–1249. https://doi.org/10.1609/aaai.v34i01.5478

27. F. S. G. Constante, J. C. López, M. J. Rider, Optimal reactive power dispatch with discrete controllers using a branch-and-bound algorithm: A semidefinite relaxation approach, *IEEE Trans. Power Syst.*, **36** (2021), 4539–4550. https://doi.org/10.1109/TPWRS.2021.3056637

28. X. Zheng, Y. Pan, Z. Hu, Perspective Reformulations of Semicontinuous Quadratically Constrained Quadratic Programs, *INFORMS J. Comput.*, **33** (2020), 163–179. https://doi.org/10.1287/ijoc.2019.0925

29. X. Deng, W. Li, Y. Liu, Hesitant fuzzy portfolio selection model with score and novel hesitant semi-variance, *Comput. Ind. Eng.*, **164** (2022), 107879. https://doi.org/10.1016/j.cie.2021.107879

30. S. Nayak, S. Maharana, An efficient fuzzy mathematical approach to solve multi-objective fractional programming problem under fuzzy environment, *J. Appl. Math. Comput.*, **69** (2023), 2873–2899. https://doi.org/10.1007/s12190-023-01860-0

31. S. Orlovsky, Decision-making with a fuzzy preference relation, *Fuzzy Sets Syst.*, **1** (1978), 155–167. http://dx.doi.org/10.1016/B978-1-4832-1450-4.50077-8

32. Y. Xu, Q. Wang, F. Chiclana, E. Herrera-Viedma, A local adjustment method to improve multiplicative consistency of fuzzy reciprocal preference relations, *IEEE Trans. Syst. Man Cybern. Syst.*, **53** (2023), 5702–5714. https://doi.org/10.1109/TSMC.2023.3275167

33. T. Tanino, Fuzzy preference orderings in group decision making, *Fuzzy Sets Syst.*, **12** (1984), 117–131. http://dx.doi.org/10.1016/0165-0114(84)90032-0

34. C. Li, Y. Dong, Y. Xu, F. Chiclana, E. Herrera-Viedma, F. Herrera, An overview on managing additive consistency of reciprocal preference relations for consistency-driven decision making and fusion: Taxonomy and future directions, *Inf. Fusion*, **52** (2019), 143–156. http://dx.doi.org/10.1016/j.inffus.2018.12.004

35. J. Ma, Z. P. Fan, Y. P. Jiang, J. Y. Mao, L. Ma, A method for repairing the inconsistency of fuzzy preference relations, *Fuzzy Sets Syst.*, **157** (2006), 20–33. https://doi.org/10.1016/j.fss.2005.05.046

36. Z. Xu, Uncertain multi-attribute decision making: Methods and applications, Springer, Berlin Heidelberg, 2015.

37. W. Guo, Z. Gong, X. Xu, E. Herrera-Viedma, Additive and multiplicative consistency modeling for incomplete linear uncertain preference relations and its weight acquisition, *IEEE Trans. Fuzzy Syst.*, **29** (2020), 805–819. https://doi.org/10.1109/TFUZZ.2020.2965909

38. A. Yazidi, M. Ivanovska, F. M. Zennaro, P. G. Lind, E. Herrera-Viedma, A new decision making model based on Rank Centrality for GDM with fuzzy preference relations, *Eur. J. Oper. Res.*, **297** (2022), 1030–1041. https://doi.org/10.1016/j.ejor.2021.05.030

39. S. S. Dey, A. Kazachkov, A. Lodi, G. Munoz, Cutting plane generation through sparse principal component analysis, *SIAM J. Optim.*, **32** (2022), 1319–1343. https://doi.org/10.1137/21M1399956

40. F. S. Gharehchopogh, Quantum-inspired metaheuristic algorithms: comprehensive survey and classification, *Artif. Intell. Rev.*, **56** (2023), 5479–5543. https://doi.org/10.1007/s10462-022-10280-8

41. W. Xu, J. Tang, K. F. C. Yiu, J. W. Peng, An efficient global optimal method for cardinality-constrained portfolio optimization, *INFORMS J. Comput.*, **36** (2024), 690–704. https://doi.org/10.1287/ijoc.2022.0344

42. W. Guo, G. Zhang, X. Chen, Portfolio selection models considering fuzzy preference relations of decision makers, *IEEE Trans. Syst. Man Cybern. Syst.*, **54** (2024), 2254–2265. https://doi.org/10.1109/TSMC.2023.3342038

43. D. Maringer, O. Oyewumi, Index tracking with constrained portfolios, *Intell. Syst. Account. Finance Manag.*, **15** (2007), 57–71. https://doi.org/10.1002/isaf.285

44. A. Scozzari, F. Tardella, S. Paterlini, T. Krink, Exact and heuristic approaches for the index tracking problem with UCITS constraints, *Ann. Oper. Res.*, **205** (2013), 235–250. https://doi.org/10.1007/s10479-012-1207-1

45. K. J. Oh, T. Y. Kim, S. Min, Using genetic algorithm to support portfolio optimization for index fund management, *Expert Syst. Appl.*, **28** (2005), 371–379. https://doi.org/10.1016/j.eswa.2004.10.014

46. A. Takeda, M. Niranjan, J. y. Gotoh, Y. Kawahara, Simultaneous pursuit of out-of-sample performance and sparsity in index tracking portfolios, *Comput. Manag. Sci.*, **10** (2013), 21–49. https://doi.org/10.1007/s10287-012-0158-y

47. K. Andriosopoulos, M. Doumpos, N. C. Papapostolou, P. K. Pouliasis, Portfolio optimization and index tracking for the shipping stock and freight markets using evolutionary algorithms, *Transp. Res. Part E Logist. Transp. Rev.*, **52** (2013), 16–34. https://doi.org/10.1016/j.tre.2012.11.006

48. R. Ruiz-Torrubiano, A. Suárez, A hybrid optimization approach to index tracking, *Ann. Oper. Res.*, **166** (2009), 57–71. https://doi.org/10.1007/s10479-008-0404-4

49. D. X. Shaw, S. Liu, L. Kopman, Lagrangian relaxation procedure for cardinality-constrained portfolio optimization, *Optim. Methods Softw.*, **23** (2008), 411–420. https://doi.org/10.1080/10556780701722542

50. J. Gao, D. Li, Optimal cardinality-constrained portfolio selection, *Oper. Res.*, **61** (2013), 745–761. https://doi.org/10.1287/opre.2013.1170

AIMS Press

© 2026 the Author(s), licensee AIMS Press. This is an open access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0)