Journal of Industrial and Management Optimization

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

JIMO, 22(4): 1726–1757.
DOI:10.3934/jimo.2026064
Received: 15 December 2025
Revised: 08 February 2026
Accepted: 24 February 2026
Published: 11 March 2026

Research article

An efficient ADMM for multi-period sparse behavioral portfolio optimization based on cumulative prospect theory

Qingyang Wang[1], Kunpeng Zhu[2], Yanjing Guo[2,*], Liu Yang[2] and Zhongming Wu[2]

[1] School of Management and Engineering, Nanjing University, Nanjing 210093, China

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

* **Correspondence:** Email: 202311240002@nuist.edu.cn.

Abstract: Traditional portfolio optimization approaches typically presume a completely rational market, overlooking investors' behavioral inclinations and the complexity of sparsity in high-dimensional data. To tackle these concerns, this study presents a novel multi-period sparse behavioral portfolio optimization model. In a multi-period context, the model incorporates a utility function based on cumulative prospect theory, effectively capturing the irrational behavioral characteristics of investors. By leveraging ℓ_1-norm regularization, it attains portfolio sparsity in each period and minimizes turnover across periods. Next, we proposed a hybrid algorithm that integrates the alternating direction method of multipliers with the pooling-adjacent-violators algorithm to efficiently solve the newly formulated model. Furthermore, the framework incorporates environmental, social, and governance factors to evaluate their influence on investors' behavioral portfolios. Numerical experiments and empirical analyses demonstrated that the proposed method can efficiently solve the model, and the new model is capable of reducing risk and transaction costs.

Keywords: portfolio optimization; multi-period; ℓ_1-norm regularization; cumulative prospect theory; alternating direction method of multipliers
Mathematics Subject Classification: 91G10, 65K05, 90C06

1. Introduction

Portfolio optimization is a fundamental challenge in investment management, seeking an optimal balance between risk and return through the strategic allocation of available funds [1]. However, traditional portfolio optimization models are typically based on the assumption that the market is perfectly rational. They focus on objective rationality while often overlooking the influence of investors' behavioral preferences on decision-making [2]. Recently, behavioral finance has received increasing

attention. A growing body of studies indicates that investors frequently display irrational behaviors, such as loss aversion and probability weighting regarding tail events [3,4]. Consequently, it is extremely important to incorporate behavioral preferences into portfolio optimization models, not only to capture irrationality but also to enhance robustness against downside risks. Moreover, as data dimensionality and the complexity of multi-period investments increase, effectively identifying a small subset of crucial investment opportunities in high-dimensional spaces to manage costs has become a pressing issue [5]. Addressing these difficulties, this study incorporates the value function from cumulative prospect theory (CPT) into a multi-period portfolio selection framework and leverages regularization approaches. This integration aims to investigate how sparse optimization can discipline behavioral portfolios, balancing theoretical utility with practical investability.

In practical investment scenarios, single-period portfolio optimization models frequently prove inadequate in handling crucial aspects such as transaction costs, taxes, the dynamic fluctuations of asset returns, and the delicate balance between short-term and long-term interests. In contrast, multi-period portfolio optimisation models are better equipped to manage these limitations, as noted by [6]. As a result, scholars have endeavored to expand single-period models into multi-period frameworks. Li [7] was the earliest to propose the multi-period mean-variance model. Cui [8] later advanced this work by formulating a multi-stage mean-variance problem without allowing short sales. Chen [9] examined multi-period risk parity techniques and optimal investment decisions under dynamic settings. Additionally, Moghadam [10] introduced a robust multi-period framework that controls risk through interval semi-absolute deviation and solved it with meta-heuristic optimization methods. A multi-period risk-parity portfolio was designed by Li [6], with optimization performed through mean-variance model predictive control. Dai [11] explored the optimization of uncertain multi-period investment portfolios by integrating dynamic risk preferences, with a particular emphasis on minimizing transaction frequency within an uncertain environment.

The multi-period models outperform single-period models due to their capacity to enable investors to dynamically adjust their portfolios in response to the ever-changing market information. This adaptability effectively lessens the influence of the stock market's inherent volatility and uncertainty on investment decisions. One critical limitation of multi-period portfolio optimization lies in its vulnerability to the curse of dimensionality. This phenomenon increases the complexity of the problem and extends the solution time. To address these issues and account for the unique features of financial return data — particularly the strong correlations between portfolio weights in consecutive periods — Tibshirani et al. [12] first introduced the least absolute shrinkage and selection operator (LASSO) model, and later extended it to the fused LASSO [13]. Building on this foundation, Corsaro [14] designed a regularization framework with two separate penalty terms, where the first one restricts the magnitude of portfolio weights to mitigate overly concentrated positions arising from parameter estimation errors. In this way, it achieves sparse solutions and cuts down holding costs [15, 16]. The second penalty factor imposes a penalty on the differences between weight vectors in adjacent periods. Consequently, this approach lowers portfolio turnover and explicitly incorporates transaction costs into the optimization process, thereby enhancing the overall cost-effectiveness of the allocation [17]. Recently, the solving methods such as stochastic gradient descent and the Lagrange–Newton algorithm for the variants of sparse portfolio selection problems have also been investigated [18, 19].

While these regularization techniques effectively address the computational and structural challenges of high-dimensional portfolios, they often rely on simplified assumptions regarding investor

behavior. In traditional portfolio selection theory, investors are generally assumed to be rational entities that solely concentrate on minimizing risk and maximizing returns, disregarding individual preference disparities [20]. However, real-world investors are often driven by asymmetric risk preferences, exhibiting distinct loss aversion and sensitivity to tail events. Elements such as cognitive capabilities, accumulated experience, and psychological characteristics exert a substantial influence on investment decisions. Subjective preferences, which can have a positive impact on investment results, ought not to be neglected in portfolio construction. They offer personalized strategies that can lead to higher returns in dynamic and uncertain markets. To rigorously model these behavioral mechanisms, CPT [21,22] has emerged as the dominant paradigm in behavioral finance, providing valuable perspectives on decision-making under risk and uncertainty. Unlike standard utility functions, CPT incorporates probability weighting, which overweights extreme outcomes. This mechanism is particularly valuable in portfolio optimization for capturing tail risks that mean-variance models often overlook. CPT plays a significant part in behavioral finance and finds applications in stock allocation [23], efficiency assessment [24], and consumer behavior models [25]. Beyond CPT, researchers have explored other avenues to capture behavioral features. For instance, Rodriguez et al. [26] proposed a diversified behavioral portfolio model using fuzzy logic to handle investor satisfaction levels regarding return, risk (including skewness and kurtosis), and entropy. However, CPT remains the dominant paradigm for explicitly modeling loss aversion and probability weighting.

Scholars have incorporated CPT into portfolio selection. For example, Vanbilsen and Luxenberg [27,28] derived optimal consumption-investment policies under cumulative prospect theory. In a related multi-period setting, Shi et al. [29] developed and analyzed dynamic behavioral portfolio choices explicitly based on the CPT preference structure. Srivastava [30] investigated portfolio selection via the risk criteria of CPT. Moreover, certain scholars have merged behavioral finance theory with traditional models to devise more comprehensive portfolio selection frameworks. Hens et al. [31] performed a detailed comparison between asset allocations obtained by applying CPT optimization on the mean-variance efficient frontier. Fulga [32] proposed a unified portfolio selection model that explicitly embeds investor-specific preferences into the classical mean-risk framework, thereby offering deeper insight into the decision-making process of loss-averse individuals. Bi et al. [33] introduced probability distortion into the traditional mean-variance portfolio selection problem, constructing a behavioral mean-variance model that takes into account biases in investors' probability judgments. Acknowledging the significance of sustainability principles, Kaucic et al. [34] developed a bi-objective model that optimizes both financial and sustainable CPT value functions by incorporating environmental, social, and governance (ESG) information. Wei et al. [35] proposed a dynamic sentiment adjustment model for multi-period sparse portfolio selection. This model introduces prospect theory to gauge investors' asymmetric sentiments and employs an ℓ_0-regularization term to encourage sparsity, with the aim of minimizing portfolio risk. More recently, Wu et al. [36] developed a CPT-based portfolio model that incorporates an asset exploration mechanism and solved it via a symmetric alternating direction method of multipliers (ADMM) algorithm. While their work demonstrates the efficacy of ADMM-type methods for CPT problems, it primarily focuses on expanding the asset universe rather than controlling turnover and stability in high-dimensional settings.

Despite these advancements, few studies have systematically integrated behavioral preferences, multi-period constraints, and sparse regularization into a unified framework. This paper endeavors to fill this gap by developing a multi-period sparse behavioral portfolio selection model that embeds

CPT-based utility within the established fused LASSO model. This framework accommodates not only asymmetric risk attitudes but also leverages sparse constraints to discipline the high turnover often associated with behavioral strategies.

Specifically, our contributions are summarized as follows:

(i) We construct a multi-period sparse behavioral portfolio framework by integrating CPT-based utility with fused LASSO regularization, which accommodates asymmetric risk preferences while mitigating high turnover and transaction costs inherent in dynamic environments.

(ii) We adopt a tailored ADMM approach combined with the pooled adjacent violators (PAV) algorithm to effectively solve the resulting nonconvex optimization problem, ensuring computational tractability and providing rigorous convergence analysis for the solution procedure.

(iii) We extend the model to incorporate ESG factors for socially responsible investing and demonstrate via extensive numerical experiments that the proposed approach yields superior risk-adjusted returns and sparsity compared to traditional benchmarks.

The paper is organized in the subsequent sections as follows. In Section 2, we outline the theoretical foundations and mathematical framework of cumulative prospect theory within the context of behavioral preferences in portfolio optimization. Section 3 then develops the multi-period sparse behavioral portfolio optimization model. Section 4 delineates the solution algorithm for the proposed model. Section 5 analyzes the convergence characteristics of the algorithm. Section 6 expands the model by incorporating ESG factors into the multi-period sparse behavioral portfolio framework. Section 7 conducts numerical experiments and empirical analyses to validate the efficacy of the model. Lastly, Section 8 concludes the paper.

2. Cumulative prospect theory

Within the framework of cumulative prospect theory (CPT), investors' behavioral preferences are characterized by a value function and a probability weighting function [22]. Consistent with prospect theory, the CPT value function is nonconvex over gains and convex over losses, reflecting diminishing sensitivity. It is given by

$$U(z) = \begin{cases} -\mu(B - z)^{\alpha}, & z \le B, \\ (z - B)^{\rho}, & z > B, \end{cases}$$

where α and ρ denote the risk attitude parameters in the loss and gain domains, respectively, $\mu > 1$ is the loss aversion coefficient, and B is the reference point.

Since the value function is non-differentiable at $z = B$, the reference point is fixed at $B = 0$ [22]. This choice is well suited for multi-period portfolio optimization, as it avoids the overfitting and computational burden associated with dynamic reference points, which may induce excessive path dependence and weaken intertemporal consistency [37]. A fixed reference point provides a stable benchmark for evaluating gains and losses, thereby simplifying the model and improving robustness in long-term dynamic investment.

Moreover, the CPT value function and probability weighting mechanism are particularly suitable for multi-period decision-making, as they capture investors' aversion to extreme losses while preserving

intertemporal consistency [22]. Loss aversion mitigates the accumulation of short-term losses due to frequent rebalancing, whereas the overweighting of low-probability extreme events enhances tail risk management [38]. These features motivate the incorporation of sparse regularization within the CPT framework.

Based on these considerations, we formulate the following CPT-based portfolio optimization model:

$$\min_{\mathbf{x}} -\sum_{i=1}^{N} c_i U((\boldsymbol{R}\mathbf{x})_{(i)}), \quad \text{s.t.} \quad \mathbf{x} \in \mathcal{X}.$$

This model is a sample average approximation (SAA) of a continuous portfolio optimization problem. Here, N is the number of return scenarios, $\boldsymbol{R} \in \mathbb{R}^{N\times d}$ is the return scenario matrix, $\mathbf{x}$ denotes the portfolio weight vector, and $(\boldsymbol{R}\mathbf{x})_{(i)}$ is the i-th order statistic of portfolio returns. The decision weights c_i are defined via the probability weighting function $\omega(\cdot)$ as $c_i = \omega\left(\frac{N-i+1}{N}\right) - \omega\left(\frac{N-i}{N}\right)$, and $\mathcal{X}$ denotes the feasible set, typically including non-negativity and budget constraints.

The main challenges of this problem stem from the nonconvex and nonsmooth structure of the value and probability weighting functions, as well as the computational complexity induced by sorting operations. A tailored solution framework is developed in the subsequent sections to efficiently address these challenges.

3. Multi-period sparse behavioral portfolio optimization model

In practical investment processes, particularly within complex market environments and when dealing with high-dimensional data, [14] introduced the fused LASSO model to efficiently promote sparse portfolio selection. Let m denote the number of rebalancing dates, with decisions made in period j retained within the interval $[j, j+1)$. Let n represent the number of assets available for selection at each rebalancing date. The investment weights for the portfolio selection over the entire investment cycle take the form $\mathbf{w} = [\mathbf{w}_1^\top, \mathbf{w}_2^\top, \ldots, \mathbf{w}_m^\top]^\top \in \mathbb{R}^N$, where $N = mn$. The expected return vector associated with period j is denoted as $\mathbf{r}_j \in \mathbb{R}^n$ and the covariance matrix in period j is defined as $\boldsymbol{H_j} \in \mathbb{R}^{n\times n}$. Consequently, the fused LASSO model proposed by [14] takes the following form:

$$\begin{aligned}
&\min_{\mathbf{w}} \sum_{j=1}^{m}\left[\frac{1}{2}\mathbf{w}_j'\boldsymbol{H}_j\mathbf{w}_j + \tau_1\left\|\mathbf{w}_j\right\|_1\right] + \tau_2\sum_{j=1}^{m-1}\left\|\mathbf{w}_{j+1} - \mathbf{w}_j\right\|_1 \\
&\text{s.t. } \mathbf{w}_1^\top 1_n = \xi_{\text{init}}, \\
&\quad \mathbf{w}_j^\top 1_n = \left(1_n + \mathbf{r}_{j-1}\right)^\top \mathbf{w}_{j-1},\ j = 2, \ldots, m, \\
&\quad (1_n + \mathbf{r}_m)^\top \mathbf{w}_m = \xi_{\text{term}},
\end{aligned} \tag{3.1}$$

where $\tau_1 > 0$ and $\tau_2 > 0$ are regularization parameters that govern the strength of the two penalty terms. Appropriate selection of these hyperparameters is crucial to obtain a portfolio that is both financially meaningful from a financial perspective and consistent with the observed return distribution. Here, ξ_{init} represents the initial wealth, ξ_{term} the target terminal wealth, and $\mathbf{1}_n \in \mathbb{R}^n$ the n-dimensional vector of ones.

Although the fused LASSO model is grounded in traditional investment principles and effectively enhances portfolio sparsity and cost control, it still relies on the assumption of fully rational investors

with complete information. As a result, it fails to account for psychological factors and behavioral biases that systematically influence investment behavior. To relax this assumption, CPT is incorporated to capture asymmetric risk attitudes and tail-risk sensitivity. Moreover, embedding CPT preferences within a multi-period sparse regularization framework helps discipline behavioral distortions, mitigating overreaction and excessive portfolio turnover in dynamic investment environments. Thus, we combine the fused LASSO regularization technique with the utility function of cumulative prospect theory, thus developing a multi-period sparse behavioral portfolio (MSBP) model that incorporates investor preferences:

$$\begin{aligned} &\min_{\mathbf{w}} \sum_{j=1}^{m}\left[\frac{1}{2}\mathbf{w}_j'\boldsymbol{H}_j\mathbf{w}_j - c_j U\left((\boldsymbol{G}\mathbf{w})_{[j]}\right) + \tau_1\left\|\mathbf{w}_j\right\|_1\right] + \tau_2\sum_{j=1}^{m-1}\left\|\mathbf{w}_{j+1}-\mathbf{w}_j\right\|_1 \\ \text{(MSBP)}\quad &\text{s.t. } \mathbf{w}_1^\top 1_n = \xi_{\text{init}}\,, \\ &\mathbf{w}_j^\top 1_n = \left(1_n + \mathbf{r}_{j-1}\right)^\top \mathbf{w}_{j-1},\ j=2,\ldots,m, \\ &(1_n+\mathbf{r}_m)^\top \mathbf{w}_m = \xi_{\text{term}}, \end{aligned} \tag{3.2}$$

where the objective function integrates the risk control, the utility of return, and the sparse-induced regularization terms for portfolios in each period and turnover across periods. Note that the matrix $\boldsymbol{G} \in \mathbb{R}^{m\times N}$ in the objective function consists of n scenarios of the return vector $\mathbf{r}$, structured as follows:

$$\boldsymbol{G} = \begin{pmatrix} \mathbf{r}_1^\top & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{r}_2^\top & \ddots & \vdots \\ \vdots & \ddots & \ddots & \mathbf{0} \\ \mathbf{0} & \cdots & \mathbf{0} & \mathbf{r}_m^\top \end{pmatrix}.$$

When the utility functions are constant, model (3.2)reduces to the convex optimization model (3.1), which can be efficiently solved by classical convex optimization methods.

Following the methodology of [39], (3.2) can be rewritten in a more compact form as follows:

$$\begin{aligned} &\min_{\mathbf{w}} \sum_{j=1}^{m}\left[\frac{1}{2}\mathbf{w}_j^\top\boldsymbol{H}_j\mathbf{w}_j - c_j U\left((\boldsymbol{G}\mathbf{w})_{[j]}\right) + \tau_1\left\|\mathbf{w}_j\right\|_1\right] + \tau_2\sum_{j=1}^{m-1}\left\|\mathbf{w}_{j+1}-\mathbf{w}_j\right\|_1 \\ &\text{s.t. } \boldsymbol{E}\mathbf{w} = \mathbf{b}, \end{aligned} \tag{3.3}$$

where $(\boldsymbol{G}\mathbf{w})_{[j]}$ represents the j-th smallest component of $\boldsymbol{G}\mathbf{w}$, the matrix $\boldsymbol{E} \in \mathbb{R}^{(m+1)\times N}$ has a natural block representation formed by stacking $(m+1)$ blocks of dimension $1\times n$:

$$\begin{aligned} &\boldsymbol{E}_{i,i} = -1_n^\top, \quad i = 1,2,\ldots,m, \\ &\boldsymbol{E}_{i,i-1} = (1_n + \mathbf{r}_{i-1})^\top, \quad i = 2,3,\ldots,m+1, \end{aligned} \tag{3.4}$$

and $\mathbf{b} = \begin{pmatrix} -\xi_{\text{init}} \\ \mathbf{0} \\ \vdots \\ \mathbf{0} \\ \xi_{\text{term}} \end{pmatrix} \in \mathbb{R}^{m+1}$. The objective function in (3.3) includes two regularization components: the first promotes sparsity in portfolio weights, whereas the second penalizes large deviations in allocated wealth

between consecutive rebalancing dates, thereby implicitly controlling transaction costs. Furthermore, model (3.3) incorporates three groups of linear constraints. The initial constraint enforces the budget balance at the outset, the subsequent constraint $2-m$ upholds self-financing across intermediate rebalancing stages, and the final $(m+1)$-th condition sets the desired terminal wealth.

We define the discrete difference operator $\boldsymbol{F} \in \mathbb{R}^{(N-n)\times N}$ as

$$\boldsymbol{F} = \begin{pmatrix} -\boldsymbol{I} & \boldsymbol{I} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & -\boldsymbol{I} & \boldsymbol{I} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \boldsymbol{I} & \mathbf{0} \\ \mathbf{0} & \cdots & \mathbf{0} & -\boldsymbol{I} & \boldsymbol{I} \end{pmatrix},$$

where $\boldsymbol{I}$ represents the identity matrix, and $\mathbf{0}$ denotes a matrix of zeros. The covariance matrix $\boldsymbol{H} \in \mathbb{R}^{N\times N}$ is given by

$$\boldsymbol{H} = \begin{pmatrix} \boldsymbol{H}_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \boldsymbol{H}_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \mathbf{0} \\ \mathbf{0} & \cdots & \mathbf{0} & \boldsymbol{H}_m \end{pmatrix},$$

where $\boldsymbol{H}_i$ are sub-matrices representing individual periods.

Using these definitions, the multi-period sparse behavioral portfolio (MSBP) model can be formulated as

$$\text{(MSBP)} \quad \begin{aligned} \min_{\mathbf{w}} \quad & \frac{1}{2}\mathbf{w}^\top \boldsymbol{H}\mathbf{w} - \mathbf{c}U(\boldsymbol{G}\mathbf{w}) + \tau_1 \|\mathbf{w}\|_1 + \tau_2 \|\boldsymbol{F}\mathbf{w}\|_1 \\ \text{s.t.} \quad & \boldsymbol{E}\mathbf{w} = \mathbf{b}, \end{aligned} \tag{3.5}$$

where $\mathbf{w}$ denotes the vector of portfolio weights, $\mathbf{c}$ is the vector of rank-dependent weights associated with the utility function, $U(\boldsymbol{G}\mathbf{w})$ represents the utility function based on cumulative prospect theory, τ_1 and τ_2 are regularization parameters that promote sparsity and control turnover, respectively, and $\boldsymbol{E}\mathbf{w} = \mathbf{b}$ represents the constraint on the portfolio. Due to the nonconvexity and nonconcavity of the utility function $U(\boldsymbol{G}\mathbf{w})$ with respect to $\mathbf{w}$, the model cannot be directly solved using traditional convex optimization methods that ensure convergence.

4. Solving method combining the ADMM and PAV for model (3.5)

In this section, we introduce a customized algorithm that combines the alternating direction method of multipliers (ADMM) with the pooled-adjacent-violators (PAV) technique to handle the newly formulated model (3.5). Specifically, the ADMM is utilized to solve problem (3.5), whereas the PAV method is incorporated to efficiently deal with the difficult subproblem involved.

4.1. ADMM for solving model (3.5)

Model (3.5) exhibits nonconvexity and lack of smoothness. Earlier works have therefore depended on standard nonlinear programming solvers or evolutionary algorithms for its resolution. However, these methods often suffer from slow convergence speeds and lack theoretical convergence guarantees. Fortunately, the ADMM is an efficient first-order method to solve separable structural optimization

problems with linear constraints, which has been well-studied in the literature [40–42]. Consequently, we construct a specialized ADMM framework capable of efficiently solving the nonconvex model (3.5) with established convergence properties. For this purpose, auxiliary variables $\mathbf{f} \in \mathbb{R}^{N-n}$, $\mathbf{p} \in \mathbb{R}^N$, and $\mathbf{y} \in \mathbb{R}^m$ are introduced, which permit the reformulation of (3.5) as shown below:

$$\begin{aligned} \min_{\mathbf{f},\mathbf{p},\mathbf{y},\mathbf{w}} & \frac{1}{2}\mathbf{w}^\top \boldsymbol{H}\mathbf{w} - \mathbf{c}U(\mathbf{y}) + \tau_1 \|\mathbf{p}\|_1 + \tau_2 \|\mathbf{f}\|_1 \\ \text{s.t. } & \boldsymbol{E}\mathbf{w} = \mathbf{b}, \\ & \boldsymbol{F}\mathbf{w} = \mathbf{f}, \\ & \mathbf{w} = \mathbf{p}, \\ & \boldsymbol{G}\mathbf{w} = \mathbf{y}. \end{aligned} \tag{4.1}$$

Next, we define the following matrices and vectors:

$$\boldsymbol{A} = \begin{pmatrix} \boldsymbol{E} \\ \boldsymbol{F} \\ \boldsymbol{I} \\ \boldsymbol{G} \end{pmatrix}, \quad \boldsymbol{B} = \begin{pmatrix} \mathbf{0} \\ -\boldsymbol{I} \\ \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \quad \boldsymbol{C} = \begin{pmatrix} \mathbf{0} \\ \mathbf{0} \\ -\boldsymbol{I} \\ \mathbf{0} \end{pmatrix}, \quad \boldsymbol{D} = \begin{pmatrix} \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \\ -\boldsymbol{I} \end{pmatrix}, \quad \mathbf{q} = \begin{pmatrix} \mathbf{b} \\ \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \end{pmatrix}. \tag{4.2}$$

We can then rewrite model (4.1) in the following compact form:

$$\begin{aligned} \min_{\mathbf{f},\mathbf{p},\mathbf{y},\mathbf{w}} & \frac{1}{2}\mathbf{w}^\top \boldsymbol{H}\mathbf{w} - \mathbf{c}U(\mathbf{y}) + \tau_1 \|\mathbf{p}\|_1 + \tau_2 \|\mathbf{f}\|_1 \\ \text{s.t. } & \boldsymbol{A}\mathbf{w} + \boldsymbol{B}\mathbf{f} + \boldsymbol{C}\mathbf{p} + \boldsymbol{D}\mathbf{y} = \mathbf{q}. \end{aligned} \tag{4.3}$$

We form the augmented Lagrangian of (4.3) as

$$\begin{aligned} L_\beta(\mathbf{f},\mathbf{p},\mathbf{y},\mathbf{w},\boldsymbol{\lambda}) = & \frac{1}{2}\mathbf{w}^\top \boldsymbol{H}\mathbf{w} - \mathbf{c}U(\mathbf{y}) + \tau_1 \|\mathbf{p}\|_1 + \tau_2 \|\mathbf{f}\|_1 - \boldsymbol{\lambda}^\top(\boldsymbol{A}\mathbf{w} + \boldsymbol{B}\mathbf{f} + \boldsymbol{C}\mathbf{p} + \boldsymbol{D}\mathbf{y} - \mathbf{q}) \\ & + \frac{\beta}{2}\|\boldsymbol{A}\mathbf{w} + \boldsymbol{B}\mathbf{f} + \boldsymbol{C}\mathbf{p} + \boldsymbol{D}\mathbf{y} - \mathbf{q}\|_2^2, \end{aligned} \tag{4.4}$$

where $\boldsymbol{\lambda} = (\boldsymbol{\lambda}_f, \boldsymbol{\lambda}_p, \boldsymbol{\lambda}_y, \boldsymbol{\lambda}_w)^\top$ denotes the vector of Lagrange multipliers associated with the equality constraints in (4.3), and $\beta > 0$ denotes the penalty parameter. Starting from $(\mathbf{f}^k, \mathbf{p}^k, \mathbf{y}^k, \mathbf{w}^k, \boldsymbol{\lambda}^k)$, the ADMM scheme produces the next iterate $(\mathbf{f}^{k+1}, \mathbf{p}^{k+1}, \mathbf{y}^{k+1}, \mathbf{w}^{k+1}, \boldsymbol{\lambda}^{k+1})$ according to:

$$\mathbf{f}^{k+1} = \arg\min_{\mathbf{f}} \left\{ \tau_2 \|\mathbf{f}\|_1 + \frac{\beta}{2}\left\| \boldsymbol{F}\mathbf{w}^k - \mathbf{f} - \frac{\boldsymbol{\lambda}_f^k}{\beta} \right\|^2 \right\}, \tag{4.5a}$$

$$\mathbf{p}^{k+1} = \arg\min_{\mathbf{p}} \left\{ \tau_1 \|\mathbf{p}\|_1 + \frac{\beta}{2}\left\| \mathbf{w}^k - \mathbf{p} - \frac{\boldsymbol{\lambda}_p^k}{\beta} \right\|^2 \right\}, \tag{4.5b}$$

$$\mathbf{y}^{k+1} = \arg\min_{\mathbf{y}} \left\{ -\mathbf{c}U(\mathbf{y}) + \frac{\beta}{2}\left\| \boldsymbol{G}\mathbf{w}^k - \mathbf{y} - \frac{\boldsymbol{\lambda}_y^k}{\beta} \right\|^2 \right\}, \tag{4.5c}$$

$$\mathbf{w}^{k+1} = \arg\min_{\mathbf{w}} \left\{ Q\left(\mathbf{w}; \mathbf{f}^{k+1}, \mathbf{p}^{k+1}, \mathbf{y}^{k+1}\right) \right\}, \tag{4.5d}$$

$$\boldsymbol{\lambda}^{k+1} = \boldsymbol{\lambda}^{k} - \beta\left(\boldsymbol{A}\mathbf{w}^{k+1} + \boldsymbol{B}\mathbf{f}^{k+1} + \boldsymbol{C}\mathbf{p}^{k+1} + \boldsymbol{D}\mathbf{y}^{k+1} - \mathbf{q}\right), \tag{4.5e}$$

where $Q(\cdot;\cdot,\cdot)$ is the following quadratic majorization:

$$\begin{aligned} Q\left(\mathbf{w};\mathbf{f}^{k+1},\mathbf{p}^{k+1},\mathbf{y}^{k+1}\right) = \frac{1}{2}\mathbf{w}^{\top}\boldsymbol{H}\mathbf{w} + \frac{\beta}{2}\left\|\boldsymbol{E}\mathbf{w} - \mathbf{b} - \frac{\boldsymbol{\lambda}_w^k}{\beta}\right\|_2^2 + \frac{\beta}{2}\left\|\boldsymbol{F}\mathbf{w} - \mathbf{f}^{k+1} - \frac{\boldsymbol{\lambda}_f^k}{\beta}\right\|_2^2 \\ + \frac{\beta}{2}\left\|\mathbf{w} - \mathbf{p}^{k+1} - \frac{\boldsymbol{\lambda}_p^k}{\beta}\right\|_2^2 + \frac{\beta}{2}\left\|\boldsymbol{G}\mathbf{w} - \mathbf{y}^{k+1} - \frac{\boldsymbol{\lambda}_y^k}{\beta}\right\|_2^2. \end{aligned} \tag{4.6}$$

Note that $\boldsymbol{\lambda}_f$, $\boldsymbol{\lambda}_p$, $\boldsymbol{\lambda}_y$, and $\boldsymbol{\lambda}_w$ are the components of $\boldsymbol{\lambda}$ corresponding to the Lagrangian multipliers associated with the constraints $\boldsymbol{F}\mathbf{w} = \mathbf{f}$, $\mathbf{w} = \mathbf{p}$, $\boldsymbol{G}\mathbf{w} = \mathbf{y}$, and $\boldsymbol{E}\mathbf{w} = \mathbf{b}$ in (3.5). The iterative update of $\boldsymbol{\lambda}$ in (4.5e) can be decomposed to

$$\begin{aligned} \boldsymbol{\lambda}_f^{k+1} &= \boldsymbol{\lambda}_f^{k} - \beta\left(\boldsymbol{F}\mathbf{w}^{k+1} - \mathbf{f}^{k+1}\right), \\ \boldsymbol{\lambda}_p^{k+1} &= \boldsymbol{\lambda}_p^{k} - \beta\left(\mathbf{w}^{k+1} - \mathbf{p}^{k+1}\right), \\ \boldsymbol{\lambda}_y^{k+1} &= \boldsymbol{\lambda}_y^{k} - \beta\left(\boldsymbol{G}\mathbf{w}^{k+1} - \mathbf{y}^{k+1}\right), \\ \boldsymbol{\lambda}_w^{k+1} &= \boldsymbol{\lambda}_w^{k} - \beta\left(\boldsymbol{E}\mathbf{w}^{k+1} - \mathbf{b}\right). \end{aligned} \tag{4.7}$$

Next, we describe how to efficiently solve the subproblems (4.5a)–(4.5d). Specifically, the **f**-subproblem (4.5a) and **p**-subproblem (4.5b) are given by

$$\mathbf{f}^{k+1} = \arg\min_{\mathbf{f}}\left\{\tau_2\|\mathbf{f}\|_1 + \frac{\beta}{2}\left\|\boldsymbol{F}\mathbf{w}^{k} - \mathbf{f} - \frac{\boldsymbol{\lambda}_f^k}{\beta}\right\|^2\right\}$$

and

$$\mathbf{p}^{k+1} = \arg\min_{\mathbf{p}}\left\{\tau_1\|\mathbf{p}\|_1 + \frac{\beta}{2}\left\|\mathbf{w}^{k} - \mathbf{p} - \frac{\boldsymbol{\lambda}_p^k}{\beta}\right\|^2\right\}.$$

Both the **f**-subproblem and **p**-subproblem admit closed-form solutions via the well-known soft-thresholding operator. In particular, the soft-thresholding operator $S(\mathbf{u},\tau)$ is obtained as the minimizer of the proximal problem

$$\min_{\mathbf{x}}\left\{\|\mathbf{x}\|_1 + \frac{1}{2\tau}\|\mathbf{x}-\mathbf{u}\|^2\right\},$$

where $\mathbf{u}$ is a given vector and τ is a prescribed parameter. Its closed-form expression is

$$(S(\mathbf{u},\tau))_i = \frac{u_i}{|u_i|}\max\{\,|u_i| - \tau,\ 0\,\}.$$

The **y**-subproblem admits a closed-form solution and reduces to

$$\mathbf{y}^{k+1} = \arg\min_{\mathbf{y}}\left\{-\mathbf{c}U(\mathbf{y}) + \frac{\beta}{2}\left\|\boldsymbol{G}\mathbf{w}^{k} - \mathbf{y} - \frac{\boldsymbol{\lambda}_y^k}{\beta}\right\|^2\right\},$$

which can be solved using the pooling-adjacent-violators (PAV) algorithm, with detailed methodology to be provided later.

The **w**-subproblem involves solving the following linear system:

$$(\boldsymbol{H} + \beta \boldsymbol{A}^\top \boldsymbol{A})\,\mathbf{w} = \boldsymbol{A}^\top \boldsymbol{\lambda}^k - \beta \boldsymbol{A}^\top \left(\boldsymbol{B}\mathbf{f}^{k+1} + \boldsymbol{C}\mathbf{p}^{k+1} + \boldsymbol{D}\mathbf{y}^{k+1} - \boldsymbol{q}\right). \tag{4.8}$$

Since the coefficient matrix remains positive definite, each linear system is solved inexpensively via Cholesky factorization following [39]. A complete description of the ADMM procedure for model (4.3) is given in Algorithm 1.

Algorithm 1 ADMM for CPT-based sparse behavioral portfolio problem (3.5)

Input: $\beta > \hat{\beta} := \max\left\{\frac{-\rho'_{\min} + \sqrt{\rho'^2_{\min} + 8\rho'_{\max}}}{2\rho_{\min}}, \frac{\rho'^2_{\max}}{\rho'_{\min}\rho_{\min}}\right\}$, where $\rho_{\min}$ denotes the smallest eigenvalue of $\boldsymbol{A}^\top \boldsymbol{A}$, and $\rho'_{\min}$ and $\rho'_{\max}$ denote the smallest and largest eigenvalues of $\boldsymbol{H}$. We are given $\mathbf{f}_0$, $\mathbf{p}_0$, $\mathbf{y}_0$, $\mathbf{w}_0$, τ_1, τ_2, $\boldsymbol{\lambda}_0$, and set $k = 0$.

1: **repeat**
2: Solve (4.5a) to obtain $\mathbf{f}^{k+1}$;
3: Solve (4.5b) to obtain $\mathbf{p}^{k+1}$;
4: Solve (4.5c) to obtain $\mathbf{y}^{k+1}$;
5: Solve (4.5d) to obtain $\mathbf{w}^{k+1}$;
6: Update $\boldsymbol{\lambda}^{k+1} = \boldsymbol{\lambda}^k - \beta\left(\boldsymbol{A}\mathbf{w}^{k+1} + \boldsymbol{B}\mathbf{f}^{k+1} + \boldsymbol{C}\mathbf{p}^{k+1} + \boldsymbol{D}\mathbf{y}^{k+1} - \mathbf{q}\right)$.
7: $k = k + 1$
8: **until** the stopping criterion is satisfied.

4.2. Solving **y***-subproblem with the ADMM*

As discussed earlier, the **f**-subproblem, **p**-subproblem, and **w**-subproblem each yield closed-form updates within the ADMM framework. We now turn to the **y**-subproblem, which incorporates the CPT utility function. Without loss of generality, the minimization task in (4.5c) can be rewritten as:

$$\Gamma(\mathbf{y}) = -\mathbf{c}U(\mathbf{y}) + \frac{\beta}{2}\left\|\boldsymbol{G}\mathbf{w}^k - \mathbf{y} - \frac{\boldsymbol{\lambda}_y^k}{\beta}\right\|^2.$$

This subsection introduces the pooling-adjacent-violators (PAV) algorithm as an efficient tool for solving the **y**-subproblem. Renowned for its ability to address convex separable problems subject to order constraints, PAV traces back to [43], with a formal and thorough analysis provided by [44] in the literature on monotone estimation. The method's wide applicability has been demonstrated in various contexts, including the works of [45–47]. Building upon this foundation, [48] introduced a novel algorithm based on the PAV framework to address the CPT utility maximization problem under monotonicity constraints. Despite the nonconvexity of equation (4.5c) and its reformulation, the **y**-subproblem can be efficiently solved by leveraging the structured form of $f_i(y)$.

For simplicity, we reformulate the **y**-subproblem by introducing $\mathbf{s}^{k+1} = \boldsymbol{G}\mathbf{w}^{k+1} - \frac{\boldsymbol{\lambda}^k}{\beta}$. Since the term $c_i U(y_{[i]})$ is invariant to permutations of **y**, the **y**-subproblem (4.5c) can be reformulated—after suitable reordering—as the following isotonic regression problem:

$$\min \sum_{i=1}^{m} f_i(y_i) \quad \text{s.t. } y_1 \le y_2 \le \cdots \le y_m, \tag{4.9}$$

where $f_i(y_i) = -c_i U(y_i) + \frac{\beta}{2}(y_i - s_i)^2$. Additionally, the constraints in Equation (4.9) comply with the restrictions defined by the simple chain constraint framework [49].

To efficiently address the $\mathbf{y}$-subproblem, we employ the modified PAV procedure introduced in [48]. Let $J = \{1, 2, \ldots, M\}$ and begin with the finest possible partition consisting of M singletons $[1, 1], [2, 2], \ldots, [M, M]$. Each block $[p, q]$ subsequently represents the consecutive indices from p to q, inclusive. A block $[p, q]$ is called single-valued if all y_i within this block take an identical value in the optimal solution to the following problem:

$$\min \sum_{i=p}^{q} f_i(y_i) \quad \text{s.t. } y_p \le y_{p+1} \le \cdots \le y_q,$$

where specifically $y_p^* = y_{p+1}^* = \cdots = y_q^*$, denoted as $V_{[p,q]}$. If two adjacent blocks $[p, q]$ and $[q + 1, r]$ violate the monotonicity constraint, i.e., $V_{[p,q]} > V_{[q+1,r]}$, they are combined into a larger block $[p, r]$. When adjacent blocks preserve non-decreasing values, the optimal solution assigns $y_i = V_{[p,q]}$ for all $p \le i \le q$, where $[p, q] \in J$. The PAV algorithm ends once adjacent blocks exhibit non-decreasing values. Notably, Cui [48] found that binary search can still identify the minimizer of a nonconvex f_i by utilizing its monotonicity, even though classical PAV methods assume convexity. According to Fact 1 of [48], the function $\sum_{i=p}^{q} f_i(y_i)$ can have at most two local minima, so the global minimizer is determined by evaluating and comparing these two points.

Lemma 1 (Finite subgradient)**.** *Let* $\mathbf{y}^*$ *represent an optimal solution to the problem in* (4.9)*. Then* $U'(y_i^*) < \infty$ *for all* $i = 1, \ldots, m$*.*

Lemma 1 guarantees that, at optimality, the value function $U(\cdot)$ is differentiable (in the usual sense) at every component of $\mathbf{y}^*$. Combined with Lemma 2, this ensures that the Clarke generalized subgradient of the objective is a singleton at $\mathbf{y}^*$, a property that proves essential for both the correctness of the proposed algorithm and the subsequent convergence analysis.

Lemma 2. *Assume that* $\mathbf{y}^*$ *is an optimal solution of problem* (4.9)*. Consequently, we obtain*

$$s_1 \le y_1^* \le \cdots \le y_m^* \le \max_{i=1,\ldots,m} \left\{ \max \left(B + 1, s_i + \frac{b_i U'(B + 1)}{\beta} \right) \right\}.$$

Lemma 2 additionally establishes that any optimal solution to problem (4.9) lies within a given set of predefined constants, where

$$l_b = s_1 \quad \text{and} \quad u_b = \max_{i=1,\ldots,m} \left\{ \max \left(B + 1, s_i + \frac{b_i U'(B + 1)}{\beta} \right) \right\},$$

which provide the starting lower and upper bounds for the binary search incorporated into the PAV solver of the $\mathbf{y}$-subproblem. In line with [48], the search is confined to the fixed interval $[l_b, u_b]$ and iteratively bisected according to the sign of the derivative of the pooled objective, thereby ensuring finite termination at a local minimizer and greatly facilitating the subsequent convergence analysis.

Remark 1. *While the modified PAV algorithm does not guarantee convergence to the global optimum, it does yield a stationary point of the associated optimization problem:*

$$\min_{\mathbf{y}} \Gamma(\mathbf{y}) := \sum_{i=1}^{m} -c_i U\left(y_{[i]}\right) + \frac{\sigma}{2}\left(y_{[i]} - s_i\right)^2, \tag{4.10}$$

where (4.10) *corresponds to* (4.9). *Specifically, the output* $\bar{\mathbf{y}}$ *produced by the PAV algorithm satisfies the stationarity condition* $0 \in \partial\Gamma(\bar{\mathbf{y}})$. *Detailed derivations and technical proofs of the results stated above are provided in [48]. The concrete PAV method for solving the* $\mathbf{y}$*-subproblem can be found in the Algorithm 2 of [48].*

5. Convergence analysis

Due to the nonconvexity of $\mathbf{c}U(\cdot)$, problem (4.3) remains nonconvex, and the ADMM framework presented in Algorithm 1 does not fit within existing methodologies. Consequently, our focus now shifts to analyzing the theoretical convergence of the ADMM to ensure its efficiency.

5.1. Preliminaries

For a proper and closed function $g : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$, a vector $\mathbf{u} \in \partial g(\mathbf{x})$ is called a subgradient of g at $\mathbf{x} \in \operatorname{dom} g$, where the subdifferential ∂g is defined as

$$\partial g(\mathbf{x}) := \left\{\mathbf{u} \in \mathbb{R}^n \mid \exists \mathbf{x}^k \to \mathbf{x}, \widehat{\partial} g(\mathbf{x}^k) \ni \mathbf{u}^k \to \mathbf{u}, g(\mathbf{x}^k) \to g(\mathbf{x})\right\}, \tag{5.1}$$

where $\widehat{\partial} g(\mathbf{x})$ denotes the set of regular subgradients of g at $\mathbf{x}$:

$$\widehat{\partial} g(\mathbf{x}) := \{\mathbf{u} \in \mathbb{R}^n \mid g(\mathbf{y}) \geq g(\mathbf{x}) + \langle \mathbf{u}, \mathbf{y} - \mathbf{x}\rangle + o(\|\mathbf{y} - \mathbf{x}\|), \forall \mathbf{y} \in \mathbb{R}^n\}.$$

If $f : \mathbb{R}^n \to \mathbb{R}$ is C^1 and $g : \mathbb{R}^n \to \overline{\mathbb{R}} := \mathbb{R} \cup \{+\infty\}$ is proper and lower semicontinuous, then [50] implied the sum rule $\partial(f + g)(\mathbf{x}) = \nabla f(\mathbf{x}) + \partial g(\mathbf{x})$ for all $\mathbf{x} \in \mathbb{R}^n$. A point $\mathbf{x}^*$ is termed a *limiting critical point* of a function $F : \mathbb{R}^n \to \overline{\mathbb{R}}$ if $0 \in \partial F(\mathbf{x}^*)$. The set of all such points is denoted by $\operatorname{crit}(F)$. Define the auxiliary function $\Omega(\mathbf{y}) = -\sum_{i=1}^m c_i U(y_{[i]})$, which constitutes the smooth component of the overall objective $\Gamma(\mathbf{y})$. Under the mild assumptions stated in Lemma 1 of [48], both $\Omega(\mathbf{y})$ and $\Gamma(\mathbf{y})$ are locally Lipschitz continuous. Consequently, critical points of these functions can be characterized via the Clarke subdifferential. When the derivative $U'(y_i)$ remains finite for all i, the Clarke generalized gradient of Ω at $\mathbf{y}$, denoted by $\partial\Omega(\mathbf{y})$, is well-defined and nonempty [51].

Lemma 3 (Outer semicontinuity of the Clarke subdifferential)**.** *Suppose* Ω *is locally Lipschitz in a neighborhood of* $\mathbf{y}^*$. *Let* $\{\mathbf{y}^k\}_{k\in\mathbb{N}}$ *and* $\{\mathbf{v}^k\}_{k\in\mathbb{N}}$ *be sequences satisfying*

$$\mathbf{y}^k \to \mathbf{y}^*, \quad \mathbf{v}^k \to \mathbf{v}^*, \quad \textit{and} \quad \mathbf{v}^k \in \partial\Omega(\mathbf{y}^k) \quad \forall k.$$

Then $\mathbf{v}^* \in \partial\Omega(\mathbf{y}^*)$.

Definition 1. *Denote* $\Phi := \|\cdot\|_1$. *We say that* $(\mathbf{f}^*, \mathbf{p}^*, \mathbf{y}^*, \mathbf{w}^*, \boldsymbol{\lambda}^*)$ *is a critical point of the augmented Lagrangian function* $L_\beta(\cdot)$ *in 4.4 if it satisfies*

$$\begin{cases} 0 \in \tau_2 \partial\Phi(\mathbf{f}^*) - \boldsymbol{B}^\top \boldsymbol{\lambda}^*, \\ 0 \in \tau_1 \partial\Phi(\mathbf{p}^*) - \boldsymbol{C}^\top \boldsymbol{\lambda}^*, \\ 0 \in \partial_{\mathbf{y}}\Omega(\mathbf{y}^*) - \boldsymbol{D}^\top \boldsymbol{\lambda}^*, \\ 0 = \frac{1}{2}\boldsymbol{H}\mathbf{w}^* - \boldsymbol{A}^\top \boldsymbol{\lambda}^*, \\ 0 = \boldsymbol{A}\mathbf{w}^* + \boldsymbol{B}\mathbf{f}^* + \boldsymbol{C}\mathbf{p}^* + \boldsymbol{D}\mathbf{y}^* - \mathbf{q}. \end{cases} \tag{5.2}$$

A stationary point of the augmented Lagrangian associated with (4.3) is equivalent to a Karush–Kuhn–Tucker (KKT) point of the original problem. Throughout this work, we assume that (4.3) admits at least one such KKT point.

5.2. Convergence

In this subsection, we investigate the convergence properties of Algorithm 1. With $\Phi := |\cdot|_1$ and the update rules in (4.5a)–(4.5d) in mind, we begin by outlining the first-order optimality conditions associated with each subproblem in Algorithm 1:

$$\begin{cases} 0 \in \tau_2 \partial\Phi\left(\mathbf{f}^{k+1}\right) - \boldsymbol{B}^\top \boldsymbol{\lambda}^k + \beta \boldsymbol{B}^\top \left(\boldsymbol{A}\mathbf{w}^k + \boldsymbol{B}\mathbf{f}^{k+1} + \boldsymbol{C}\mathbf{p}^k + \boldsymbol{D}\mathbf{y}^k - \mathbf{q}\right), \\ 0 \in \tau_1 \partial\Phi\left(\mathbf{p}^{k+1}\right) - \boldsymbol{C}^\top \boldsymbol{\lambda}^k + \beta \boldsymbol{C}^\top \left(\boldsymbol{A}\mathbf{w}^k + \boldsymbol{B}\mathbf{f}^{k+1} + \boldsymbol{C}\mathbf{p}^{k+1} + \boldsymbol{D}\mathbf{y}^k - \mathbf{q}\right), \\ 0 \in \partial_{\mathbf{y}}\Omega\left(\mathbf{y}^{k+1}\right) - \boldsymbol{D}^\top \boldsymbol{\lambda}^k + \beta \boldsymbol{D}^\top \left(\boldsymbol{A}\mathbf{w}^k + \boldsymbol{B}\mathbf{f}^{k+1} + \boldsymbol{C}\mathbf{p}^{k+1} + \boldsymbol{D}\mathbf{y}^{k+1} - \mathbf{q}\right), \\ 0 = \frac{1}{2}\boldsymbol{H}\mathbf{w}^{k+1} - \boldsymbol{A}^\top \boldsymbol{\lambda}^k + \beta \boldsymbol{A}^\top \left(\boldsymbol{A}\mathbf{w}^{k+1} + \boldsymbol{B}\mathbf{f}^{k+1} + \boldsymbol{C}\mathbf{p}^{k+1} + \boldsymbol{D}\mathbf{y}^{k+1} - \mathbf{q}\right), \\ \boldsymbol{\lambda}^{k+1} = \boldsymbol{\lambda}^k - \beta\left(\boldsymbol{A}\mathbf{w}^{k+1} + \boldsymbol{B}\mathbf{f}^{k+1} + \boldsymbol{C}\mathbf{p}^{k+1} + \boldsymbol{D}\mathbf{y}^{k+1} - \mathbf{q}\right), \end{cases} \tag{5.3}$$

which is equivalent to

$$\begin{cases} 0 \in \tau_2 \partial\Phi\left(\mathbf{f}^{k+1}\right) - \boldsymbol{\lambda}_f^k - \beta\left(\boldsymbol{F}\mathbf{w}^k - \mathbf{f}^{k+1}\right), \\ 0 \in \tau_1 \partial\Phi\left(\mathbf{p}^{k+1}\right) - \boldsymbol{\lambda}_p^k - \beta\left(\mathbf{w}^k - \mathbf{p}^{k+1}\right), \\ 0 \in \partial_{\mathbf{y}}\Omega\left(\mathbf{y}^{k+1}\right) - \boldsymbol{\lambda}_y^k - \beta\left(\boldsymbol{G}\mathbf{w}^k - \mathbf{y}^{k+1}\right), \\ 0 = \frac{1}{2}H\mathbf{w}^{k+1} - \boldsymbol{A}^\top \boldsymbol{\lambda}^k + \beta \boldsymbol{A}^\top \left(\boldsymbol{A}\mathbf{w}^{k+1} + \boldsymbol{B}\mathbf{f}^{k+1} + \boldsymbol{C}\mathbf{p}^{k+1} + \boldsymbol{D}\mathbf{y}^{k+1} - \mathbf{q}\right), \\ \boldsymbol{\lambda}^{k+1} = \boldsymbol{\lambda}^k - \beta\left(\boldsymbol{A}\mathbf{w}^{k+1} + \boldsymbol{B}\mathbf{f}^{k+1} + \boldsymbol{C}\mathbf{p}^{k+1} + \boldsymbol{D}\mathbf{y}^{k+1} - \mathbf{q}\right). \end{cases} \tag{5.4}$$

Assumption 5.1. *The* **f***-subproblem,* **p***-subproblem, and* **w***-subproblem admit global solutions. The* **y***-subproblem is handled according to the following conditions:*

$$\begin{aligned} &L_\beta\left(\mathbf{f}^{k+1}, \mathbf{p}^{k+1}, \mathbf{y}^k, \mathbf{w}^k, \boldsymbol{\lambda}^k\right) - L_\beta\left(\mathbf{f}^{k+1}, \mathbf{p}^{k+1}, \mathbf{y}^{k+1}, \mathbf{w}^k, \boldsymbol{\lambda}^k\right) \geq 0, \quad \textit{and} \\ &\mathbf{0} \in \partial\Omega\left(\mathbf{y}^{k+1}\right) + \beta\left(\mathbf{y}^{k+1} - \boldsymbol{G}\mathbf{w}^k + \frac{\boldsymbol{\lambda}^k}{\beta}\right). \end{aligned} \tag{5.5}$$

Assumption 5.1 governs the solutions of the **f**-subproblem, **p**-subproblem, **w**-subproblem, and **y**-subproblem. Unlike the others, the **y**-subproblem can only be driven to a stationary point $\mathbf{y}^{k+1}$ whose objective value is no larger than that at the incoming iterate $\mathbf{y}^k$. To handle this subproblem efficiently, we employ the pooled-adjacent-violators (PAV) algorithm, with a complete description provided in Section 4.2. According to Remark 1, we can see that the second condition in (5.5) can be satisfied theoretically, whereas the first condition is more easily satisfied via a termination criterion in implementation. Furthermore, when the **y**-subproblem attains an exact solution, (5.5) naturally holds.

Lemma 4. *Let* $\{(\mathbf{f}^k, \mathbf{p}^k, \mathbf{y}^k, \mathbf{w}^k, \boldsymbol{\lambda}^k)\}$ *be the sequence generated by Algorithm 1. Then, for any* $k > 0$*, we have*

$$\left\|\boldsymbol{\lambda}^{k+1} - \boldsymbol{\lambda}^k\right\|^2 \leq \frac{1}{\rho_{\min}} \left\|A^\top(\boldsymbol{\lambda}^{k+1} - \boldsymbol{\lambda}^k)\right\|^2,$$

where $\boldsymbol{\rho}_{\min}$ *is the smallest eigenvalue of* $\boldsymbol{A}^\top \boldsymbol{A}$*.*

Proof. The proof is similar to that of Lemma 1 in [52]. □

Lemma 5. *Let* $\{(\mathbf{f}^k, \mathbf{p}^k, \mathbf{y}^k, \mathbf{w}^k, \boldsymbol{\lambda}^k)\}$ *be the sequence generated by Algorithm 1, and then the sequence* $\{L_\beta(\mathbf{f}^k, \mathbf{p}^k, \mathbf{y}^k, \mathbf{w}^k, \boldsymbol{\lambda}^k)\}$ *is decreasing, i.e.,*

$$L_\beta\left(\mathbf{f}^{k+1}, \mathbf{p}^{k+1}, \mathbf{y}^{k+1}, \mathbf{w}^{k+1}, \boldsymbol{\lambda}^{k+1}\right) - L_\beta\left(\mathbf{f}^k, \mathbf{p}^k, \mathbf{y}^k, \mathbf{w}^k, \boldsymbol{\lambda}^k\right) \leq -b\left\|\mathbf{w}^{k+1} - \mathbf{w}^k\right\|^2, \tag{5.6}$$

where $b > 0$ *is a certain positive constant.*

Proof. From the definition of $L_\beta(\cdot)$ in (4.4) and the $\boldsymbol{\lambda}$-updating rule in (4.5e), we have

$$\begin{aligned} &L_\beta\left(\mathbf{f}^{k+1}, \mathbf{p}^{k+1}, \mathbf{y}^{k+1}, \mathbf{w}^{k+1}, \boldsymbol{\lambda}^{k+1}\right) - L_\beta\left(\mathbf{f}^{k+1}, \mathbf{p}^{k+1}, \mathbf{y}^{k+1}, \mathbf{w}^{k+1}, \boldsymbol{\lambda}^{k}\right) \\ &= \langle \boldsymbol{\lambda}^{k+1} - \boldsymbol{\lambda}^k, \boldsymbol{A}\mathbf{w}^{k+1} + \boldsymbol{B}\mathbf{f}^{k+1} + \boldsymbol{C}\mathbf{p}^{k+1} + \boldsymbol{D}\mathbf{y}^{k+1} - \mathbf{q}\rangle \\ &= \frac{1}{\beta}\|\boldsymbol{\lambda}^{k+1} - \boldsymbol{\lambda}^k\|^2. \end{aligned} \tag{5.7}$$

The first-order condition associated with the $\mathbf{w}$-subproblem (4.5d) is given by

$$0 = \boldsymbol{H}\mathbf{w}^{k+1} - \boldsymbol{A}^\top\boldsymbol{\lambda}^k - \boldsymbol{A}^\top(\boldsymbol{\lambda}^{k+1} - \boldsymbol{\lambda}^k). \tag{5.8}$$

Hence, we have $\boldsymbol{A}^\top\boldsymbol{\lambda}^{k+1} = \boldsymbol{H}\mathbf{w}^{k+1}$. Similarly, $\boldsymbol{A}^\top\boldsymbol{\lambda}^k = \boldsymbol{H}\mathbf{w}^k$. Because $\boldsymbol{H}$ is positive definite, it follows that

$$\|\boldsymbol{A}^\top(\boldsymbol{\lambda}^{k+1} - \boldsymbol{\lambda}^k)\|^2 = \|\boldsymbol{H}^\top(\mathbf{w}^{k+1} - \mathbf{w}^k)\|^2 \le \rho'^2_{\max}\|\mathbf{w}^{k+1} - \mathbf{w}^k\|^2,$$

where $\rho'_{\max}$ denotes the largest eigenvalue of H. Combining this with Lemma 4, we get

$$\frac{1}{\beta}\|\boldsymbol{\lambda}^{k+1} - \boldsymbol{\lambda}^k\|^2 \le \frac{\rho'^2_{\max}}{\beta\rho_{\min}}\|\mathbf{w}^{k+1} - \mathbf{w}^k\|^2. \tag{5.9}$$

Next, the $\mathbf{w}$-subproblem (4.5d) in Algorithm 1 simplifies to

$$(\boldsymbol{H} + \beta\boldsymbol{A}^\top\boldsymbol{A})\mathbf{w} = \boldsymbol{A}^\top\boldsymbol{\lambda}^k - \beta\boldsymbol{A}^\top(\boldsymbol{B}\mathbf{f}^{k+1} + \boldsymbol{C}\mathbf{p}^{k+1} + \boldsymbol{D}\mathbf{y}^{k+1} - \mathbf{q}).$$

Since $\boldsymbol{H}$ is positive definite and $\boldsymbol{A}$ has full rank, the perturbed Hessian $\boldsymbol{H} + \beta\boldsymbol{A}^\top A$ is clearly invertible for any $\beta > 0$. Consequently, the subproblem in $\boldsymbol{w}$ is strongly convex, with a strong-convexityy modulus at least $\rho'_{\min} + \beta\rho_{\min}$. Therefore, we have

$$\begin{aligned} &L_\beta\left(\mathbf{f}^{k+1}, \mathbf{p}^{k+1}, \mathbf{y}^{k+1}, \mathbf{w}^{k+1}, \boldsymbol{\lambda}^{k}\right) - L_\beta\left(\mathbf{f}^{k+1}, \mathbf{p}^{k+1}, \mathbf{y}^{k+1}, \mathbf{w}^{k}, \boldsymbol{\lambda}^{k}\right) \\ &= -\frac{\rho'_{\min} + \beta\rho_{\min}}{2}\|\mathbf{w}^{k+1} - \mathbf{w}^k\|^2. \end{aligned} \tag{5.10}$$

Furthermore, as $\mathbf{f}^{k+1}$ is the optimal solution to the $\mathbf{f}$-subproblem (4.5a) and $\mathbf{p}^{k+1}$ solves the $\mathbf{p}$-subproblem (4.5b) in Algorithm 1, we have

$$L_\beta\left(\mathbf{f}^{k+1}, \mathbf{p}^{k}, \mathbf{y}^{k}, \mathbf{w}^{k}, \boldsymbol{\lambda}^{k}\right) - L_\beta\left(\mathbf{f}^{k}, \mathbf{p}^{k}, \mathbf{y}^{k}, \mathbf{w}^{k}, \boldsymbol{\lambda}^{k}\right) \le 0, \tag{5.11}$$

and

$$L_\beta\left(\mathbf{f}^{k+1}, \mathbf{p}^{k+1}, \mathbf{y}^{k}, \mathbf{w}^{k}, \boldsymbol{\lambda}^{k}\right) - L_\beta\left(\mathbf{f}^{k+1}, \mathbf{p}^{k}, \mathbf{y}^{k}, \mathbf{w}^{k}, \boldsymbol{\lambda}^{k}\right) \le 0. \tag{5.12}$$

According to Assumption 5.1, we know

$$L_\beta\left(\mathbf{f}^{k+1}, \mathbf{p}^{k+1}, \mathbf{y}^{k+1}, \mathbf{w}^{k}, \boldsymbol{\lambda}^{k}\right) - L_\beta\left(\mathbf{f}^{k+1}, \mathbf{p}^{k+1}, \mathbf{y}^{k}, \mathbf{w}^{k}, \boldsymbol{\lambda}^{k}\right) \le 0. \tag{5.13}$$

Summing (5.7), (5.10), (5.11), (5.12), and (5.13), we get

$$\begin{aligned} &L_\beta\left(\mathbf{f}^{k+1}, \mathbf{p}^{k+1}, \mathbf{y}^{k+1}, \mathbf{w}^{k+1}, \boldsymbol{\lambda}^{k+1}\right) - L_\beta\left(\mathbf{f}^{k}, \mathbf{p}^{k}, \mathbf{y}^{k}, \mathbf{w}^{k}, \boldsymbol{\lambda}^{k}\right) \\ &\le \frac{\rho'^2_{\max}}{\beta\rho_{\min}}\|\mathbf{w}^{k+1} - \mathbf{w}^k\|^2 - \frac{\rho'_{\min} + \beta\rho_{\min}}{2}\|\mathbf{w}^{k+1} - \mathbf{w}^k\|^2 \\ &= -b\|\mathbf{w}^{k+1} - \mathbf{w}^k\|^2, \end{aligned} \tag{5.14}$$

where $b = \frac{\rho'_{\min}+\beta\rho_{\min}}{2} - \frac{\rho'^2_{\max}}{\beta\rho_{\min}}$. We establish that $b > 0$ if $\beta > \frac{-\rho'_{\min}+\sqrt{\rho'^2_{\min}+8\rho'^2_{\max}}}{2\rho_{\min}}$ as specified in Algorithm 1. Consequently, it follows that the sequence $\{L_\beta(\mathbf{f}^k, \mathbf{p}^k, \mathbf{y}^k, \mathbf{w}^k, \boldsymbol{\lambda}^k)\}$ is decreasing. This completes the proof. □

Lemma 6. *The sequence* $\{(\mathbf{f}^k, \mathbf{p}^k, \mathbf{y}^k, \mathbf{w}^k, \boldsymbol{\lambda}^k)\}$ *generated by Algorithm 1 is bounded.*

Proof. From Lemma 5, for $k \geq 1$, we have

$$
\begin{aligned}
&L_\beta\left(\mathbf{f}^1, \mathbf{p}^1, \mathbf{y}^1, \mathbf{w}^1, \boldsymbol{\lambda}^1\right) \\
&\geq L_\beta\left(\mathbf{f}^k, \mathbf{p}^k, \mathbf{y}^k, \mathbf{w}^k, \boldsymbol{\lambda}^k\right) \\
&= \frac{1}{2}\mathbf{w}^k H \mathbf{w}^k + \tau_1\Phi(\mathbf{f}^k) + \tau_2\Phi(\mathbf{p}^k) + \Omega(\mathbf{y}^k) - \lambda^\top(\boldsymbol{A}\mathbf{w}^k + \boldsymbol{B}\mathbf{f}^k + \boldsymbol{C}\mathbf{p}^k + \boldsymbol{D}\mathbf{y}^k - \mathbf{q}) \\
&\quad + \frac{\beta}{2}\left\|\boldsymbol{A}\mathbf{w}^k + \boldsymbol{B}\mathbf{f}^k + \boldsymbol{C}\mathbf{p}^k + \boldsymbol{D}\mathbf{y}^k - \mathbf{q}\right\|_2^2 \\
&= \frac{1}{2}\mathbf{w}^k \boldsymbol{H} \mathbf{w}^k + \tau_1\Phi(\mathbf{f}^k) + \tau_2\Phi(\mathbf{p}^k) + \Omega(\mathbf{y}^k) + \frac{\beta}{2}\left\|\boldsymbol{A}\mathbf{w}^k + \boldsymbol{B}\mathbf{f}^k + \boldsymbol{C}\mathbf{p}^k + \boldsymbol{D}\mathbf{y}^k - \mathbf{q} - \frac{\boldsymbol{\lambda}^k}{\beta}\right\|_2^2 \qquad (5.15) \\
&\quad - \frac{1}{2\beta}\|\boldsymbol{\lambda}^k\|^2 \\
&\geq \frac{1}{2}\left(\rho'_{\min} - \frac{\rho'^2_{\max}}{\beta\rho_{\min}}\right)\|\mathbf{w}^k\|^2 + \tau_1\Phi(\mathbf{f}^k) + \tau_2\Phi(\mathbf{p}^k) + \Omega(\mathbf{y}^k) \\
&\quad + \frac{\beta}{2}\left\|\boldsymbol{A}\mathbf{w}^k + \boldsymbol{B}\mathbf{f}^k + \boldsymbol{C}\mathbf{p}^k + \boldsymbol{D}\mathbf{y}^k - \mathbf{q} - \frac{\boldsymbol{\lambda}^k}{\beta}\right\|_2^2,
\end{aligned}
$$

where the last inequality can be derived using the fact that $\boldsymbol{A}^\top\boldsymbol{\lambda}^k = \boldsymbol{H}\mathbf{w}^k$ from (5.8). Specifically, it follows from $\boldsymbol{A}^\top\boldsymbol{\lambda}^k = \boldsymbol{H}\mathbf{w}^k$ that

$$
-\rho'^2_{\max}\|\mathbf{w}^k\|^2 \leq -\|\boldsymbol{H}\mathbf{w}^k\|^2 = -\|\boldsymbol{A}^\top\boldsymbol{\lambda}^k\|^2 \leq -\rho_{\min}\|\boldsymbol{\lambda}^k\|^2,
$$

which implies

$$
-\|\boldsymbol{\lambda}^k\|^2 \geq -\frac{\rho'^2_{\max}}{\rho_{\min}}\|\mathbf{w}^k\|^2.
$$

Given the assumption on β in Algorithm 1, it follows that $\rho'_{\min} - \frac{\rho'^2_{\max}}{\beta\rho_{\min}} > 0$. Considering that $\Omega(\mathbf{y}^k)$ is concave for gains and convex for losses, with a finite upper and lower bound due to loss aversion, we infer that $\mathbf{y}^k$ must also be bounded. If not, the boundedness of $U(\mathbf{y}^k)$ would be violated. Furthermore, since $\Phi(\mathbf{f}^k)$ and $\Phi(\mathbf{p}^k)$ are nonnegative, and $\Omega(\mathbf{y}^k)$ has both upper and lower bounds, we conclude that the sequences $\left\{\mathbf{f}^k\right\}$, $\left\{\mathbf{p}^k\right\}$, $\left\{\mathbf{y}^k\right\}$, $\left\{\mathbf{w}^k\right\}$, and $\left\{\boldsymbol{A}\mathbf{w}^k + \boldsymbol{B}\mathbf{f}^k + \boldsymbol{C}\mathbf{p}^k + \boldsymbol{D}\mathbf{y}^k - \mathbf{q} - \frac{\boldsymbol{\lambda}^k}{\beta}\right\}$ are bounded.

Moreover, it holds that

$$
\left\|\boldsymbol{\lambda}^k\right\| \leq \beta\left\|\boldsymbol{A}\mathbf{w}^k + \boldsymbol{B}\mathbf{f}^k + \boldsymbol{C}\mathbf{p}^k + \boldsymbol{D}\mathbf{y}^k - \mathbf{q}\right\| + \beta\left\|\boldsymbol{A}\mathbf{w}^k + \boldsymbol{B}\mathbf{f}^k + \boldsymbol{C}\mathbf{p}^k + \boldsymbol{D}\mathbf{y}^k - \mathbf{q} - \frac{\boldsymbol{\lambda}^k}{\beta}\right\|, \qquad (5.16)
$$

which implies that the sequence $\left\{\boldsymbol{\lambda}^k\right\}$ is also bounded. Consequently, the sequence $\left\{\left(\mathbf{f}^k, \mathbf{p}^k, \mathbf{y}^k, \mathbf{w}^k, \boldsymbol{\lambda}^k\right)\right\}$ is bounded. This completes the proof. □

Remark 2. *Conventional convergence analysis often relies on the coercivity of objective components such as* $\Phi(\mathbf{f}^k)$, $\Phi(\mathbf{p}^k)$, *and* $\Omega(\mathbf{y}^k)$ *to ensure boundedness of the iterates. Because* $\Omega(\mathbf{y}^k)$ *lacks coercivity, it becomes necessary to impose boundedness on the sequence* $\{\mathbf{y}^k\}$ *to guarantee convergence.*

Lemma 7. *Let* $\{(\mathbf{f}^k, \mathbf{p}^k, \mathbf{y}^k, \mathbf{w}^k, \boldsymbol{\lambda}^k)\}$ *be the sequence generated by Algorithm 1 and* $\beta > \widetilde{\beta}$. *If* $\{\mathbf{y}^k\}$ *is bounded, then*

$$\lim_{k\to\infty}\left(\left\|\mathbf{f}^{k+1}-\mathbf{f}^k\right\|+\left\|\mathbf{p}^{k+1}-\mathbf{p}^k\right\|+\left\|\mathbf{y}^{k+1}-\mathbf{y}^k\right\|+\left\|\mathbf{w}^{k+1}-\mathbf{w}^k\right\|+\left\|\boldsymbol{\lambda}^{k+1}-\boldsymbol{\lambda}^k\right\|\right)=0.$$

Proof. Suppose that $(\mathbf{f}^*, \mathbf{p}^*, \mathbf{y}^*, \mathbf{w}^*, \boldsymbol{\lambda}^*)$ is a cluster point of a sequence $(\mathbf{f}^k, \mathbf{p}^k, \mathbf{y}^k, \mathbf{w}^k, \boldsymbol{\lambda}^k)$ generated by Algorithm 1, and let $(\mathbf{f}^{k_i}, \mathbf{p}^{k_i}, \mathbf{y}^{k_i}, \mathbf{w}^{k_i}, \boldsymbol{\lambda}^{k_i})$ be a corresponding subsequence such that

$$\lim_{i\to\infty}\left(\mathbf{f}^{k_i}, \mathbf{p}^{k_i}, \mathbf{y}^{k_i}, \mathbf{w}^{k_i}, \boldsymbol{\lambda}^{k_i}\right)=\left(\mathbf{f}^*, \mathbf{p}^*, \mathbf{y}^*, \mathbf{w}^*, \boldsymbol{\lambda}^*\right).$$

Summing (5.6) from $k = 1$ to $k = k_i - 1$, we obtain

$$L_\beta\left(\mathbf{f}^{k_i}, \mathbf{p}^{k_i}, \mathbf{y}^{k_i}, \mathbf{w}^{k_i}, \boldsymbol{\lambda}^{k_i}\right)-L_\beta\left(\mathbf{f}^1, \mathbf{p}^1, \mathbf{y}^1, \mathbf{w}^1, \boldsymbol{\lambda}^1\right)\le -b\sum_{k=1}^{k_i-1}\|\mathbf{w}^{k+1}-\mathbf{w}^k\|^2,$$

which implies

$$b\sum_{k=1}^{\infty}\|\mathbf{w}^{k+1}-\mathbf{w}^k\|^2\le L_\beta\left(\mathbf{f}^1, \mathbf{p}^1, \mathbf{y}^1, \mathbf{w}^1, \boldsymbol{\lambda}^1\right)-L_\beta\left(\mathbf{f}^*, \mathbf{p}^*, \mathbf{y}^*, \mathbf{w}^*, \boldsymbol{\lambda}^*\right)<\infty.$$

Since $b > 0$, it follows that $\mathbf{w}^{k+1} - \mathbf{w}^k \to 0$. Summing the recurrence (5.9) over $k = 1, 2, \ldots, k_i - 1$ and subsequently passing to the limit yields

$$\sum_{k=1}^{\infty}\|\boldsymbol{\lambda}^{k+1}-\boldsymbol{\lambda}^k\|^2\le\frac{\rho'^2_{\max}}{\rho_{\min}}\sum_{k=1}^{\infty}\|\mathbf{w}^{k+1}-\mathbf{w}^k\|^2,$$

from which we conclude that $\boldsymbol{\lambda}^{k+1} - \boldsymbol{\lambda}^k \to 0$. Considering the following iterative form of $\boldsymbol{\lambda}^{k+1}$:

$$\begin{aligned}\boldsymbol{\lambda}^{k+1}&=\boldsymbol{\lambda}^k+\beta\left(\boldsymbol{A}\mathbf{w}^{k+1}+\boldsymbol{B}\mathbf{f}^{k+1}+\boldsymbol{C}\mathbf{p}^{k+1}+\boldsymbol{D}\mathbf{y}^{k+1}-\mathbf{q}\right)\\&=\boldsymbol{\lambda}^k+\beta\begin{pmatrix}\boldsymbol{F}\mathbf{w}^{k+1}-\mathbf{f}^{k+1}\\\mathbf{w}^{k+1}-\mathbf{p}^{k+1}\\\boldsymbol{G}\mathbf{w}^{k+1}-\mathbf{y}^{k+1}\\\boldsymbol{E}\mathbf{w}^{k+1}-\mathbf{b}\end{pmatrix},\end{aligned}$$

we can infer that $\mathbf{f}^{k+1} - \mathbf{f}^k \to 0$, $\mathbf{p}^{k+1} - \mathbf{p}^k \to 0$, and $\mathbf{y}^{k+1} - \mathbf{y}^k \to 0$. This completes the proof. □

Theorem 1. *Let* $\{(\mathbf{f}^k, \mathbf{p}^k, \mathbf{y}^k, \mathbf{w}^k, \boldsymbol{\lambda}^k)\}$ *be the sequence generated by Algorithm 1 with* $\beta > \widetilde{\beta}$. *If* $\{\mathbf{y}^k\}$ *is bounded, then any cluster point* $(\mathbf{f}^*, \mathbf{p}^*, \mathbf{y}^*, \mathbf{w}^*, \boldsymbol{\lambda}^*)$ *of the sequence* $\{(\mathbf{f}^k, \mathbf{p}^k, \mathbf{y}^k, \mathbf{w}^k, \boldsymbol{\lambda}^k)\}$ *is a stationary point of* (4.3).

Proof. It follows from Lemma 6 that the sequence $\{(\mathbf{f}^k, \mathbf{p}^k, \mathbf{y}^k, \mathbf{w}^k, \boldsymbol{\lambda}^k)\}$ generated by Algorithm 1 is bounded. Let $(\mathbf{f}^*, \mathbf{p}^*, \mathbf{y}^*, \mathbf{w}^*, \boldsymbol{\lambda}^*)$ be a cluster point of the sequence $\{(\mathbf{f}^k, \mathbf{p}^k, \mathbf{y}^k, \mathbf{w}^k, \boldsymbol{\lambda}^k)\}$, and there exists a subsequence $(\mathbf{f}^{k_i}, \mathbf{p}^{k_i}, \mathbf{y}^{k_i}, \mathbf{w}^{k_i}, \boldsymbol{\lambda}^{k_i})$ such that

$$\lim_{i\to\infty}\left(\mathbf{f}^{k_i}, \mathbf{p}^{k_i}, \mathbf{y}^{k_i}, \mathbf{w}^{k_i}, \boldsymbol{\lambda}^{k_i}\right) = (\mathbf{f}^*, \mathbf{p}^*, \mathbf{y}^*, \mathbf{w}^*, \boldsymbol{\lambda}^*)\,.$$

Since L_β is lower semicontinuous, it follows that

$$\liminf_{i\to\infty} L_\beta\left(\mathbf{f}^{k_i+1}, \mathbf{p}^{k_i+1}, \mathbf{y}^{k_i+1}, \mathbf{w}^{k_i+1}, \boldsymbol{\lambda}^{k_i}\right) \geq L_\beta\left(\mathbf{f}^*, \mathbf{p}^*, \mathbf{y}^*, \mathbf{w}^*, \boldsymbol{\lambda}^*\right). \tag{5.17}$$

On the other hand, from the definition of $\mathbf{w}^{k_i+1}$ in (4.5d), we have

$$L_\beta\left(\mathbf{f}^{k_i+1}, \mathbf{p}^{k_i+1}, \mathbf{y}^{k_i+1}, \mathbf{w}^{k_i+1}, \boldsymbol{\lambda}^{k_i}\right) \leq L_\beta\left(\mathbf{f}^{k_i+1}, \mathbf{p}^{k_i+1}, \mathbf{y}^{k_i+1}, \mathbf{w}^{*}, \boldsymbol{\lambda}^{k_i}\right).$$

By taking limits on both sides of the above inequality, we get

$$\limsup_{i\to\infty} L_\beta\left(\mathbf{f}^{k_i+1}, \mathbf{p}^{k_i+1}, \mathbf{y}^{k_i+1}, \mathbf{w}^{k_i+1}, \boldsymbol{\lambda}^{k_i}\right) \leq L_\beta\left(\mathbf{f}^*, \mathbf{p}^*, \mathbf{y}^*, \mathbf{w}^*, \boldsymbol{\lambda}^*\right). \tag{5.18}$$

Combining (5.17) and (5.18), we obtain

$$\lim_{i\to\infty} L_\beta\left(\mathbf{f}^{k_i+1}, \mathbf{p}^{k_i+1}, \mathbf{y}^{k_i+1}, \mathbf{w}^{k_i+1}, \boldsymbol{\lambda}^{k_i}\right) = L_\beta\left(\mathbf{f}^*, \mathbf{p}^*, \mathbf{y}^*, \mathbf{w}^*, \boldsymbol{\lambda}^*\right),$$

which, together with $\mathbf{f}^{k+1} - \mathbf{f}^k \to 0$, $\mathbf{p}^{k+1} - \mathbf{p}^k \to 0$, and $\mathbf{y}^{k+1} - \mathbf{y}^k \to 0$ from Lemma 7 and the definition of L_β in (4.4), implies that

$$\lim_{i\to\infty} \Phi(\mathbf{f}^{k_i+1}) = \Phi(\mathbf{f}^*), \quad \lim_{i\to\infty} \Phi(\mathbf{p}^{k_i+1}) = \Phi(\mathbf{p}^*), \quad \text{and} \quad \lim_{i\to\infty} \Omega(\mathbf{y}^{k_i+1}) = \Omega(\mathbf{y}^*). \tag{5.19}$$

Thus, passing to the limit in (4.5d)–(4.5e) along $\{(\mathbf{f}^{k_i}, \mathbf{p}^{k_i}, \mathbf{y}^{k_i}, \mathbf{w}^{k_i}, \boldsymbol{\lambda}^{k_i})\}$ and invoking Lemma 7, (5.19), and (5.1), we see that (5.2) holds. That is, $(\mathbf{f}^*, \mathbf{p}^*, \mathbf{y}^*, \mathbf{w}^*, \boldsymbol{\lambda}^*)$ is a critical point of (4.3). □

Given the Kurdyka–Łojasiewicz (KL) property [53, 54] and the earlier conclusions, the full sequence converges to a stationary point. The theorem below summarizes this fact. Its proof, similar to those in [53], [54], and [55], is omitted for brevity.

Theorem 2. *Let $\beta > \widetilde{\beta}$, suppose $\{\mathbf{y}^k\}$ is bounded, and let L_β be a KL function. Then, the sequence $\{(\mathbf{f}^k, \mathbf{p}^k, \mathbf{y}^k, \mathbf{w}^k, \boldsymbol{\lambda}^k)\}$ generated by Algorithm 1 converges to a critical point of* (4.3).

6. Extension to sparse ESG-based behavioral portfolio optimization

As the concept of sustainable development gains prominence, investors face increasing pressure to integrate environmental, social, and governance (ESG) factors into their investment decisions, making socially responsible investment ever more significant. ESG scoring serves as a vital tool in ESG investment, evaluating a company's sustainable development capabilities based on its performance in environmental, social, and corporate governance aspects. An expanding segment of the investment community regards the incorporation of ESG factors as both a moral imperative and a prudent strategy that mitigates risk and enhances long-term value creation [56, 57]. Therefore, ESG scoring is crucial

for investors committed to sustainable investing. In line with this perspective, this paper employs the A55 dataset to demonstrate the inclusion of ESG factors through a straightforward linear aggregation technique. We then explain precisely how these ESG metrics are reflected in the computation of returns and portfolio risk.

To incorporate ESG factors into asset returns, we define ESG-adjusted returns as a convex blend of traditional returns and normalized ESG scores, controlled by an investor-specific preference parameter $\kappa \in [0, 1)$. Following [58], the parameter κ quantifies the strength of the investor's sustainability orientation: $\kappa = 0$ recovers the purely profit-driven case, while values approaching 1 indicate increasingly pronounced ESG preferences. Let $ESG_{i,t}$ denote the raw ESG rating of asset i at time t, and let $\varsigma_{i,t} \in [0, 1]$ be the corresponding normalized score. To synchronize the annually updated ESG ratings with higher-frequency returns, we follow [59] and apply forward-filling: the most recent available rating is carried forward until a new score is released. The ESG ratings are sourced from the Wind database, where each score combines a management-practice submodule (total score of 7) and a controversy-event submodule (total score of 3), producing raw values in $[0, 10]$. To facilitate interpretation and ensure comparability, we rescale these ratings to the interval $[-1, 1]$ using the transformation

$$\varsigma_{i,t} = \frac{ESG_{i,t}}{5} - 1.$$

The normalized scores satisfy $\varsigma_{i,t} \in [-1, 1]$, where zero corresponds to the midpoint of the original 0–10 scale, and positive (negative) values therefore signal above-average (below-average) sustainability performance. For any investor preference parameter $\kappa \in [0, 1)$, the ESG-adjusted return of asset i at time t is constructed as

$$\zeta_{i,t}(\kappa) = (1 - \kappa)\, r_{i,t} + \kappa \cdot \frac{\varsigma_{i,t}}{c}, \tag{6.1}$$

where $r_{i,t}$ denotes the financial return and $\varsigma_{i,t}$ is the raw ESG score. This linear weighting structure is deliberately adopted to preserve model transparency and computational tractability, avoiding the estimation uncertainty inherent in complex nonlinear specifications. The scaling constant c aligns the magnitude of the ESG component with financial returns by setting it equal to the annual frequency of observations, specifically 252 for daily, 52 for weekly, and 12 for monthly data. This standardization ensures that the ESG term remains comparable to annualized financial performance regardless of the observation interval. Finally, κ quantifies the intensity of sustainability preferences, with higher values indicating a stronger commitment to ESG goals.

With ESG factors incorporated, problem (3.2) can be recast as:

$$\begin{aligned}
&\min_{\mathbf{w}} \sum_{j=1}^{m} \left[\frac{1}{2} \mathbf{w}_j^\top \widetilde{\boldsymbol{H}}_j \mathbf{w}_j - c_j U\left((\widetilde{\boldsymbol{G}}\mathbf{w})_{[j]} \right) + \tau_1 \left\| \mathbf{w}_j \right\|_1 \right] + \tau_2 \sum_{j=1}^{m-1} \left\| \mathbf{w}_{j+1} - \mathbf{w}_j \right\|_1 \\
\text{(ESG-MSBP)} \quad &\text{s.t. } \mathbf{w}_1^\top 1_n = \xi_{\text{init}}, \\
&\mathbf{w}_j^\top 1_n = \left(1_n + \zeta_{j-1} \right)^\top \mathbf{w}_{j-1},\ j = 2, \ldots, m, \\
&(1_n + \zeta_m)^\top \mathbf{w}_m = \xi_{\text{term}},
\end{aligned} \tag{6.2}$$

where the matrix $\widetilde{\boldsymbol{G}} \in \mathbb{R}^{m \times N}$ formed by m scenarios of the ESG-valued return vector ζ_j is defined as

follows:

$$\widetilde{\boldsymbol{G}} = \begin{pmatrix} \zeta_1^\top & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \zeta_2^\top & \ddots & \vdots \\ \vdots & \ddots & \ddots & \mathbf{0} \\ \mathbf{0} & \cdots & \mathbf{0} & \zeta_m^\top \end{pmatrix},$$

and $\widetilde{\boldsymbol{H}}_j \in \mathbb{R}^{n\times n}$ is the covariance matrix of the ESG-augmented returns in period j. The same algorithm used for model (3.2) can be applied to solve the above problem. Thus, the details are not elaborated here.

7. Numerical experiments

This section examines the empirical behavior of the proposed model (3.2) and its ADMM-based solution method using real stock market data. All computational experiments were performed on a 64-bit Windows 10 laptop equipped with an Intel(R) Core(TM) i5-8250U processor (1.60–1.80 GHz), 8 GB of RAM, and MATLAB 2021a.

In our experiments, we meticulously select three datasets: (1) NDX100, NASDAQ 100 stocks (95 constituents after discarding those with incomplete records), daily returns from January 2020 to January 2024; (2) FTES100, FTSE 100 stocks (98 constituents after discarding those with incomplete records), daily returns from January 2020 to January 2024; (3) A55, 55 Chinese A-share stocks with complete daily returns from January 2018 to December 2021 and corresponding annual ESG ratings.

We terminate the iteration when $\|\boldsymbol{E}\mathbf{w}^k - \mathbf{b}\|_2 \le \text{Tol}$ and $\|\boldsymbol{F}\mathbf{w}^k - \mathbf{f}^k\|_2 \le \text{Tol}$ with $\text{Tol} = 10^{-4}$. The maximum number of iterations allowed is set to 1000. We adopt the widely used prospect-theory parameters of [22]: loss-aversion coefficient $\mu = 2.25$, risk-attitude parameters $\alpha = \varrho = 0.88$, probability-weighting parameters $\gamma = 0.61$ (gains) and $\delta = 0.69$ (losses), and reference point $B = 0$. For faster convergence, the ADMM penalty parameter is adjusted adaptively as follows:

$$\beta^{k+1} = \begin{cases} \iota\beta^k, & \text{if } \left\|\mathbf{y}^{k+1} - G\mathbf{w}^{k+1}\right\| > \epsilon, \\ \beta^k, & \text{otherwise,} \end{cases}$$

where $\iota > 1$ is a constant [48]. The algorithm terminates once the primal residual requirement $\|\mathbf{y}^k - G\mathbf{w}^k\| \le \epsilon$ and the dual residual condition $\|\mathbf{y}^k - \mathbf{y}^{k-1}\| \le \epsilon$ are simultaneously satisfied. In all subsequent experiments that prioritize solution accuracy, we tighten the stopping tolerance to $\epsilon = 5\times10^{-3}$. Regarding the algorithmic hyperparameters, we adopt an adaptive penalty scheme to enhance robustness. Specifically, we initialize the penalty parameter $\beta_0 = 1$ and set the scaling factor $\iota = 3$, following the heuristics suggested in recent ADMM literature [40, 48]. This configuration dynamically adjusts the penalty to maintain a comparable magnitude between primal and dual residuals, thereby preventing stagnation and accelerating convergence in nonconvex environments. The risk-free rate is fixed at $r_f = 1.03 \times 10^{-2}$, aligning with the average market yield observed during our sample period.

7.1. Performance criteria

To examine how the main parameters influence the proposed multi-period sparse behavioral portfolio model, we compare its out-of-sample performance with the traditional equal-weight ($1/n$) strategy,

which distributes wealth evenly across all n assets at each rebalance point. The equal-weighting $(1/n)$ strategy, which we refer to as the naive benchmark, is used throughout for comparison. We normalize initial wealth to one monetary unit, i.e., $\xi_{\text{init}} = 1$. Let ξ_{naive} denote the expected terminal wealth generated by the naive portfolio under the historical return distribution. To ensure a level playing field, the terminal wealth target of the optimized portfolio is fixed at exactly this value: $\xi_{\text{term}} = \xi_{\text{naive}}$. The expected wealth of the naive portfolio is computed by repeatedly applying the $1/n$ allocation rule—namely, at each rebalancing time, the total wealth is distributed equally across all available assets [14]:

$$\xi_{naive} = \frac{1}{n}(\cdots(\frac{1}{n}(\frac{\xi_{init}}{n}(1_n + \mathbf{r}_1)^\top 1_n)(1_n + \mathbf{r}_2)^\top 1_n)\cdots)(1_n + \mathbf{r}_m)^\top 1_n.$$

To assess both the risk exposure and the implementation cost of the optimized portfolios, we employ several performance metrics. The risk of the optimal portfolio is evaluated as follows:

$$Risk = \mathbf{w}_{opt}^\top \boldsymbol{H} \mathbf{w}_{opt},$$

where $\mathbf{w}_{opt}$ represents the optimal portfolio allocation, and the term $\mathbf{w}_{opt}^\top \boldsymbol{H} \mathbf{w}_{opt}$ provides its estimated risk. This measure, denoted by $Risk$, reflects the level of risk reduction achieved. We then compute the proportion of active positions to approximate holding costs, and evaluate changes in portfolio weights to estimate transaction costs. If $(\mathbf{w}_j)i \neq (\mathbf{w}j + 1)_i$, asset i is considered traded during the interval $[j, j + 1]$, and the number of weight adjustments across periods is used to approximate the transaction cost. Accordingly, we define the matrix $V \in \mathbb{R}^{n\times(m-1)}$ as follows:

$$V_{i,j} = \begin{cases} 0, & \text{if } |(\mathbf{w}_{j+1})_i - (\mathbf{w}_j)_i)| < \varepsilon * \xi_{init}, \\ 1, & \text{otherwise,} \end{cases}$$

where $i = 1, 2, \ldots, n$ and $j = 1, 2, \ldots, m - 1$. Since turnover is measured from a practical trading perspective, we adopt a minimum materiality threshold of 10^{-3}. Weight changes smaller than this value are treated as negligible and are not counted toward turnover or rebalancing activity. The number of transactions for the optimal portfolio is then given by

$$T_{opt} = \sum_{i=1}^{n} \sum_{j=1}^{m-1} V_{i,j}.$$

The naive equal-weighting strategy maintains a uniform allocation across all assets at each rebalancing point, so:

$$T_{naive} = (m - 1) \times n.$$

We measure the transaction intensity of the optimized strategy via the following turnover metric:

$$T = \frac{T_{opt}}{T_{naive}}.$$

Additionally, *Density* (%) denotes the average proportion of active (non-zero) positions, while *Shorts* measures the average number of short positions held.

As a core metric for measuring risk-adjusted returns, the Sharpe ratio (SR) has the following standard calculation formula:

$$SR = \frac{E[R_p - R_f]}{\sigma_p},$$

where $E[R_p]$ denotes the expected return of the portfolio, R_f represents the risk-free rate of return, and σ_p is the standard deviation of the portfolio return (i.e., portfolio risk). This metric reflects the excess return obtained per unit of risk.

7.2. Empirical study

In this subsection, we conduct experiments over a 5-year period using two datasets: NDX100 and FTSE100. The selection of the regularization parameters τ_1 and τ_2 plays a critical role in determining the out-of-sample performance of the resulting portfolios. Specifically, τ_1 reduces the density and the number of short positions, while τ_2 primarily influences transaction costs. An increase in τ_2 may also influence sparsity, requiring the use of regularization parameters with different magnitudes to meet predefined financial criteria [14]. To test the robustness of the proposed model, we explore various combinations of τ_1 and τ_2, where the values of both parameters are selected from the set $\{10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}\}$ based on cross-validation for experimental analysis.

Table 1. Numerical results under different combinations of τ_1 and τ_2 for the NDX100 dataset.

τ_1	τ_2	MSBP					Fused LASSO				
		SR	*Density*(%)	*Risk*	*T*(%)	*Shorts*	*SR*	*Density*(%)	*Risk*	*T*(%)	*Shorts*
10^{-4}	10^{-4}	0.1414	79.82	0.0153	58.68	182	0.1138	97.02	0.0176	80.26	230
10^{-4}	10^{-3}	0.1245	75.96	0.0174	46.84	172	0.0861	92.11	0.0232	60.00	194
10^{-4}	10^{-2}	0.0714	70.70	0.0303	29.21	155	0.0526	95.79	0.0380	38.42	208
10^{-4}	10^{-1}	0.0494	76.14	0.0437	26.84	164	0.0505	96.84	0.0396	37.89	216
10^{-3}	10^{-4}	0.1261	55.44	0.0171	50.00	109	0.0906	61.75	0.0221	63.95	117
10^{-3}	10^{-3}	0.1126	49.82	0.0192	41.58	88	0.0748	63.33	0.0267	50.26	117
10^{-3}	10^{-2}	0.0699	49.82	0.0309	25.79	89	0.0499	68.95	0.0401	28.68	121
10^{-3}	10^{-1}	0.0479	61.23	0.0450	23.68	128	0.0478	64.21	0.0418	24.47	114
10^{-2}	10^{-4}	0.0731	21.23	0.0295	25.26	19	0.0518	24.91	0.0386	29.47	6
10^{-2}	10^{-3}	0.0712	21.58	0.0303	22.63	18	0.0498	26.49	0.0401	25.26	6
10^{-2}	10^{-2}	0.0577	21.05	0.0374	15.26	11	0.0401	25.79	0.0498	11.58	2
10^{-2}	10^{-1}	0.0401	26.32	0.0540	10.26	24	0.0382	26.32	0.0524	11.05	6
10^{-1}	10^{-4}	0.0312	8.60	0.0692	12.89	7	0.0470	22.81	0.0425	25.79	0
10^{-1}	10^{-3}	0.0311	8.60	0.0694	12.37	7	0.0457	25.26	0.0437	22.89	0
10^{-1}	10^{-2}	0.0307	11.05	0.0703	10.79	7	0.0381	27.02	0.0525	11.84	0
10^{-1}	10^{-1}	0.0354	8.07	0.0610	6.84	2	0.0377	26.49	0.0530	10.53	0

Table 2. Numerical results under different combinations of τ_1 and τ_2 for the FTSE100 dataset.

τ_1	τ_2	MSBP					Fused LASSO				
		SR	$Density(\%)$	$Risk$	$T(\%)$	$Shorts$	SR	$Density(\%)$	$Risk$	$T(\%)$	$Shorts$
10^{-4}	10^{-4}	0.1443	71.43	0.0129	55.36	146	0.1748	93.88	0.0114	81.12	227
10^{-4}	10^{-3}	0.1262	70.24	0.0147	39.54	135	0.1211	94.22	0.0165	56.38	215
10^{-4}	10^{-2}	0.0786	68.88	0.0237	27.30	146	0.0727	94.90	0.0275	36.22	216
10^{-4}	10^{-1}	0.0634	76.19	0.0293	26.53	175	0.0716	94.90	0.0280	34.95	216
10^{-3}	10^{-4}	0.1290	45.75	0.0144	46.94	66	0.0127	50.34	0.0157	59.18	85
10^{-3}	10^{-3}	0.1156	39.12	0.0161	33.67	47	0.1010	58.50	0.0197	42.86	90
10^{-3}	10^{-2}	0.0801	44.05	0.0232	19.39	87	0.0702	58.33	0.0285	22.45	114
10^{-3}	10^{-1}	0.0624	52.89	0.0298	21.43	101	0.0671	61.22	0.0298	22.19	120
10^{-2}	10^{-4}	0.0906	20.92	0.0205	25.00	10	0.0805	27.55	0.0248	32.91	6
10^{-2}	10^{-3}	0.0877	20.75	0.0212	19.90	10	0.0764	30.27	0.0262	26.02	7
10^{-2}	10^{-2}	0.0731	19.73	0.0255	11.22	6	0.0576	26.02	0.0347	12.76	4
10^{-2}	10^{-1}	0.0478	18.54	0.0389	7.91	6	0.0487	26.53	0.0410	9.95	0
10^{-1}	10^{-4}	0.0361	4.93	0.0516	6.89	2	0.0631	25.68	0.0317	28.83	0
10^{-1}	10^{-3}	0.0360	5.10	0.0517	6.89	2	0.0606	28.06	0.0330	21.94	0
10^{-1}	10^{-2}	0.0373	4.25	0.0498	5.61	2	0.0488	25.85	0.0410	10.71	0
10^{-1}	10^{-1}	0.0421	5.61	0.0441	3.32	0	0.0487	25.51	0.0410	9.95	0

Table 3. Numerical results for the MSBP model without sparsity regularization ($\tau_1 = 0$).

τ_2	NDX100					FTSE100				
	SR	$Density(\%)$	$Risk$	$T(\%)$	$Shorts$	SR	$Density(\%)$	$Risk$	$T(\%)$	$Shorts$
10^{-4}	0.0649	63.16	0.1052	29.74	7	0.0265	54.08	0.0701	25.26	1
10^{-3}	0.0649	63.33	0.1052	29.74	7	0.0265	54.08	0.0701	25.26	1
10^{-2}	0.0649	62.63	0.1051	29.47	7	0.0266	54.08	0.0699	25.26	1
10^{-1}	0.0649	59.30	0.1053	29.74	6	0.0267	52.55	0.0696	25.77	1

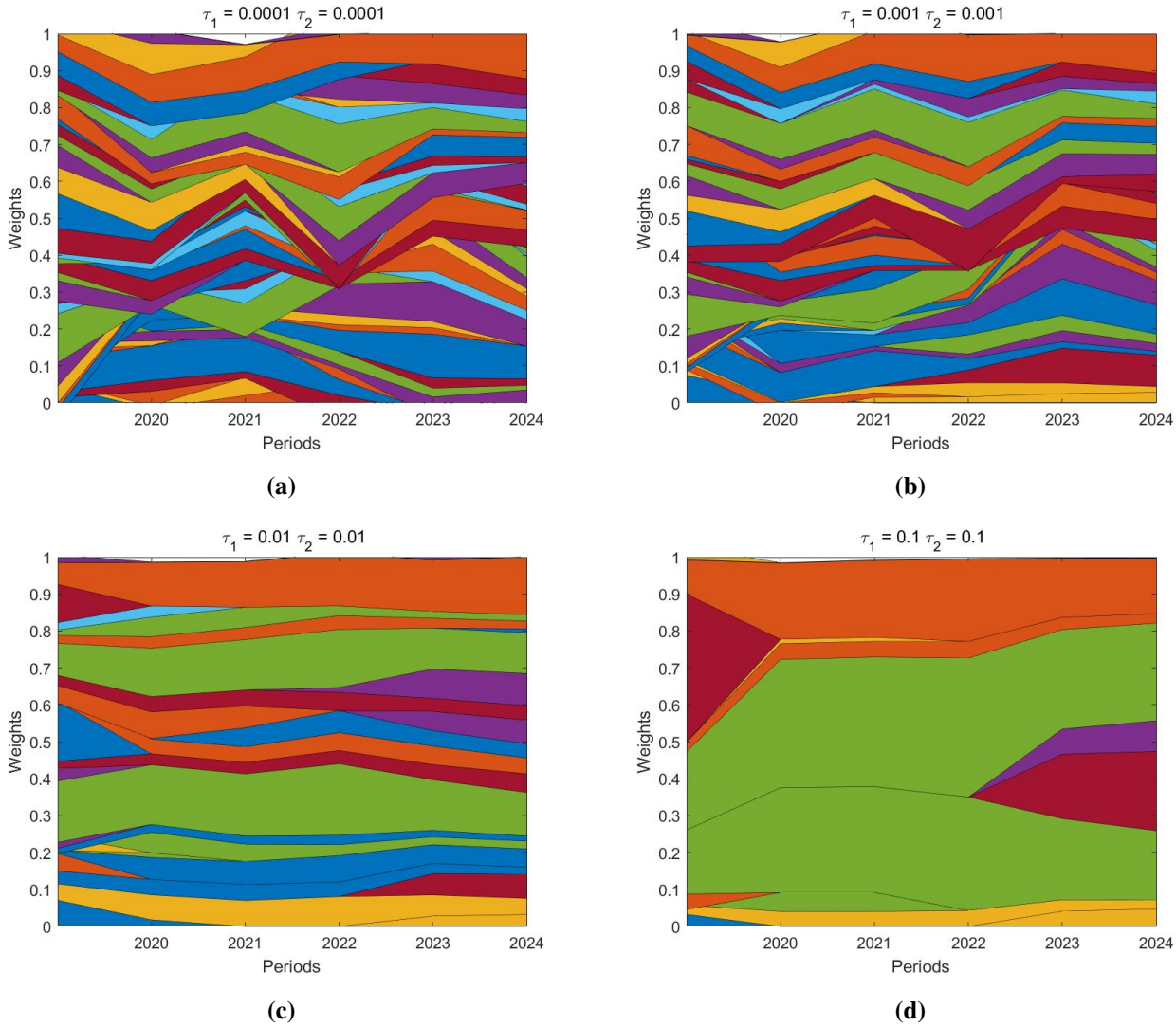

Figure 1. Evolutions of asset weights with respect to periods for the multi-period sparse behavioral portfolio model on the NDX100 dataset ((a)–(d)).

Tables 1 and 2 clearly delineate the five-year out-of-sample portfolio performance of our proposed MSBP model against the fused LASSO model, using both the NDX100 and FTSE100 datasets. Our analysis centers on five key evaluative metrics: portfolio density, realized risk, turnover rate, number of short positions, and the Sharpe ratio (SR). Across both datasets, we observed a consistent pattern: as the regularization parameters τ_1 and τ_2 increase, portfolio density, turnover rate, and the number of short positions all exhibit a monotonically decreasing trend. This phenomenon robustly validates the superior efficacy of fused LASSO regularization in enhancing portfolio sparsity and inter-period smoothing. Further investigation reveals that increasing the values of τ_1 and τ_2 significantly reduces the optimal portfolio density for both NDX100 and FTSE100, underscoring the pivotal role of the penalty terms in driving sparse solutions; stronger regularization constraints tend to yield more parsimonious and concentrated portfolios. At the risk level, while larger regularization parameters generally correlate with a slight increase in realized risk, the overall risk levels consistently remain subdued, indicating commendable model robustness in risk management.

Furthermore, as the penalty strength intensifies, the turnover rates for both portfolios decrease markedly, reaching their minimum at $\tau_1 = \tau_2 = 10^{-1}$. This reflects the salutary effect of the regularization mechanism in curbing excessive trading and mitigating potential transaction costs. Compared to the fused LASSO model, the MSBP model not only achieves marginally higher Sharpe ratios in most scenarios but also constructs sparser and more stable portfolio structures. Although in isolated parameter settings, certain metrics of the MSBP model may slightly lag behind the fused LASSO, the magnitude of these discrepancies is minimal, and its overall performance maintains a distinct advantage. Table 3 further reports the portfolio performance of the pure CPT model (i.e., $\tau_1 = 0$, without LASSO regularization). The results indicate that portfolios constructed solely based on behavioral preferences typically exhibit substantially higher realized risks and lower Sharpe ratios, explicitly demonstrating that the absence of regularization constraints compromises portfolio stability and risk-adjusted returns. This finding emphasizes the critical role of fused LASSO regularization in ameliorating portfolio risk-return characteristics and enhancing inter-period consistency.

Our empirical findings consistently demonstrate that neither the standard fused LASSO model nor the standalone CPT behavioral model surpasses the MSBP model proposed herein in terms of comprehensive performance metrics. The MSBP model achieves a more balanced and robust performance by simultaneously orchestrating risk levels, turnover rates, and portfolio complexity. This advantage is particularly pronounced within the context of multi-period portfolio selection, where strategy stability, operational feasibility, and constraint satisfaction are equally paramount as the maximization of a singular performance metric. The superior performance of the MSBP model stems from the complementary synergy between CPT behavioral preferences and LASSO regularization: the former provides directional guidance for asset allocation, while the latter reinforces risk control and model robustness through structural constraints.

Figures 1 and 2 display the time evolution of optimal portfolio weights in the sparse behavioral model for the NDX100 and FTSE100 datasets, respectively, under varying levels of the regularization parameters τ_1 and τ_2. In each panel, a colored band represents a single asset that receives positive allocation, the width of the band is fixed, and its vertical height is proportional to the weight assigned to that asset. As τ_1 and τ_2 increase, the number of colored bands visibly declines and progressively shrinks, whereas the height of the remaining bands tends to rise. This pattern reflects growing portfolio concentration and confirms that stronger regularization induces markedly higher sparsity. The visual evidence thus corroborates our earlier quantitative findings: larger penalty parameters systematically produce sparser behavioral portfolios.

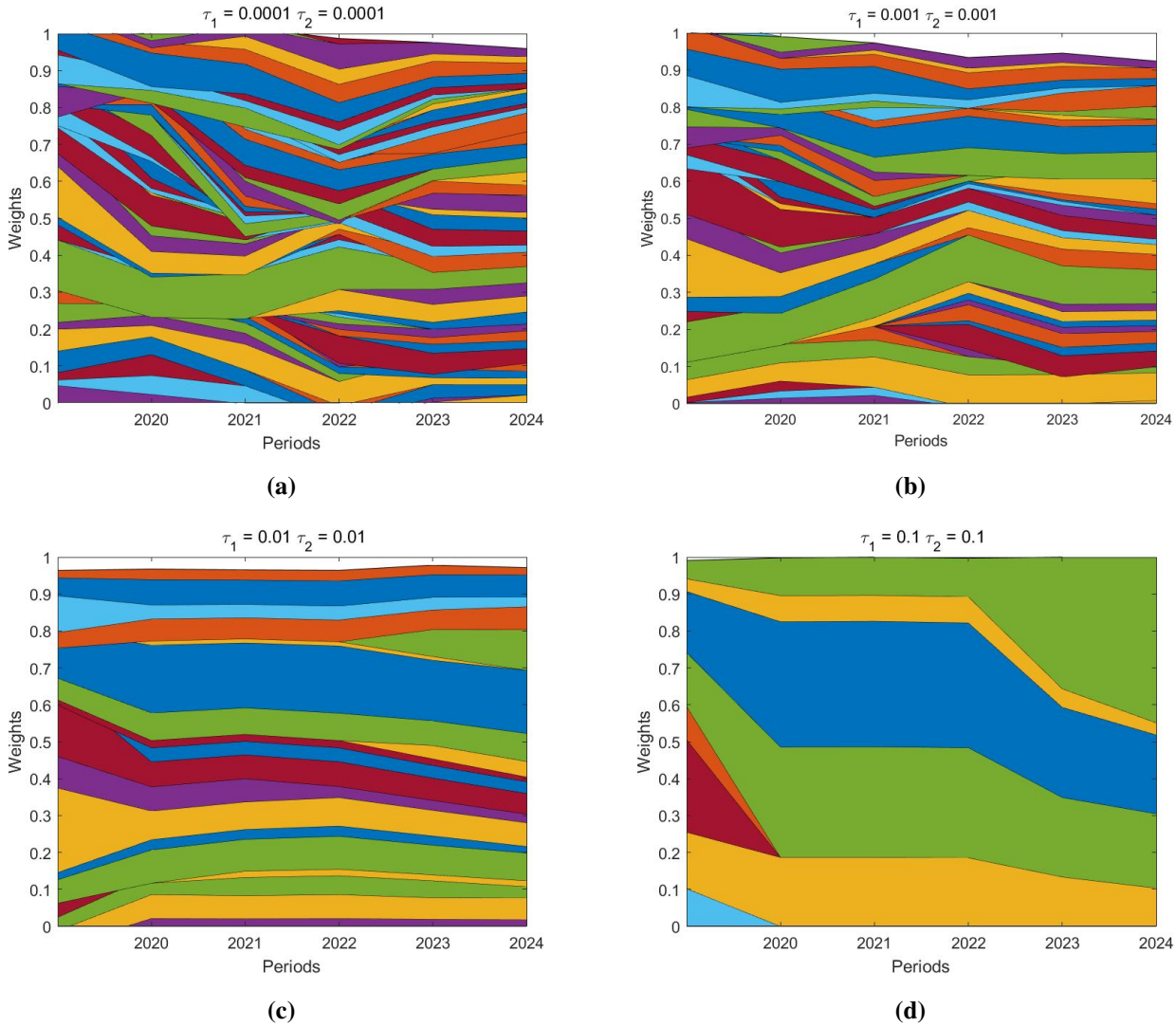

Figure 2. Evolutions of asset weights with respect to periods for the multi-period sparse behavioral portfolio model on the FTSE100 dataset ((a)–(d)).

7.3. Empirical study of considering ESG factors

In this subsection, we examine the sensitivity of the proposed model to its key parameters—most notably the ESG preference intensity κ introduced in (6.1)—using the A55 dataset. To assess robustness, we consider an extensive grid of regularization strengths $\tau_1, \tau_2 \in \{10^{-3}, 10^{-2}, 10^{-1}\}$.

Table 4. Numerical results under different combinations of κ, τ_1, and τ_2 for the A55 dataset.

τ_1	τ_2	Density (%)					Risk				
		κ=0	κ=0.25	κ=0.5	κ=0.75	κ=0.99	κ=0	κ=0.25	κ=0.5	κ=0.75	κ=0.99
10^{-3}	10^{-3}	68.36	66.18	55.64	55.27	61.45	0.0440	0.0319	0.0149	0.0050	0.000012
10^{-3}	10^{-2}	57.82	60.00	54.55	65.09	63.27	0.0557	0.0383	0.0182	0.0060	0.000012
10^{-3}	10^{-1}	63.27	69.09	56.36	69.09	70.55	0.0670	0.0448	0.0222	0.0066	0.000013
10^{-2}	10^{-3}	34.91	28.00	29.82	34.55	36.00	0.0541	0.0369	0.0181	0.0057	0.000011
10^{-2}	10^{-2}	34.55	25.82	29.82	30.18	29.45	0.0617	0.0414	0.0198	0.0067	0.000013
10^{-2}	10^{-1}	36.00	24.00	26.18	22.91	25.82	0.0720	0.0445	0.0219	0.0071	0.000012
10^{-1}	10^{-3}	21.82	21.45	20.00	15.64	15.64	0.0746	0.0455	0.0240	0.0073	0.000012
10^{-1}	10^{-2}	22.91	21.45	20.36	16.36	17.45	0.0727	0.0445	0.0229	0.0076	0.000012
10^{-1}	10^{-1}	26.55	23.64	24.73	24.73	21.09	0.0799	0.0494	0.0254	0.0073	0.000013
τ_1	τ_2	T (%)					Shorts				
		κ=0	κ=0.25	κ=0.5	κ=0.75	κ=0.99	κ=0	κ=0.25	κ=0.5	κ=0.75	κ=0.99
10^{-3}	10^{-3}	74.55	69.70	58.79	53.33	52.12	67	69	39	18	16
10^{-3}	10^{-2}	46.06	46.06	40.61	41.82	41.92	41	58	45	41	34
10^{-3}	10^{-1}	47.27	46.67	38.18	44.85	43.64	60	80	60	67	72
10^{-2}	10^{-3}	47.88	36.36	36.36	37.58	38.79	10	1	1	0	0
10^{-2}	10^{-2}	31.52	23.64	24.85	24.24	24.24	3	0	2	2	2
10^{-2}	10^{-1}	31.52	26.06	23.03	19.39	16.97	10	3	5	4	3
10^{-1}	10^{-3}	31.25	29.70	27.27	22.42	20.61	0	0	0	0	0
10^{-1}	10^{-2}	25.45	22.42	21.82	15.76	16.36	0	0	0	0	0
10^{-1}	10^{-1}	23.03	21.82	18.18	18.79	13.94	0	0	0	0	0

Table 4 summarizes the four-year out-of-sample results for the sparse behavioral portfolio on the A55 dataset, highlighting the effect of the ESG preference parameter κ. The portfolio becomes less dense as the regularization terms (τ_1 and τ_2) increase, and this sparsity intensifies further with larger values of κ, indicating that ESG preference itself promotes sparsity beyond the impact of the penalty parameters. In particular, higher emphasis on ESG ratings leads to a more concentrated portfolio. Holding κ fixed, the estimated risk of portfolio A55 rises with increasing τ_1 and τ_2, yet stays below 0.1, remaining within an acceptable level. Conversely, varying κ leads to a general decrease in estimated risk, reaching its lowest value when κ equals 0.99, suggesting that ESG ratings can effectively mitigate portfolio risk. The transaction percentage for portfolio A55 consistently decreases with higher τ_1 and τ_2 values and declines significantly as κ increases. From a financial standpoint, ESG ratings play an important role in managing the portfolio's transaction costs. Moreover, the occurrence of short positions in portfolio A55 generally decreases as τ_1 and τ_2 grow, and exhibits an even sharper reduction when κ takes larger values. Notably, when κ is 0.75, no short sales occur at $\tau_1 = 10^{-2}$, demonstrating the significant role of ESG ratings in limiting short positions, thereby reducing extreme cases and minimizing errors. In summary, ESG ratings positively affect the transaction costs and number of short positions in the behavioral portfolio, promoting sparsity and meeting financial requirements.

Figures 3 (a)–(e) illustrate the evolution of optimal weights in the behavioral portfolio over time for various κ values, maintaining $\tau_1 = \tau_2 = 0.1$. The colored regions denote asset allocations based on wealth, with each region's height representing the allocated amount. As κ increases, the number of colored regions decreases, while their respective heights generally rise. This trend indicates an increase in portfolio sparsity due to a stronger emphasis on the ESG rating. This visualization reinforces the conclusion that ESG ratings have a pronounced effect on portfolio sparsity. Increasing emphasis on ESG criteria leads to a noticeably sparser behavioral portfolio.

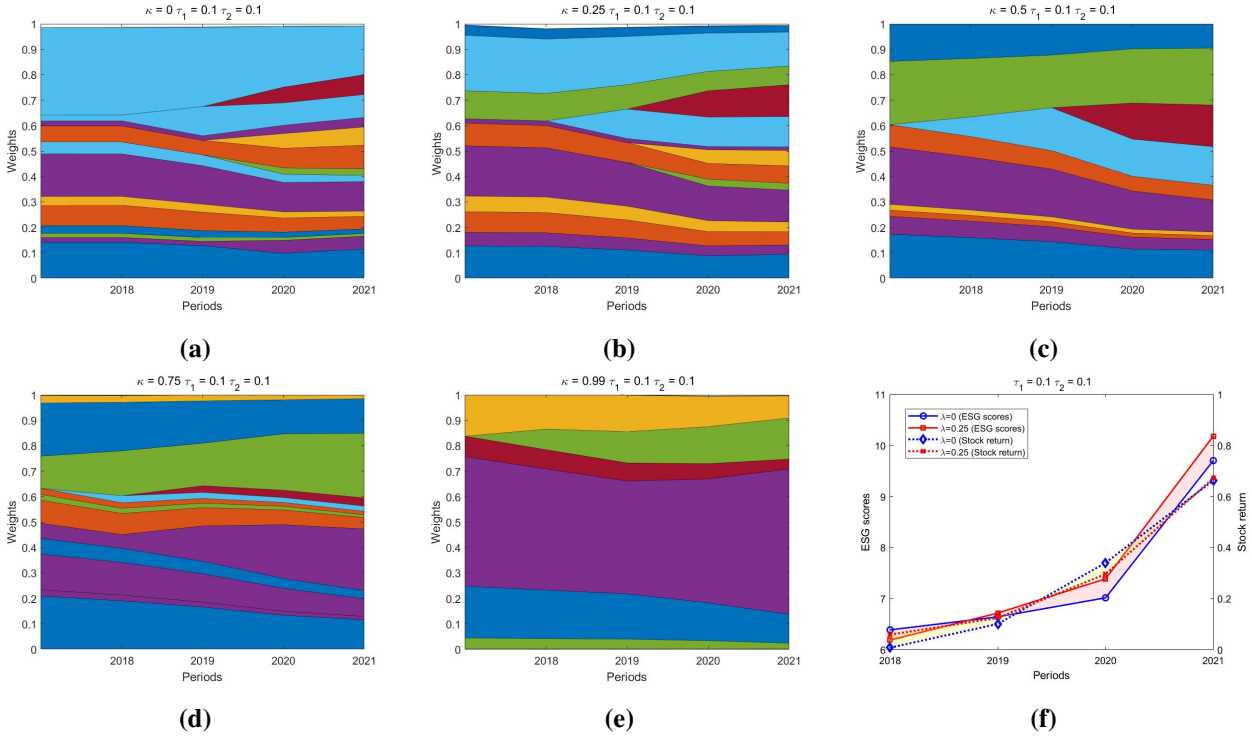

Figure 3. Figures ((a)–(e)) depict the period-by-period evolution of asset weights for the multi-period sparse ESG-enhanced behavioral portfolio on the A55 dataset. Figure 3(f) contrasts the weighted average ESG scores and returns for portfolios with varying ESG preference intensities.

8. Conclusion

In this study, we recognized that investors exhibit systematic behavioral biases—particularly asymmetric attitudes toward gains and losses—that materially affect their portfolio decisions. To capture these real-world preferences more faithfully than traditional expected-utility frameworks, we embedded the value function and probability-weighting scheme of cumulative prospect theory (CPT) into a fused-LASSO regularization framework. The resulting multi-period sparse behavioral portfolio model simultaneously enforces cardinality constraints, encourages grouping effects across consecutive periods, and reflects genuine investor psychology. This model captured investors' irrational behavioral traits through a multi-period optimization framework. We improved portfolio sparsity within each period and lowered inter-period turnover by incorporating two ℓ_1 regularization terms, offering a versatile

framework for managing portfolios across diverse asset sets. Numerically, we employed the alternating direction method of multipliers (ADMM) and the pooling-adjacent-violators (PAV) algorithm to solve the multi-period sparse behavioral portfolio model and provided a convergence analysis, empirically demonstrating their effectiveness and superiority. We carried out several experiments on diverse real-world stock datasets to demonstrate the benefits of our proposed multi-period sparse behavioral portfolio model. Furthermore, given the growing emphasis on sustainable development and socially responsible investing, we extended our model to a multi-period sparse ESG-based behavioral portfolio by incorporating ESG ratings and a linear combination of traditional returns as ESG-valued returns. This extension analyzed the impact of ESG factors on the structure and performance of sparse behavioral portfolios. Our results show that the proposed sparse behavioral portfolio optimization model improves investment efficiency and lowers risk relative to the fused LASSO benchmark. Moreover, ESG ratings were found to mitigate risk, enhance the Sharpe ratio, and provide useful guidance for selecting sparse behavioral portfolios within a sustainability-oriented framework. For future work, we plan to examine the impact of alternative regularization schemes in multi-period behavioral settings. We also aim to study how additional risk measures can be incorporated into the behavioral portfolio optimization framework.

Author contributions

Qingyang Wang: Conceptualization, methodology, formal analysis, writing–original draft; Kunpeng Zhu: Data curation, investigation, writing—original draft, formal analysis; Yanjing Guo: Methodology, formal analysis, validation, writing–review and editing; Liu Yang: Visualization, software, writing—original draft; Zhongming Wu: Supervision, writing—review and editing.

Use of Generative–AI tools declaration

In the preparation of this work, the authors used Generative AI tools such as ChatGPT to assist in improving the clarity of the language. All AI-assisted content was carefully reviewed and revised by the authors, who take full responsibility for the final version of the manuscript.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (Nos. 12471291, 12001286), the Natural Science Foundation of Jiangsu Province (No. BK20241899), and the China Postdoctoral Science Foundation (No. 2022M711672).

Conflict of interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

1. A. Gunjan, S. Bhattacharyya, A brief review of portfolio optimization techniques, *Artif. Intell. Rev.*, **56** (2023), 3847–3886. http://doi.org/10.1007/s10462-022-10273-7

2. G. Curtis, Modern portfolio theory and behavioural finance, *J. Wealth Manag.*, **7** (2004), 16–22. http://doi.org/10.3905/jwm.2004.434562
3. V. N. C. Mushinada, Are individual investors irrational or adaptive to market dynamics? *J. Behav. Exp. Finance*, **25** (2020), 100243. http://doi.org/10.1016/j.jbef.2019.100243
4. Y. Chen, B. Li, An uncertainty theory based tri-objective behavioral portfolio selection model with loss aversion and reference level using a modified evolutionary root system growth algorithm, *J. Comput. Appl. Math.*, **446** (2024), 115859. http://doi.org/10.1016/j.cam.2024.115859
5. X. Li, B. Li, T. Jin, P. Zheng, Uncertain random portfolio optimization with non-dominated sorting genetic algorithm-II and optimal solution criterion, *Artif. Intell. Rev.*, **56** (2023), 8511–8546. http://doi.org/10.1007/s10462-022-10388-x
6. X. Li, A. S. Uysal, J. M. Mulvey, Multi-period portfolio optimization using model predictive control with mean-variance and risk parity frameworks, *Eur. J. Oper. Res.*, **299** (2022), 1158–1176. http://doi.org/10.1016/j.ejor.2021.10.002
7. D. Li, W. L. Ng, Optimal dynamic portfolio selection: Multiperiod mean-variance formulation, *Math. Finance*, **10** (2000), 387–406. http://doi.org/10.1111/1467-9965.00100
8. X. Cui, J. Gao, X. Li, D. Li, Optimal multi-period mean–variance policy under no-shorting constraint, *Eur. J. Oper. Res.*, **234** (2014), 459–468. http://doi.org/10.1016/j.ejor.2013.02.040
9. Z. Chen, G. Consigli, J. Liu, G. Li, T. Fu, Q. Hu, Multi-period risk measures and optimal investment policies, In: *Optimal financial decision making under uncertainty*, 2016. http://doi.org/10.1007/978-3-319-41613-7_1
10. M. A. Moghadam, S. B. Ebrahimi, D. Rahmani, A constrained multi-period robust portfolio model with behavioral factors and an interval semi-absolute deviation, *J. Comput. Appl. Math.*, **374** (2020), 112742. http://doi.org/10.1016/j.cam.2020.112742
11. Y. Dai, Z. Qin, Multi-period uncertain portfolio optimization model with minimum transaction lots and dynamic risk preference, *Appl. Soft Comput.*, **109** (2021), 107519. http://doi.org/10.1016/j.asoc.2021.107519
12. R. Tibshirani, Regression shrinkage and selection via the lasso, *J. R. Stat. Soc. Ser. B*, **58** (1996), 267–288. http://doi.org/10.1111/j.2517-6161.1996.tb02080.x
13. R. Tibshirani, M. Saunders, S. Rosset, J. Zhu, K. Knight, Sparsity and smoothness via the fused LASSO, *J. R. Stat. Soc. Ser. B*, **67** (2005), 91–108. http://doi.org/10.1111/j.1467-9868.2005.00490.x
14. S. Corsaro, V. De Simone, Z. Marino, Fused lasso approach in portfolio selection, *Ann. Oper. Res.*, **299** (2021), 47–59. http://doi.org/10.1007/s10479-019-03289-w
15. F. Cesarone, J. Moretti, F. Tardella, Optimally chosen small portfolios are better than large ones, *Econ. Bull.*, **36** (2016), 1876–1891.
16. M. Rezaei, H. Nezamabadi-Pour, A taxonomy of literature reviews and experimental study of deep reinforcement learning in portfolio management, *Artif. Intell. Rev.*, **58** (2025), 94. http://doi.org/10.1007/s10462-024-11066-w

17. P. N. Kolm, R. Tümentüncü, F. J. Fabozzi, 60 years of portfolio optimization: Practical challenges and current trends, *Eur. J. Oper. Res.*, **234** (2014), 356–371. http://doi.org/10.1016/j.ejor.2013.10.060
18. Z. Yang, Y. Zhao, Hybrid SGD algorithms to solve stochastic composite optimization problems with application in sparse portfolio selection problems, *J. Comput. Appl. Math.*, **436** (2024), 115425. http://doi.org/10.1016/j.cam.2023.115425
19. Y. Wang, L. Kong, H. Qi, An efficient Lagrange–Newton algorithm for long-only cardinality constrained portfolio selection on real data sets, *J. Comput. Appl. Math.*, **461** (2025), 116453. http://doi.org/10.1016/j.cam.2024.116453
20. W. Guo, W. G. Zhang, X. Chen, Portfolio selection models considering fuzzy preference relations of decision makers, *IEEE Trans. Syst. Man Cybern. Syst.*, **54** (2024), 2254–2265. http://doi.org/10.1109/TSMC.2023.3342038
21. D. Kahneman, Prospect theory: An analysis of decisions under risk, *Econometrica*, **47** (1979), 278. http://doi.org/10.2307/1914185
22. A. Tversky, D. Kahneman, Advances in prospect theory: Cumulative representation of uncertainty, *J. Risk Uncertain.*, **5** (1992), 297–323. http://doi.org/10.1007/BF00122574
23. X. D. He, X. Y. Zhou, Portfolio choice under cumulative prospect theory: An analytical treatment, *Manag. Sci.*, **57** (2011), 315–331. http://doi.org/10.1287/mnsc.1100.1269
24. X. Chen, X. Liu, Q. Zhu, Z. Wang, DEA cross-efficiency models with prospect theory and distance entropy: An empirical study on high-tech industries, *Expert Syst. Appl.*, **244** (2024), 122941. http://doi.org/10.1016/j.eswa.2023.122941
25. Y. Shi, X. Cui, J. Yao, D. Li, Dynamic trading with reference point adaptation and loss aversion, *Oper. Res.*, **63** (2015), 789–806. http://doi.org/10.1287/opre.2015.1399
26. Y. E. Rodríguez, J. M. Gómez, J. Contreras, Diversified behavioral portfolio as an alternative to modern portfolio theory, *North Am. J. Econ. Finance*, **58** (2021), 101508. http://doi.org/10.1016/j.najef.2021.101508
27. S. van Bilsen, R. J. A. Laeven, Dynamic consumption and portfolio choice under prospect theory, *Insur. Math. Econ.*, **91** (2020), 224–237. http://doi.org/10.1016/j.insmatheco.2020.02.004
28. E. Luxenberg, P. Schiele, S. Boyd, Portfolio Optimization with Cumulative Prospect Theory Utility via Convex Optimization, *Comput. Econ.*, **64** (2024), 3027–3047. https://doi.org/10.1007/s10614-024-10556-x
29. Y. Shi, X. Cui, D. Li, Discrete-time behavioural portfolio selection under cumulative prospect theory, *J. Econ. Dyn. Control*, **61** (2015), 283–302. http://doi.org/10.1016/j.jedc.2015.10.002
30. S. Srivastava, A. Aggarwal, A. Mehra, Portfolio selection by cumulative prospect theory and its comparison with mean-variance model, *Granul. Comput.*, **7** (2022), 903–916. http://doi.org/10.1007/s41066-021-00302-1
31. T. Hens, J. Mayer, Cumulative prospect theory and mean variance analysis: A rigorous comparison, *Swiss Finance Institute Research Paper*, 2014. http://doi.org/10.2139/ssrn.2417191
32. C. Fulga, Portfolio optimization under loss aversion, *Eur. J. Oper. Res.*, **251** (2016), 310–322. http://doi.org/10.1016/j.ejor.2015.11.038

33. J. Bi, H. Jin, Q. Meng, Behavioural mean-variance portfolio selection, *Eur. J. Oper. Res.*, **271** (2018), 644–663. http://doi.org/10.1016/j.ejor.2018.05.065
34. M. Kaucic, F. Piccotto, G. Sbaiz, G. Valentinuz and others, Optimal portfolio with sustainable attitudes under cumulative prospect theory, *J. Appl. Finance Bank.*, **13** (2023), 65–86. http://doi.org/10.47260/jafb/1344
35. J. Wei, Y. Yang, M. Jiang, J. Liu, Dynamic multi-period sparse portfolio selection model with asymmetric investors' sentiments, *Expert Syst. Appl.*, **177** (2021), 114945. http://doi.org/10.1016/j.eswa.2021.114945
36. Z. Wu, L. Yang, V. De Simone, Behavioral portfolio optimization via cumulative prospect theory with a symmetric alternating direction method of multipliers, *Comput. Optim. Appl.*, **93** (2025), 303–334. http://doi.org/10.1007/s10589-025-00721-9
37. K. Blay, A. Gosh, S. Kusiak, H. Markowitz, N. Savoulides, Q. Zheng, Multiperiod portfolio selection: A practical simulation-based framework, *J. Invest. Manag.*, **18** (2020), 94–129.
38. G. N. Raj, Adaptive and Regime-Aware RL for Portfolio Optimization, arXiv:2509.14385, 2025. http://doi.org/10.48550/arXiv.2509.14385
39. S. Corsaro, V. De Simone, Z. Marino, Split Bregman iteration for multi-period mean variance portfolio optimization, *Appl. Math. Comput.*, **392** (2021), 125715. http://doi.org/10.1016/j.amc.2020.125715
40. S. Boyd, N. Parikh, E. Chu, B. Peleato, J. Eckstein, Distributed optimization and statistical learning via the alternating direction method of multipliers, *Found. Trends Mach. Learn.*, **3** (2011), 1–122.
41. K. Guo, D. Han, T. Wu, Convergence of alternating direction method for minimizing sum of two nonconvex functions with linear constraints, *Int. J. Comput. Math.*, **94** (2017), 1653–1669. http://doi.org/10.1080/00207160.2016.1227432
42. D. Han, A survey on some recent developments of alternating direction method of multipliers, *J. Oper. Res. Soc. China*, **10** (2022), 1–52. http://doi.org/10.1007/s40305-021-00368-3
43. M. Ayer, H. D. Brunk, G. M. Ewing, W. T. Reid, E. Silverman, An empirical distribution function for sampling with incomplete information, *Ann. Math. Stat.*, **26** (1955), 641–647. http://doi.org/10.1214/aoms/1177728423
44. H. D. Brunk, R. E. Barlow, D. J. Bartholomew, J. M. Bremner, Statistical inference under order restrictions the theory and application of isotonic regression, *Int. Stat. Rev.*, **41** (1972), 195.
45. R. K. Ahuja, J. B. Orlin, A fast scaling algorithm for minimizing separable convex functions subject to chain constraints, *Oper. Res.*, **49** (2001), 784–789. http://doi.org/10.1287/opre.49.5.784.10601
46. Z. Cui, C. Lee, L. Zhu, Y. Zhu, Non-convex isotonic regression via the Myersonian approach, *Stat. Probab. Lett.*, **179** (2021), 109210. http://doi.org/10.1016/j.spl.2021.109210
47. Z. Yu, X. Chen, X. Li, A dynamic programming approach for generalized nearly isotonic optimization, *Math. Program. Comput.*, **15** (2023), 195–225. http://doi.org/10.1007/s12532-022-00229-x
48. X. Cui, R. Jiang, Y. Shi, R. Xiao, Y. Yan, Decision making under cumulative prospect theory: An alternating direction method of multipliers, *INFORMS J. Comput.*, **37** (2024), 856–873. http://doi.org/10.1287/ijoc.2023.0243

49. M. J. Best, N. Chakravarti, V. A. Ubhaya, Minimizing separable convex functions subject to simple chain constraints, *SIAM J. Optim.*, **10** (2000), 658–672. http://doi.org/10.1137/S1052623497314970
50. R. T. Rockafellar, R. J. B. Wets, Variational Analysis, In: *Grundlehren der mathematischen Wissenschaften,* Springer-Verlag, Berlin, 1998. http://doi.org/10.1007/978-3-642-02431-3
51. F. H. Clarke, *Optimization and Nonsmooth Analysis*, SIAM, Philadelphia, PA, 1990. http://doi.org/10.1137/1.9781611971309
52. 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. http://doi.org/10.1016/j.ejor.2024.07.017
53. H. Attouch, J. Bolte, B. F. Svaiter, Convergence of descent methods for semi-algebraic and tame problems: proximal algorithms, forward–backward splitting, and regularized Gauss–Seidel methods, *Math. Program.*, **137** (2013), 91–129. http://doi.org/10.1007/s10107-011-0484-9
54. J. Bolte, S. Sabach, M. Teboulle, Proximal alternating linearized minimization for nonconvex and nonsmooth problems, *Math. Program.*, **146** (2014), 459–494. http://doi.org/10.1007/s10107-013-0701-9
55. Z. Wu, M. Li, D. Wang, D. Han, A symmetric alternating direction method of multipliers for separable nonconvex minimization problems, *Asia-Pac. J. Oper. Res.*, **34** (2017), 1750030. http://doi.org/10.1142/S0217595917500300
56. S. Utz, M. Wimmer, R. Steuer, Tri-criterion modeling for constructing more-sustainable mutual funds, *Eur. J. Oper. Res.*, **246** (2015), 331–338. http://doi.org/10.1016/j.ejor.2015.04.035
57. L. H. Pedersen, S. Fitzgibbons, L. Pomorski, Responsible investing: The ESG-efficient frontier, *J. Financ. Econ.*, **142** (2021), 572–597. http://doi.org/10.1016/j.jfineco.2020.11.001
58. Z. Wu, L. Yang, Y. Fei, X. Wang, Regularization methods for sparse ESG-valued multi-period portfolio optimization with return prediction using machine learning, *Expert Syst. Appl.*, **232** (2023), 120850. http://doi.org/10.1016/j.eswa.2023.120850
59. D. Lauria, W. Brent Lindquist, S. Mittnik, Svetlozar T. Rachev, ESG-valued portfolio optimization and dynamic asset pricing, arXiv:2206.02854, 2022. http://doi.org/10.48550/arXiv.2206.02854

AIMS Press