Journal of Industrial and Management Optimization

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

JIMO, 22(11): 6003–6033.
DOI:10.3934/jimo.2026202
Received: 07 August 2026
Revised: 22 October 2026
Accepted: 23 October 2026
Published: 27 October 2026

Research article

Mean-constrained portfolio optimization under lambda expected shortfall: A loss-level decomposition approach

Kunpeng Zhu and Yanjing Guo*

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

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

Abstract: Conventional expected shortfall portfolio optimization requires an externally specified confidence level, which can make optimized portfolios sensitive to the chosen confidence level. We developed a mean-constrained lambda expected shortfall portfolio model that replaces the fixed confidence level with a loss-dependent confidence function, together with an ℓ_2-regularized extension intended to stabilize portfolio weights. The resulting joint empirical formulations are generally nonlinear and nonconvex. We addressed this difficulty through a loss-level decomposition that reduces the original problems to convex fixed-confidence expected shortfall subproblems and a one-dimensional outer optimization over the loss level. For the basic model, the monotone-crossing structure yields a bisection procedure. We also developed an interval branch-and-bound procedure applicable to both models, including the regularized case in which the monotone-crossing structure is not guaranteed. Numerical experiments on four equity datasets showed that the proposed models endogenously select effective confidence levels and achieve competitive out-of-sample performance relative to fixed-confidence benchmarks, with the regularized extension further reducing portfolio turnover.

Keywords: lambda expected shortfall; conditional value-at-risk; portfolio optimization; loss-level decomposition; interval branch-and-bound
Mathematics Subject Classification: Primary 90C26; Secondary 90C25, 90C90, 91B30, 91G10

1. Introduction

Portfolio selection balances expected return against risk. In Markowitz's mean–variance model [1], risk is measured by the variance of returns. Expected shortfall (ES), also called conditional value-at-risk (CVaR), instead averages losses in the upper $1-\alpha$ tail at confidence level α and is a coherent risk measure [2, 3]. CVaR is also used in reinsurance with premium budgets and reinsurer risk limits [4],

and in mean-CVaR newsvendor and supply-chain models [5]. The Rockafellar–Uryasev representation makes ES minimization tractable under common convex portfolio constraints [6–8]. The portfolio formulation still requires α to be chosen before optimization.

The confidence level α determines both the losses emphasized by ES and the number of tail observations available for its estimation. As α rises, ES focuses on more severe losses, while fewer observations remain in the tail of a finite sample. The estimated ES and the resulting portfolio can therefore be sensitive to sampling variation [9–11]. A fixed α applies the same tail probability in every ES evaluation, regardless of the loss level being assessed. Previous work has examined the role of the confidence level in CVaR optimization [12] and distributional uncertainty in CVaR portfolio models [13]. Here, we allow the confidence level used by ES to depend on the loss level at which risk is evaluated.

Lambda risk measures represent this dependence through a confidence function $\Lambda(\cdot)$. Frittelli et al. [14] introduced this approach for lambda value-at-risk (Λ-VaR) models. Subsequent studies examined Λ-VaR in risk measurement and backtesting [15, 16]. Bellini et al. [17] extended the construction to expected shortfall. Let $X \in L^\infty$ denote a loss, with larger values indicating more severe outcomes, and let $\Lambda : \mathbb{R} \to [0, 1)$ be decreasing. Throughout this paper, decreasing means nonincreasing. Lambda expected shortfall (Λ-ES) is defined by

$$\mathrm{ES}^{\Lambda}(X) = \sup_{x \in \mathbb{R}} \{\mathrm{ES}_{\Lambda(x)}(X) \wedge x\}, \tag{1.1}$$

where $a \wedge b = \min\{a, b\}$. At each loss level x, $\Lambda(x)$ is the confidence level used in $\mathrm{ES}_{\Lambda(x)}(X)$. When $\Lambda(x) \equiv \alpha$, the definition reduces to ordinary ES_α. Otherwise, the confidence level varies with the loss level. In the portfolio model developed below, Λ is calibrated before optimization. The optimized loss level x^* obtained from an equivalent minimization representation determines the effective confidence level $\Lambda(x^*)$.

Bellini et al. [17] established the properties and optimization representations of Λ-ES, including cases where the loss depends on a decision. In related Λ-VaR work, Boonen et al. [18] optimized insurance decisions at fixed loss levels. Peri and Wunderlich [19] developed numerical methods for lambda quantiles and applied them to portfolio optimization. Our setting combines the Λ-ES representation with empirical losses determined by portfolio weights, a minimum expected return constraint, and a confidence function calibrated to the feasible loss range. Portfolio norm constraints and regularization have been used to address estimation error and weight concentration [20, 21]. For ES portfolios, Papp et al. [22] studied the tradeoff between bias and variance under ℓ_2 regularization. In our model, the ℓ_2 penalty preserves convexity of each portfolio subproblem at a fixed loss level but changes the objective minimized over loss levels. This change affects the scalar outer optimization studied below.

We study long-only, fully invested portfolios subject to a minimum expected return constraint. The mean-constrained lambda expected shortfall (MΛES) model minimizes Λ-ES over this feasible set. The ℓ_2-regularized mean-constrained lambda expected shortfall (MΛES-ℓ_2) model adds a squared ℓ_2-norm penalty to discourage concentrated weights. We assess its effect on portfolio stability empirically. For each estimation window, we calibrate a logistic confidence function using the feasible loss interval $[x_L, x_U]$. The endpoint confidence levels $\alpha_{\min}$ and $\alpha_{\max}$ are specified in advance, while the returns observed in that window and the feasible portfolio set determine the interval. This calibration places

the prescribed confidence range on the loss scale of the current window. The optimized loss level x^* then yields the effective confidence level $\Lambda(x^*)$.

The main computational difficulty lies in optimizing the loss level and portfolio weights jointly. In the empirical Rockafellar–Uryasev formulation, the fixed coefficient $1/(1-\alpha)$ becomes $q(x) = 1/(1-\Lambda(x))$. When x varies, $q(x)$ multiplies tail excess variables determined by the portfolio, making the joint formulation nonlinear and generally nonconvex. At a fixed x, $\Lambda(x)$ is constant. The portfolio subproblem is then a linear program for MΛES and a convex quadratic program for MΛES-ℓ_2. The loss-level decomposition reduces the remaining task to a scalar search over x.

For MΛES, the optimal value of the portfolio subproblem is nonincreasing in x, while x itself increases. This monotone crossing supports bisection. The ℓ_2 penalty can remove this crossing, so the bisection guarantee does not extend to the regularized model. We therefore derive an interval relaxation and use it in an interval branch-and-bound procedure [23, 24]. In exact arithmetic, the relaxation gives a valid lower bound on each interval and becomes tight as the interval shrinks. In the numerical implementation, conservative objective evaluations of checked feasible portfolios provide upper bounds, while verified interval bounds give a global lower bound. A certified objective gap is reported only when the difference between these bounds meets the stopping tolerance. Adaptive loss-level discretization provides a finite-grid benchmark.

Our main contributions can be summarized as follows.

1. We formulate MΛES and MΛES-ℓ_2 for portfolios with a minimum expected return. In each estimation window, we calibrate the confidence function to the feasible loss interval using prescribed endpoint levels. The optimized loss level determines the effective confidence level.
2. We specialize the Λ-ES representation to empirical portfolio losses, establish a compact outer interval, and prove attainment and monotone crossing for MΛES. We show that an additive decision term preserves the decomposition, while an ℓ_2 counterexample shows that it need not preserve crossing.
3. We construct an interval relaxation for both models and prove that its exact bound lies within the interval width of the interval minimum. Verified lower bounds and feasible portfolios allow interval branch-and-bound to certify the objective gap.
4. Against fixed-confidence and standard portfolio benchmarks, MΛES achieves the highest cumulative wealth and Sharpe ratio in most of the four datasets. Its regularized extension reduces turnover in every dataset and can improve terminal wealth when transaction costs are included.

The remainder of the paper is organized as follows. Section 2 presents the optimization representation of Λ-ES, its decision-dependent reformulation, and the scalar outer problems. Section 3 develops the two mean-constrained portfolio models and the calibration of the confidence function. Section 4 gives the bisection, adaptive-grid, and relaxation-guided interval procedures, together with the exact relaxation results and finite-precision implementation. Section 5 reports the numerical experiments and rolling-window portfolio tests, and Section 6 concludes the paper. Proof details, algorithm pseudocode, and additional computational diagnostics are provided in the supplementary materials.

2. Lambda expected shortfall: optimization structure

This section recalls the optimization representation of Λ-ES and its attainment conditions. It then applies the representation to decision-dependent losses and derives the scalar outer problem.

2.1. Optimization representation of Λ-ES

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, and let L^∞ denote the space of essentially bounded random variables. We consider a loss $X \in L^\infty$, where larger values represent more severe losses. This setting includes the empirical portfolio losses introduced in Section 3.1. Throughout, "decreasing" means nonincreasing. We use "strictly decreasing" when the stronger condition is required.

Bellini et al. [17] established the representations for decreasing $\Lambda : \mathbb{R} \to [0, 1]$. We use the narrower range $[0, 1)$ throughout. This keeps the coefficient $1/(1 - \Lambda(x))$ finite in the formulas below. Recall from (1.1) that

$$\mathrm{ES}^\Lambda(X) = \sup_{x \in \mathbb{R}} \{\mathrm{ES}_{\Lambda(x)}(X) \wedge x\}.$$

For a fixed confidence level $\alpha \in [0, 1)$, the Rockafellar–Uryasev representation [6, 7] is

$$\mathrm{ES}_\alpha(X) = \inf_{a \in \mathbb{R}} \left\{ a + \frac{1}{1-\alpha} \mathbb{E}[(X-a)^+] \right\}.$$

Bellini et al. [17, Theorem 2, Eq. (10)] gave an equivalent infimum representation for decreasing Λ. Combining it with the Rockafellar–Uryasev formula yields

$$\begin{aligned} \mathrm{ES}^\Lambda(X) &= \inf_{x \in \mathbb{R}} \{\mathrm{ES}_{\Lambda(x)}(X) \vee x\} \\ &= \inf_{(a,x) \in \mathbb{R}^2} \left\{ \left[a + \frac{1}{1-\Lambda(x)} \mathbb{E}[(X-a)^+] \right] \vee x \right\}. \end{aligned} \tag{2.1}$$

Here, $u \wedge v = \min\{u, v\}$, $u \vee v = \max\{u, v\}$, and $(y)^+ = \max\{y, 0\}$. In (2.1), a is the threshold in the Rockafellar–Uryasev formula. The variable x indexes the outer loss-level optimization.

The equalities in (2.1) hold without continuity of Λ. Bellini et al. [17, Remark 2] showed that right-continuity ensures attainment of the outer infimum in (2.1), while left-continuity ensures attainment of the supremum in (1.1). The logistic confidence functions used in Section 3.1 are continuous and satisfy both conditions. For these functions, the joint infimum over (a, x) in (2.1) is also attained [17, Theorem 4].

For each fixed x, minimizing the Rockafellar–Uryasev term over a gives $\mathrm{ES}_{\Lambda(x)}(X)$. The outer problem then minimizes $\mathrm{ES}_{\Lambda(x)}(X) \vee x$ over x. This nested structure is used in the decision-dependent reformulation below.

2.2. Decision-dependent loss-level decomposition

Let Θ be a nonempty feasible set, and suppose that each decision $\theta \in \Theta$ induces a bounded loss variable $X_\theta \in L^\infty$. We consider the decision-dependent Λ-ES problem

$$\inf_{\theta \in \Theta} \mathrm{ES}^\Lambda(X_\theta). \tag{2.2}$$

By *loss-level decomposition*, we mean the exact reformulation of (2.2) as a scalar outer problem over the loss level x. At each fixed x, the inner problem takes the infimum of $\mathrm{ES}_{\Lambda(x)}(X_\theta)$ over $\theta \in \Theta$, with $\Lambda(x)$ held constant. The outer problem minimizes the maximum of x and this inner infimal value. The following proposition establishes this identity.

Proposition 1. *Let Θ be nonempty. Let $\Lambda : \mathbb{R} \to [0, 1)$ be decreasing, and suppose that $X_\theta \in L^\infty$ for every $\theta \in \Theta$. Then*

$$\inf_{\theta\in\Theta} \mathrm{ES}^{\Lambda}(X_\theta) = \inf_{x\in\mathbb{R}} \left\{ \left(\inf_{\theta\in\Theta} \mathrm{ES}_{\Lambda(x)}(X_\theta) \right) \vee x \right\}. \tag{2.3}$$

Proof. Applying (2.1) pointwise and using the fact that both iterated infima equal the infimum of the same objective over $\Theta \times \mathbb{R}$, we obtain

$$\begin{aligned} \inf_{\theta\in\Theta} \mathrm{ES}^{\Lambda}(X_\theta) &= \inf_{\theta\in\Theta} \inf_{x\in\mathbb{R}} \{\mathrm{ES}_{\Lambda(x)}(X_\theta) \vee x\} \\ &= \inf_{x\in\mathbb{R}} \inf_{\theta\in\Theta} \{\mathrm{ES}_{\Lambda(x)}(X_\theta) \vee x\}. \end{aligned}$$

For a fixed loss level x, define

$$m(x) = \inf_{\theta\in\Theta} \mathrm{ES}_{\Lambda(x)}(X_\theta).$$

If $m(x)$ is finite, the monotonicity and continuity of $t \mapsto t \vee x$ give

$$\inf_{\theta\in\Theta} \{\mathrm{ES}_{\Lambda(x)}(X_\theta) \vee x\} = m(x) \vee x.$$

If $m(x) = -\infty$, choose a sequence $(\theta_n) \subseteq \Theta$ such that $\mathrm{ES}_{\Lambda(x)}(X_{\theta_n}) \to -\infty$. The corresponding maximum terms equal x for all sufficiently large n, while every such term is bounded below by x. Hence, the same identity holds under the convention $(-\infty) \vee x = x$. Substitution into the preceding expression proves (2.3). □

For each loss level x, define the fixed-confidence value function

$$G(x) = \inf_{\theta\in\Theta} \mathrm{ES}_{\Lambda(x)}(X_\theta). \tag{2.4}$$

With this notation, Proposition 1 becomes

$$\inf_{\theta\in\Theta} \mathrm{ES}^{\Lambda}(X_\theta) = \inf_{x\in\mathbb{R}} \{G(x) \vee x\}. \tag{2.5}$$

For a fixed loss level x, the Rockafellar–Uryasev representation of ES at confidence level $\Lambda(x)$ gives

$$G(x) = \inf_{\theta\in\Theta,\, a\in\mathbb{R}} \left\{ a + \frac{1}{1-\Lambda(x)} \mathbb{E}[(X_\theta - a)^+] \right\}. \tag{2.6}$$

The loss-level decomposition is an equality of infimal values. Beyond the assumptions in Proposition 1, it requires no continuity of Λ, convexity or compactness of Θ, or attainment of the infima. In the basic portfolio model, the inner problem is convex because fixed-confidence ES is convex, the empirical loss is affine in the portfolio weights, and the feasible set is convex. Section 3 establishes optimizer existence under its application-specific assumptions.

To recover an optimizer, the outer infimum must be attained at some x^*, and the inner infimum $G(x^*)$ must be attained at some θ^*. For each fixed X_θ, right-continuity of Λ ensures attainment of the outer infimum in (2.1). It does not ensure attainment after optimization over θ. The following corollary states the required conditions explicitly.

Corollary 1. *Assume that the outer infimum in* (2.5) *is attained at* x^* *and that the infimum defining* $G(x^*)$ *is attained at* $\theta^* \in \Theta$*. Then* θ^* *is optimal for* (2.2)*.*

Proof. Since θ^* attains $G(x^*)$, we have $\mathrm{ES}_{\Lambda(x^*)}(X_{\theta^*}) = G(x^*)$. By (2.1) and the attainment of the outer infimum in (2.5),

$$\begin{aligned}\mathrm{ES}^{\Lambda}(X_{\theta^*}) &\le \mathrm{ES}_{\Lambda(x^*)}(X_{\theta^*}) \vee x^* \\ &= G(x^*) \vee x^* \\ &= \inf_{\theta\in\Theta} \mathrm{ES}^{\Lambda}(X_\theta).\end{aligned}$$

Since $\theta^* \in \Theta$, the reverse inequality also holds. Hence, θ^* is optimal for (2.2). □

2.3. Scalar outer problems

With G denoting the fixed-confidence value function in (2.4), the scalar outer problem (2.5) takes the form

$$\inf_{x\in\mathbb{R}} H(x), \qquad H(x) = G(x) \vee x. \tag{2.7}$$

The next proposition shows that G is nonincreasing and H is quasiconvex.

Proposition 2. *Let* G *and* H *be defined by* (2.4) *and* (2.7)*, respectively. Let*

$$D(x) = G(x) - x.$$

Suppose that G *is finite-valued on* $\mathbb{R}$*. If* Λ *is decreasing, then* G *is nonincreasing,* D *is strictly decreasing, and* H *is quasiconvex on* $\mathbb{R}$*.*

Proof. Let $x_1 < x_2$. Since Λ is decreasing and the fixed-confidence ES functional is nondecreasing in the confidence level, for every $\theta \in \Theta$,

$$\mathrm{ES}_{\Lambda(x_1)}(X_\theta) \ge \mathrm{ES}_{\Lambda(x_2)}(X_\theta).$$

Taking the infimum over $\theta \in \Theta$ gives $G(x_1) \ge G(x_2)$, so G is nonincreasing. Moreover,

$$D(x_1) - D(x_2) = G(x_1) - G(x_2) + x_2 - x_1 > 0,$$

and hence D is strictly decreasing.

For any $c \in \mathbb{R}$,

$$\{x \in \mathbb{R} : H(x) \le c\} = \{x \in \mathbb{R} : G(x) \le c\} \cap \{x \in \mathbb{R} : x \le c\}.$$

Because G is nonincreasing, the first set is an upper interval and the second is a lower interval. Their intersection is therefore either empty or an interval, so H is quasiconvex. □

The difference function D identifies the active term in the outer objective:

$$D(x) \ge 0 \iff H(x) = G(x), \qquad D(x) \le 0 \iff H(x) = x.$$

Since D is strictly decreasing, its sign can change at most once, and any solution to $G(x) = x$ is unique. On a bounded interval $[x_L, x_U]$, x_L minimizes H if $D(x_L) \le 0$. Likewise, x_U minimizes H if $D(x_U) \ge 0$.

Otherwise, $D(x_L) > 0 > D(x_U)$, and the signs determine a single boundary between the two regimes. If G is continuous on $[x_L, x_U]$, this boundary is the unique solution of $G(x) = x$ and a minimizer of H. Bisection locates the crossing under this continuity condition, which holds for the basic portfolio model in Section 3.

The monotone-crossing argument uses the form $H(x) = G(x) \vee x$. An additive decision term $\Phi(\theta)$ that is independent of x preserves the loss-level decomposition, but it can change the outer value function. The next proposition establishes the decomposition for this case.

Proposition 3. *Let Θ be a nonempty feasible set, and let $\Lambda : \mathbb{R} \to [0, 1)$ be decreasing. For each decision $\theta \in \Theta$, let $X_\theta \in L^\infty$ denote the induced loss. Let $\Phi : \Theta \to \mathbb{R} \cup \{+\infty\}$ be an extended real-valued function that depends on θ but not on the loss level x. Then*

$$\begin{aligned}\inf_{\theta\in\Theta}\left\{\mathrm{ES}^{\Lambda}(X_\theta) + \Phi(\theta)\right\} &= \inf_{\theta\in\Theta}\inf_{x\in\mathbb{R}}\left\{\left(\mathrm{ES}_{\Lambda(x)}(X_\theta) \vee x\right) + \Phi(\theta)\right\} \\ &= \inf_{x\in\mathbb{R}}\inf_{\theta\in\Theta}\left\{\left(\mathrm{ES}_{\Lambda(x)}(X_\theta) \vee x\right) + \Phi(\theta)\right\}. \end{aligned} \tag{2.8}$$

For each fixed loss level x, define

$$V_\Phi(x) = \inf_{\theta\in\Theta}\left\{\left(\mathrm{ES}_{\Lambda(x)}(X_\theta) \vee x\right) + \Phi(\theta)\right\}. \tag{2.9}$$

The additive problem therefore has the scalar outer representation

$$\inf_{x\in\mathbb{R}} V_\Phi(x). \tag{2.10}$$

Proof. Because $\Phi(\theta)$ is independent of x and does not take the value $-\infty$, (2.1) gives

$$\mathrm{ES}^{\Lambda}(X_\theta) + \Phi(\theta) = \inf_{x\in\mathbb{R}}\left\{\left(\mathrm{ES}_{\Lambda(x)}(X_\theta) \vee x\right) + \Phi(\theta)\right\}.$$

Taking the infimum over $\theta \in \Theta$ gives the first equality in (2.8). The second follows by exchanging the two infima, since both expressions equal the infimum of the same extended real-valued objective over $\Theta \times \mathbb{R}$. Equation (2.10) follows from the definition of V_Φ. □

Remark 1. When an additive objective term is independent of the loss level, Proposition 3 shows that the loss-level decomposition remains valid, with the additive term retained in each fixed-level subproblem. In the portfolio application, the ℓ_2 regularization term is

$$\Phi(\boldsymbol{w}) = \frac{\gamma}{2}\|\boldsymbol{w}\|_2^2.$$

Terms or constraints that depend explicitly on x are not covered by Proposition 3 and require separate analysis.

Preservation of the loss-level decomposition does not imply preservation of the monotone-crossing structure. In particular, the identity

$$\inf_{\theta\in\Theta}\left\{\left(\mathrm{ES}_{\Lambda(x)}(X_\theta) \vee x\right) + \Phi(\theta)\right\} = \left[\inf_{\theta\in\Theta}\left\{\mathrm{ES}_{\Lambda(x)}(X_\theta) + \Phi(\theta)\right\}\right] \vee x$$

need not hold. Thus, the right-hand side cannot generally be used as a crossing representation of V_Φ. Proposition 4 gives a regularized portfolio instance for which no nonincreasing $\widetilde{G}$ satisfies $V_\Phi(x) = \widetilde{G}(x) \vee x$.

3. Mean-constrained Λ-ES portfolio optimization

This section applies the loss-level decomposition from Section 2 to empirical mean-constrained portfolios. We define the loss-level interval and calibrate the logistic confidence function before formulating the MΛES and MΛES-ℓ_2 models. We then establish optimizer existence and identify the source of nonconvexity in their joint formulations.

3.1. Empirical portfolio setting

In this setting, the decision θ and loss X_θ from Section 2.2 correspond to portfolio weights and their scenario losses. Let $\boldsymbol{r}_t \in \mathbb{R}^N$ be the return vector in scenario t, and let $\boldsymbol{R} \in \mathbb{R}^{T\times N}$ have row t equal to $\boldsymbol{r}_t^\top$. For portfolio weights $\boldsymbol{w} \in \mathbb{R}^N$, the return vector is $\boldsymbol{Rw}$, and the empirical loss vector is

$$L(\boldsymbol{w}) = -\boldsymbol{Rw}.$$

Let

$$\ell_t(\boldsymbol{w}) = -\boldsymbol{r}_t^\top \boldsymbol{w}, \qquad t = 1, \ldots, T,$$

denote the scenario losses. Throughout the paper, empirical ES is evaluated under the uniform distribution over these T observations. In particular, for $\alpha \in [0, 1)$,

$$\mathrm{ES}_\alpha(L(\boldsymbol{w})) = \min_{a\in\mathbb{R}} \left\{ a + \frac{1}{(1-\alpha)T} \sum_{t=1}^{T} (\ell_t(\boldsymbol{w}) - a)^+ \right\}.$$

For $\alpha \in (0, 1)$, we use the left empirical quantile convention

$$\mathrm{VaR}_\alpha(L(\boldsymbol{w})) = \inf \left\{ y : \frac{1}{T} \sum_{t=1}^{T} \mathbf{1}_{\{\ell_t(\boldsymbol{w})\le y\}} \ge \alpha \right\}.$$

The Rockafellar–Uryasev representation assigns the required probability mass at a tied VaR threshold without ordering the tied observations.

We consider the standard long-only, fully invested portfolio setting with a minimum expected return constraint [8, 25]. With expected returns estimated by the sample mean vector $\hat{\boldsymbol{\mu}} = T^{-1}\boldsymbol{R}^\top \mathbf{1}_T$, the feasible investment set is

$$\mathcal{W}_\eta = \left\{ \boldsymbol{w} \in \mathbb{R}^N : \mathbf{1}^\top \boldsymbol{w} = 1,\ \boldsymbol{w} \ge \mathbf{0},\ \boldsymbol{w}^\top \hat{\boldsymbol{\mu}} \ge \eta \right\},$$

where η is the minimum expected return specified before solving the portfolio problem. We assume that $\mathcal{W}_\eta$ is nonempty. Under the long-only and fully invested constraints, this is equivalent to

$$\eta \le \max_{i=1,\ldots,N} \hat{\mu}_i.$$

As a closed subset of the unit simplex, $\mathcal{W}_\eta$ is compact. This compactness ensures attainment of the loss-interval endpoints below and is also used later to establish existence of portfolio optimizers.

The scenario losses generated by feasible portfolios form the attainable loss set

$$\mathcal{L}(\mathcal{W}_\eta) = \left\{ -\boldsymbol{r}_t^\top \boldsymbol{w} : t = 1, \ldots, T,\ \boldsymbol{w} \in \mathcal{W}_\eta \right\}.$$

Its extrema define the loss-level interval, whose endpoints are

$$x_L = \min_{t=1,\dots,T} \min_{\boldsymbol{w}\in\mathcal{W}_\eta} \{-\boldsymbol{r}_t^\top \boldsymbol{w}\}, \qquad x_U = \max_{t=1,\dots,T} \max_{\boldsymbol{w}\in\mathcal{W}_\eta} \{-\boldsymbol{r}_t^\top \boldsymbol{w}\}. \tag{3.1}$$

Equivalently,

$$[x_L, x_U] = \Big[\min \mathcal{L}(\mathcal{W}_\eta), \max \mathcal{L}(\mathcal{W}_\eta)\Big].$$

Compactness of $\mathcal{W}_\eta$ ensures that both endpoints are finite and attained. For each scenario, the inner minimization or maximization in (3.1) is a linear program, so constructing the interval requires at most $2T$ such problems. In the rolling-window implementation, this construction is treated as a preprocessing step that precedes calibration of Λ and finding the solution of the scalar outer problem. Because all scenario losses generated by $\boldsymbol{w} \in \mathcal{W}_\eta$ lie in $[x_L, x_U]$,

$$x_L \le \mathrm{ES}_\alpha(L(\boldsymbol{w})) \le x_U, \qquad \alpha \in [0, 1).$$

For any decreasing confidence function Λ, these bounds allow us to restrict the scalar outer problem to $x \in [x_L, x_U]$. For any $\boldsymbol{w} \in \mathcal{W}_\eta$, if $x < x_L$, then

$$\begin{aligned} \mathrm{ES}_{\Lambda(x)}(L(\boldsymbol{w})) &\ge \mathrm{ES}_{\Lambda(x_L)}(L(\boldsymbol{w})) \ge x_L > x, \\ \mathrm{ES}_{\Lambda(x)}(L(\boldsymbol{w})) \vee x &\ge \mathrm{ES}_{\Lambda(x_L)}(L(\boldsymbol{w})) \vee x_L. \end{aligned}$$

If $x > x_U$, every empirical ES value is at most x_U, and therefore

$$\mathrm{ES}_{\Lambda(x)}(L(\boldsymbol{w})) \vee x = x > x_U = \mathrm{ES}_{\Lambda(x_U)}(L(\boldsymbol{w})) \vee x_U.$$

Thus, no loss level outside $[x_L, x_U]$ can improve the outer objective. The same comparisons remain valid after adding an objective term that depends on $\boldsymbol{w}$ but not on x.

For the empirical model, we use the logistic confidence function considered in the Λ-ES literature [17]:

$$\Lambda(x) = \frac{1}{1 + \exp\{\beta(x - c)\}}, \qquad \beta > 0. \tag{3.2}$$

This function is continuous and strictly decreasing, with values in $(0, 1)$.

If $x_L = x_U$, the scalar outer domain contains only the loss level x_L, and only the value $\Lambda(x_L)$ is relevant. For a prescribed $\bar{\alpha} \in (0, 1)$, the condition $\Lambda(x_L) = \bar{\alpha}$ can be imposed by choosing any $\beta > 0$ and setting

$$c = x_L + \beta^{-1} \log\left(\frac{\bar{\alpha}}{1 - \bar{\alpha}}\right).$$

Suppose now that $x_L < x_U$. We choose two endpoint confidence levels satisfying

$$0 < \alpha_{\min} < \alpha_{\max} < 1$$

and impose

$$\Lambda(x_L) = \alpha_{\max}, \qquad \Lambda(x_U) = \alpha_{\min}.$$

The resulting parameters are

$$\begin{aligned} \beta &= \frac{\mathrm{logit}(\alpha_{\max}) - \mathrm{logit}(\alpha_{\min})}{x_U - x_L}, \\ c &= x_L + \beta^{-1} \mathrm{logit}(\alpha_{\max}), \end{aligned} \qquad \mathrm{logit}(p) = \log\left(\frac{p}{1 - p}\right).$$

Because $\alpha_{\max} > \alpha_{\min}$, the calibrated value of β is positive.

The endpoint levels $(\alpha_{\min}, \alpha_{\max})$ are specified before portfolio optimization. The current return sample and feasible set determine $[x_L, x_U]$, which in turn determines β and c. Thus, each estimation window has its own calibrated logistic confidence function.

More explicitly, let

$$\tau(x) = \frac{x - x_L}{x_U - x_L}.$$

Then

$$\operatorname{logit}(\Lambda(x)) = (1 - \tau(x)) \operatorname{logit}(\alpha_{\max}) + \tau(x) \operatorname{logit}(\alpha_{\min}),$$

so the log-odds of the confidence level vary linearly with the normalized loss level.

The optimized loss level x^* determines the effective confidence level $\Lambda(x^*)$. In the rolling-window analysis, the notation suppresses the window index. Specifically, $[x_{L,k}, x_{U,k}]$, β_k, c_k, and Λ_k are recomputed using only information available in estimation window k.

3.2. Model formulation

Using the empirical loss $L(\boldsymbol{w})$, the feasible set $\mathcal{W}_\eta$, and the calibrated confidence function Λ, we consider

$$\min_{\boldsymbol{w} \in \mathcal{W}_\eta} \left\{ \mathrm{ES}^{\Lambda}(L(\boldsymbol{w})) + \frac{\gamma}{2} \|\boldsymbol{w}\|_2^2 \right\}, \qquad \gamma \geq 0. \tag{3.3}$$

When $\gamma = 0$, this is the mean-constrained Λ-ES model, denoted by MΛES. When $\gamma > 0$, it is the ℓ_2-regularized model, denoted by MΛES-ℓ_2.

The penalty acts directly on the portfolio weights. It discourages concentration without changing the feasible set $\mathcal{W}_\eta$, the loss-level interval, or the calibration of Λ [26, 27]. Its effect on portfolio stability and turnover is evaluated empirically rather than imposed as a theoretical conclusion.

Because the ℓ_2 regularization term is independent of the loss level, Proposition 3 applies with $\Theta = \mathcal{W}_\eta$, $\theta = \boldsymbol{w}$, $X_\theta = L(\boldsymbol{w})$, and $\Phi(\boldsymbol{w}) = (\gamma/2)\|\boldsymbol{w}\|_2^2$. Combining the resulting decomposition with the interval argument in Section 3.1 yields

$$\min_{x \in [x_L, x_U]} V_\gamma(x), \tag{3.4}$$

where

$$V_\gamma(x) = \min_{\boldsymbol{w} \in \mathcal{W}_\eta} \left\{ (\mathrm{ES}_{\Lambda(x)}(L(\boldsymbol{w})) \vee x) + \frac{\gamma}{2} \|\boldsymbol{w}\|_2^2 \right\}, \qquad x \in [x_L, x_U]. \tag{3.5}$$

For each $x \in [x_L, x_U]$, we refer to (3.5) as the *fixed-level subproblem*. It is a fixed-confidence ES portfolio problem evaluated at confidence level $\Lambda(x)$. Using the Rockafellar–Uryasev representation together with an epigraph variable for the maximum term, the fixed-level subproblem is a linear program when $\gamma = 0$ and a convex quadratic program when $\gamma > 0$.

We now verify the application-specific conditions that turn the abstract infima in Section 2 into minima. The feasible set $\mathcal{W}_\eta$ and the loss-level interval $[x_L, x_U]$ are compact, and the calibrated logistic function is continuous. Its image on this interval is

$$\Lambda([x_L, x_U]) = \begin{cases} [\alpha_{\min}, \alpha_{\max}], & x_L < x_U, \\ \{\bar{\alpha}\}, & x_L = x_U. \end{cases}$$

In either case, the image is a compact subset of $(0, 1)$. Moreover, the mapping

$$(\boldsymbol{w}, \alpha) \longmapsto \mathrm{ES}_\alpha(L(\boldsymbol{w}))$$

is continuous on $\mathcal{W}_\eta \times \Lambda([x_L, x_U])$. Hence, the objective defining $V_\gamma(x)$ is continuous in $(\boldsymbol{w}, x)$ on $\mathcal{W}_\eta \times [x_L, x_U]$.

It follows from Berge's maximum theorem that the minimum defining $V_\gamma(x)$ is attained for every x and that V_γ is continuous on $[x_L, x_U]$. Compactness of this interval then guarantees the existence of an outer minimizer x^*. If $\boldsymbol{w}^*$ attains $V_\gamma(x^*)$, the optimizer-recovery argument in Corollary 1, applied through Proposition 3, shows that $\boldsymbol{w}^*$ solves (3.3).

For $\gamma = 0$, specializing (2.4) to $\Theta = \mathcal{W}_\eta$ and $X_\theta = L(\boldsymbol{w})$ gives

$$V_0(x) = G_\eta(x) \vee x, \qquad G_\eta(x) = \min_{\boldsymbol{w} \in \mathcal{W}_\eta} \mathrm{ES}_{\Lambda(x)}(L(\boldsymbol{w})).$$

By Proposition 2, G_η is nonincreasing and

$$D_\eta(x) = G_\eta(x) - x$$

is strictly decreasing. This gives the monotone-crossing structure used by the bisection procedure. When $\gamma > 0$, the loss-level decomposition remains valid, but the representation $V_\gamma(x) = \widetilde{G}(x) \vee x$ with nonincreasing $\widetilde{G}$ need not hold. The following counterexample shows that the bisection argument cannot be transferred to the regularized model without additional assumptions.

Proposition 4. *There exists a two-asset, two-scenario instance of the MΛES-ℓ_2 model for which no nonincreasing function $\widetilde{G} : [x_L, x_U] \to \mathbb{R}$ satisfies*

$$V_\gamma(x) = \widetilde{G}(x) \vee x, \qquad x \in [x_L, x_U].$$

Thus, the monotone-crossing representation established for MΛES is not preserved in general under ℓ_2 regularization, and the associated bisection method is not guaranteed to solve MΛES-ℓ_2 without additional structural assumptions.

A constructive proof is given in Section A.1 of the supplementary materials. The result rules out a general bisection guarantee for MΛES-ℓ_2; it does not imply that every regularized instance lacks a crossing pattern.

Continuity and compactness guarantee the existence of an outer minimizer, but not its uniqueness. For MΛES, the equation $G_\eta(x) = x$ has at most one solution because D_η is strictly decreasing, although a flat boundary regime may still admit more than one minimizing loss level. For MΛES-ℓ_2, the outer objective need not have a monotone-crossing form, and multiple minimizers may occur. Accordingly, x^* denotes any outer minimizer unless uniqueness is stated explicitly. When evaluated objective values tie exactly, the implementation applies a fixed tie-breaking rule.

The loss-level decomposition avoids optimizing the loss level and the portfolio variables simultaneously. To see why this separation is useful, apply the Rockafellar–Uryasev representation to

(3.5) and collect all variables in the following joint formulation:

$$\begin{aligned}
\min_{x\in[x_L,x_U],\ \boldsymbol{w}\in\mathbb{R}^N,\ a\in\mathbb{R},\ z\in\mathbb{R},\ \boldsymbol{u}\in\mathbb{R}^T} \quad & z+\frac{\gamma}{2}\|\boldsymbol{w}\|_2^2 \\
\text{s.t.} \quad & z \ge x, \\
& z \ge a+\frac{1}{(1-\Lambda(x))T}\sum_{t=1}^{T} u_t, \\
& u_t \ge -\boldsymbol{r}_t^{\top}\boldsymbol{w}-a, \qquad t=1,\dots,T, \\
& u_t \ge 0, \qquad t=1,\dots,T, \\
& \boldsymbol{w}\in\mathcal{W}_\eta.
\end{aligned} \tag{3.6}$$

Here, z is the epigraph variable for the maximum term, a is the threshold variable in the fixed-confidence ES representation, and u_t is the tail excess variable for scenario t. Since z is minimized and each u_t enters the ES epigraph constraint with a positive coefficient, an optimal solution can be chosen to satisfy

$$u_t^* = \max\{-\boldsymbol{r}_t^{\top}\boldsymbol{w}^* - a^*, 0\}, \qquad t=1,\dots,T.$$

The joint reformulation in (3.6) makes the main computational difficulty explicit. Unlike the constant coefficient $1/(1-\alpha)$ in a standard fixed-confidence ES portfolio problem, the present model contains

$$q(x) = \frac{1}{1-\Lambda(x)},$$

which depends nonlinearly on x. Under the logistic specification in (3.2),

$$q(x) = 1+\exp\{-\beta(x-c)\}, \qquad q'(x) = -\beta\exp\{-\beta(x-c)\} \neq 0.$$

Let $s = T^{-1}\sum_{t=1}^{T} u_t \ge 0$. The loss-level-dependent term in the ES constraint is $f(x,s) = q(x)s$, whose Hessian satisfies

$$\nabla^2 f(x,s) = \begin{pmatrix} sq''(x) & q'(x) \\ q'(x) & 0 \end{pmatrix}, \qquad \det\Big(\nabla^2 f(x,s)\Big) = -\big(q'(x)\big)^2 < 0.$$

Because $q'(x) \neq 0$, the determinant is negative, and the Hessian is indefinite. Thus, $q(x)s$ is not jointly convex in the loss level and the aggregate tail excess. The corresponding epigraph constraint is therefore nonconvex when x and the portfolio variables are optimized jointly. For $x_L < x_U$, the joint formulation (3.6) is consequently nonlinear and generally nonconvex.

For a fixed loss level, however, $q(x)$ is constant and the remaining portfolio problem is convex. If $x_L = x_U$, this fixed-level case is the entire problem. The computational approach in Section 4 uses precisely this contrast: convex fixed-level subproblems are combined with a one-dimensional search over the continuous loss-level interval.

4. Algorithms based on loss-level decomposition

4.1. Solution methods for the scalar outer problem

The reformulation in (3.4) expresses both portfolio problems as the minimization of $V_\gamma(x)$ over $[x_L, x_U]$. At a fixed loss level, evaluating $V_\gamma(x)$ requires solving (3.5). This is a linear program when $\gamma = 0$ and a convex quadratic program when $\gamma > 0$. The outer search is therefore one-dimensional.

For MΛES ($\gamma = 0$), the portfolio specialization of G in (2.4) gives

$$G_\eta(x) = \min_{\boldsymbol{w}\in\mathcal{W}_\eta} \mathrm{ES}_{\Lambda(x)}(L(\boldsymbol{w})), \qquad V_0(x) = G_\eta(x) \vee x.$$

Proposition 2 shows that G_η is nonincreasing, while x increases. This monotone crossing permits the bisection procedure [28] given in Section A.2 of the supplementary materials. The ℓ_2 penalty preserves the loss-level decomposition but need not preserve this crossing, as Proposition 4 shows. We therefore use adaptive loss-level discretization and interval branch-and-bound for MΛES-ℓ_2. Both methods also apply to MΛES.

Adaptive loss-level discretization, described in Section A.3 of the supplementary materials, provides a finite-grid benchmark. It evaluates V_γ at selected loss levels, following grid-based derivative-free methods [29, 30]. A checked feasible portfolio at a sampled level gives an upper bound on the continuous minimum. The sampled levels alone provide no lower bound for the intervals between them.

To search the continuous interval without assuming monotone crossing, we use interval branch-and-bound. Checked feasible fixed-level solutions provide an incumbent upper bound. For each subinterval, the search combines an analytic bound, an inherited parent bound, and, when available, a verified dual bound from the convex interval relaxation. These safe bounds guide interval selection and screening [31–33]. Related interval and convex relaxation methods use the same bounding and partitioning principles [23, 24].

Section 4.2 establishes the exact relaxation and explains the safe bounds used in the numerical search. The algorithm reports a certified objective gap only when the global lower bound and feasible upper bound meet the stopping criterion. Algorithm 3 in the supplementary materials gives the implementation.

4.2. Interval relaxation and branch-and-bound

We construct an interval relaxation of V_γ on each loss-level subinterval. As established in Section 3, $\mathcal{W}_\eta$ is nonempty and compact, Λ is continuous and decreasing with values in $(0, 1)$ on $[x_L, x_U]$, and $\gamma \geq 0$. Hence V_γ is continuous and attains its minimum on $[x_L, x_U]$. The coefficient $q(x) = 1/(1-\Lambda(x))$ is finite and nonincreasing there.

Let $I = [l, b] \subseteq [x_L, x_U]$. For every $x \in I$, we have $x \geq l$ and $q(x) \geq q(b)$. Since $u_t \geq 0$, replacing $z \geq x$ by $z \geq l$ and $q(x)$ by $q(b)$ weakens the fixed-level epigraph constraints throughout I. Denote the resulting relaxation by $\mathsf{R}(I)$ and its exact optimal value by $L(I) = \mathrm{val}(\mathsf{R}(I))$. Explicitly,

$$
\begin{aligned}
L(I) = \min_{\boldsymbol{w}\in\mathcal{W}_\eta,\ a,z\in\mathbb{R},\ \boldsymbol{u}\in\mathbb{R}^T} \quad & z + \frac{\gamma}{2}\|\boldsymbol{w}\|_2^2 \\
\text{s.t.} \quad & z \geq l, \\
& z \geq a + \frac{q(b)}{T}\sum_{t=1}^{T} u_t, \\
& u_t \geq -\boldsymbol{r}_t^\top \boldsymbol{w} - a, \quad u_t \geq 0, \qquad t = 1, \ldots, T.
\end{aligned}
\tag{4.1}
$$

Lemma 1 (Interval lower bound)**.** *For any loss-level subinterval $I = [l, b] \subseteq [x_L, x_U]$, the value $L(I)$ defined in* (4.1) *satisfies*

$$
L(I) \leq \min_{x\in I} V_\gamma(x). \tag{4.2}
$$

Proof. Fix $x \in I$. The loss-level-dependent epigraph constraints in the fixed-loss-level subproblem defining $V_\gamma(x)$ are

$$
z \geq x, \qquad z \geq a + \frac{q(x)}{T}\sum_{t=1}^{T} u_t.
$$

Since $x \geq l$, $q(x) \geq q(b)$, and $u_t \geq 0$,

$$
a + \frac{q(b)}{T}\sum_{t=1}^{T} u_t \leq a + \frac{q(x)}{T}\sum_{t=1}^{T} u_t.
$$

Hence, every feasible solution of the fixed-loss-level subproblem is feasible for (4.1) with the same objective value, so $L(I) \leq V_\gamma(x)$. Taking the minimum over $x \in I$ proves (4.2). □

Lemma 1 concerns the exact relaxation value $L(I)$. A solver's computed primal objective $\widehat{L}(I)$ does not by itself certify a lower bound. We therefore do not use it to form the global lower bound or to screen and stop the search.

For $I = [l, b]$, let L_p be the safe bound inherited from its parent interval, with $L_p = -\infty$ at the root. Let $L_d(I)$ be a verified dual bound for $\mathsf{R}(I)$, and set $L_d(I) = -\infty$ if no verified dual bound is available. The search uses

$$
L_s(I) = \max\left\{ l + \frac{\gamma}{2N},\ L_p,\ L_d(I) \right\}.
$$

The analytic term follows from $z \geq l$ and $\|\boldsymbol{w}\|_2^2 \geq 1/N$ on the long-only unit simplex. Each child relaxation is tighter than its parent relaxation, so the inherited bound remains valid. By weak duality, a verified dual bound does not exceed $L(I)$. Consequently,

$$
L_s(I) \leq L(I) \leq \min_{x\in I} V_\gamma(x).
$$

Proposition 5. *For any interval $I = [l, b] \subseteq [x_L, x_U]$, let $L(I)$ denote the exact optimal value of the interval relaxation $\mathsf{R}(I)$. Then*

$$
L(I) \leq \min_{x\in I} V_\gamma(x) \leq L(I) + (b - l).
$$

Consequently,

$$0 \le \min_{x \in I} V_\gamma(x) - L(I) \le \operatorname{diam}(I),$$

and the exact interval relaxation becomes tight as the interval width converges to zero.

Proof. The first inequality follows from Lemma 1. For $\epsilon' > 0$, let $(\boldsymbol{w}_{\epsilon'}, a_{\epsilon'}, z_{\epsilon'}, \boldsymbol{u}_{\epsilon'})$ be an ϵ'-optimal feasible solution of $\mathsf{R}(I)$, and set

$$\bar{z}_{\epsilon'} = \max\{z_{\epsilon'}, b\}.$$

Since $z_{\epsilon'} \ge l$,

$$0 \le \bar{z}_{\epsilon'} - z_{\epsilon'} \le b - l.$$

Moreover, $\mathsf{R}(I)$ imposes

$$z_{\epsilon'} \ge a_{\epsilon'} + \frac{q(b)}{T} \sum_{t=1}^{T} u_{\epsilon',t}.$$

Thus, $(\boldsymbol{w}_{\epsilon'}, a_{\epsilon'}, \bar{z}_{\epsilon'}, \boldsymbol{u}_{\epsilon'})$ is feasible for the fixed-loss-level problem defining $V_\gamma(b)$. Since $b \in I$,

$$\min_{x \in I} V_\gamma(x) \le V_\gamma(b) \le L(I) + \epsilon' + (b - l).$$

Letting $\epsilon' \downarrow 0$ proves the result. □

The interval-width bound applies to the exact relaxation value $L(I)$. The numerical search uses $L_s(I)$ and certifies its result through the global bound gap.

Let $\mathsf{P}(x)$ denote the fixed-loss-level epigraph problem with value $V_\gamma(x)$. At iteration k, a checked feasible portfolio and a conservative evaluation of its objective provide an incumbent upper bound UB_k. Let $\mathcal{L}_k$ contain all current leaf intervals, including those screened from the active set $\mathcal{A}_k$. The leaves cover $[x_L, x_U]$, so define

$$LB_k = \min_{I \in \mathcal{L}_k} L_s(I), \qquad g_k = UB_k - LB_k.$$

Since every leaf has a safe lower bound,

$$LB_k \le V^\star \le V_\gamma(\hat{x}_k) \le UB_k, \qquad V^\star = \min_{x \in [x_L, x_U]} V_\gamma(x).$$

Thus, the incumbent objective error is at most g_k.

The stopping tolerance is

$$\tau_k = \varepsilon_{\text{abs}} + \varepsilon_{\text{rel}} \max\{1, |UB_k|\}.$$

An interval with $L_s(I) \ge UB_k - \tau_k$ may be screened, but it remains a leaf when LB_k is computed. The algorithm reports a certified objective gap only if $g_k \le \tau_k$. Reaching the minimum interval width, node limit, or time limit does not establish this certificate. If the active set becomes empty before the gap criterion is met, the run terminates without a certified-gap status.

For the logistic confidence function used in the experiments, the solver returns candidate primal solutions and dual multipliers. Numerical feasibility tolerances are used only to accept solver candidates. For upper-bound evaluation, candidate weights are made nonnegative, normalized, and repaired in rational arithmetic to satisfy the simplex constraints exactly. If necessary, a rational convex combination with the unit portfolio in the asset with the highest sample mean return enforces

the mean-return constraint. Only portfolios that pass exact rational feasibility checks are used to recompute conservative objective upper bounds with rigorous logistic-function enclosures. Dual candidates are verified using rational arithmetic and rigorous enclosures of the logistic function. The rational checks use the exact rational types provided by python-flint, with recorded floating-point inputs and parameters interpreted as exact binary rational values. Rigorous logistic-function enclosures are computed using Arb ball arithmetic at a working precision of at least 192 bits. The underlying methods and error control are described in [34]. If verification fails, the analytic and inherited lower bounds remain available. The certificate applies to the instance defined by the recorded numerical inputs and parameters.

Both $\mathsf{P}(x)$ and $\mathsf{R}(I)$ contain $N+T+2$ decision variables: N portfolio weights, T tail excess variables, one Rockafellar–Uryasev threshold variable, and one epigraph variable. Their numbers of constraints grow linearly in $N+T$. Both subproblems are linear programs for $\gamma = 0$ and convex quadratic programs for $\gamma > 0$, so standard convex optimization solvers can be used [28].

When $x_L < x_U$ and $0 < \varepsilon_x < x_U - x_L$, the monotone crossing bisection procedure for MΛES requires $O(\log((x_U - x_L)/\varepsilon_x))$ fixed-loss-level solves to reach loss-level tolerance ε_x, apart from endpoint and final candidate evaluations. Section A.2 of the supplementary materials gives the procedure. Adaptive loss-level discretization solves one subproblem at each of its N_x evaluated grid points.

For Algorithm 3 in the supplementary materials, let N_P and N_R denote the numbers of fixed-loss-level subproblems and interval relaxations, respectively. If $C_P(N,T)$ and $C_R(N,T)$ are their per-instance solution costs, the solver workload during the search is summarized by

$$N_P C_P(N,T) + N_R C_R(N,T),$$

where the realized counts depend on the safe bounds, stopping tolerances, and shape of V_γ. The expression excludes bound verification and preprocessing. The runtime comparisons in Section 5.2 include interval construction, calibration of Λ, outer search, and certificate checks. They compare branch-and-bound (BB) with monotone crossing bisection (BIS) and adaptive loss-level discretization (GRID) using objective agreement, feasibility, certified gaps, and computational time.

5. Numerical experiments

We evaluate the proposed models and algorithms on four daily equity datasets using rolling windows. The computational study compares monotone crossing bisection (BIS), adaptive loss-level discretization (GRID), and interval branch-and-bound (BB). The portfolio study examines out-of-sample performance and the paths of the optimized loss levels and effective confidence levels. Regularization coefficients are selected on the separate NSDX100 dataset before these comparisons.

5.1. Data and experimental setup

Table 1 summarizes four daily equity datasets for algorithm verification and out-of-sample evaluation, together with the separate NSDX100 dataset used for parameter calibration. FF49 and 100SBM are obtained from the Kenneth R. French Data Library and comprise U.S. portfolios formed by industry and by size and book-to-market characteristics, respectively. FTSE100, SSE Industry, and NSDX100 provide UK, Chinese, and U.S. equity samples, respectively. Their price and return data

were obtained from the Choice Financial Terminal. NSDX100 is used for the 2015–2019 parameter calibration in Section 5.4.

Table 1 reports the asset counts used in optimization after preprocessing. FTSE100 and NSDX100 use fixed asset pools restricted to stocks that remained constituents of the respective indices throughout the study periods and met the screening criteria. Historical constituents that exited the indices during these periods are excluded. Data quality screening excludes stock return series whose zero-return proportion, computed over the full preprocessing data interval, exceeds 20%. A similar threshold has been used in portfolio optimization to filter missing values or zero returns [35]. Of the 100 original FTSE100 series, four are excluded because their zero-return proportions exceed this threshold and one because of a clearly anomalous return record, leaving 95 assets. Of the 101 original NSDX100 series, 15 are excluded because their zero-return proportions exceed the threshold, leaving 86 assets. For 100SBM, the original missing-value codes are converted to missing values, and one portfolio with missing return observations is excluded, leaving 99 portfolios. Missing returns are not imputed.

Table 1. Information on five real-world daily benchmark datasets.

Dataset	Region	Time	Days	Assets
FF49	US	Jan/2015–Jun/2026	2827	49
FTSE100	UK	Jan/2015–Jun/2026	2865	95
SSE Industry	CN	Jan/2015–Jun/2026	2797	64
100SBM	US	Jan/2015–Jun/2026	2848	99
NSDX100	US	Jan/2015–Dec/2019	1258	86

We evaluate out-of-sample performance using the rolling-window protocol standard in portfolio backtesting studies [36,37]. Consider N risky assets over T_o out-of-sample daily trading periods. With initial wealth $S^{(0)} = 1$, price-relative vector $\boldsymbol{y}_t = \mathbf{1}_N + \boldsymbol{r}_t$, and beginning-of-day portfolio weights $\boldsymbol{w}^{(t)}$, wealth evolves as

$$S^{(t)} = S^{(t-1)}\left(\boldsymbol{y}_t^\top \boldsymbol{w}^{(t)}\right).$$

We use an estimation window of $T_w = 252$ trading days and rebalance every $h = 21$ trading days, following the monthly rolling-window convention in empirical portfolio optimization [38, 39]. Let $\tau_1, \ldots, \tau_K$ denote the scheduled rebalancing dates. At each date τ_k, the portfolio is computed using only $\{\boldsymbol{r}_{\tau_k - T_w}, \ldots, \boldsymbol{r}_{\tau_k - 1}\}$. The sample mean return vector, sample covariance matrix, empirical portfolio loss distribution, and loss-level interval $[x_L, x_U]$ are computed from the same estimation window.

Between rebalancing dates, no new optimization is performed, and the portfolio weights drift with realized returns according to

$$\boldsymbol{w}^{(t+1)} = \frac{\boldsymbol{w}^{(t)} \circ \boldsymbol{y}_t}{\boldsymbol{y}_t^\top \boldsymbol{w}^{(t)}}.$$

The first T_w observations form the initial estimation window. Conditional on the fixed asset pools, all rolling-window estimates and portfolio optimizations use only return observations available before the corresponding rebalancing date [40,41].

At each rebalancing date, the minimum daily mean-return target is

$$\eta = (1.10)^{1/252} - 1.$$

This corresponds to a 10% annualized target, a level also used in portfolio optimization studies [42,43].

The mean-return constraint was feasible at the first scheduled investment date for all four datasets. If it is infeasible in a subsequent estimation window, the scheduled rebalance is skipped and existing holdings are retained, with weights drifting with realized returns. This rule applies to both proposed models and all mean-constrained benchmarks throughout the backtests, including the cost-free performance evaluation. Benchmarks without this constraint rebalance as scheduled.

Let $\{R_t^p\}_{t=1}^{T_o}$ denote the realized portfolio returns, $\{S^{(t)}\}_{t=0}^{T_o}$ the associated wealth path, and $L_t^p = -R_t^p$ the realized losses. Let $\bar{R}$ and s_R be the sample mean and standard deviation of R_t^p, respectively, and let $M_t = \max_{0\le s\le t} S^{(s)}$ be the running maximum wealth. The pre-rebalancing weights $\boldsymbol{w}^{(\tau_k,-)}$ include the drift caused by realized returns. Table 2 summarizes the performance metrics and their preferred directions.

Table 2. Out-of-sample performance metrics and preferred directions.

Metric	Definition
Cumulative wealth (higher)	$\mathrm{CW} = S^{(T_o)}$
Sharpe ratio (higher)	$\mathrm{SR} = (\bar{R}/s_R)\sqrt{252}$
Maximum drawdown (lower)	$\mathrm{DD} = \max\limits_{1\le t\le T_o}\left(1 - \dfrac{S^{(t)}}{M_t}\right)$
$\mathrm{ES}_{0.95}$ (lower)	$\mathrm{ES}_{0.95} = \min\limits_{a\in\mathbb{R}}\left\{a + \dfrac{1}{0.05T_o}\sum\limits_{t=1}^{T_o}\left(L_t^p - a\right)_+\right\}$
Turnover (lower)	$\mathrm{TO} = \dfrac{1}{2K}\sum\limits_{k=2}^{K}\left\|\boldsymbol{w}^{(\tau_k)} - \boldsymbol{w}^{(\tau_k,-)}\right\|_1$

Unless otherwise stated, wealth and risk metrics are computed before transaction costs. The Sharpe ratio is calculated assuming a zero risk-free rate. Turnover is defined as a proportion and reported as a percentage in performance tables. The average includes all K scheduled rebalancing dates, including dates with no trade, with zero assigned to the initial allocation.

We compare the two proposed MΛES models with seven benchmarks. The equal-weight portfolio $1/N$ assigns $w_i = 1/N$ at each scheduled rebalancing date. The global minimum-variance portfolio (GMV) is also long-only and fully invested, but neither $1/N$ nor GMV has a mean-return target. Let $\Delta_N = \{\boldsymbol{w} \ge \mathbf{0} : \mathbf{1}^\top \boldsymbol{w} = 1\}$ denote their investment simplex. The remaining models use the common mean-return feasible set $\mathcal{W}_\eta$: the classical mean–variance benchmark [1, 37], two fixed-confidence mean-ES models and their ℓ_2-regularized counterparts, and the two proposed MΛES models. Following standard CVaR portfolio optimization terminology, the fixed-confidence mean-ES models are denoted by MCVaR(α), with $\alpha \in \{0.90, 0.95\}$, because ES and CVaR represent the same tail risk

functional in this setting [6–8]:

$$\begin{aligned}
\text{GMV}: &\quad \min_{\boldsymbol{w}\in\Delta_N} \boldsymbol{w}^\top \widehat{\boldsymbol{\Sigma}} \boldsymbol{w}, \\
\text{MV}: &\quad \min_{\boldsymbol{w}\in\mathcal{W}_\eta} \boldsymbol{w}^\top \widehat{\boldsymbol{\Sigma}} \boldsymbol{w}, \\
\text{MCVaR}(\alpha): &\quad \min_{\boldsymbol{w}\in\mathcal{W}_\eta} \mathrm{ES}_\alpha(L(\boldsymbol{w})), \\
\text{MCVaR}(\alpha)\text{-}\ell_2: &\quad \min_{\boldsymbol{w}\in\mathcal{W}_\eta} \left\{ \mathrm{ES}_\alpha(L(\boldsymbol{w})) + \frac{\gamma}{2}\|\boldsymbol{w}\|_2^2 \right\}.
\end{aligned}$$

Here, $\widehat{\boldsymbol{\Sigma}}$ is the sample covariance matrix from the current estimation window. The model-specific coefficients are calibrated before the four-dataset comparison, as detailed in Section 5.4. We use $\gamma = 0.06$ for both regularized MCVaR benchmarks and $\gamma = 0.075$ for MΛES-ℓ_2, holding each coefficient fixed across datasets.

For the Λ-ES models, we use the logistic confidence specification in (3.2) [17] and recalibrate β and c in each rolling window according to $[x_L, x_U]$. The admissible confidence range and endpoint conditions are

$$(\alpha_{\min}, \alpha_{\max}) = (0.90, 0.99), \qquad \Lambda(x_L) = \alpha_{\max}, \qquad \Lambda(x_U) = \alpha_{\min}.$$

The endpoint levels remain fixed across windows, while the feasible loss interval sets the location and scale of the confidence profile in each window. The optimized loss level then determines the effective confidence level.

The fixed-confidence benchmarks use $\alpha = 0.90$ and 0.95, representing moderate and high tail-risk specifications in financial risk management and CVaR portfolio analysis [12, 44]. We do not include MCVaR(0.99) as a primary benchmark because a 252-trading-day estimation window leaves only approximately two to three observations in the empirical 1% tail, making the corresponding fixed-confidence ES optimization sensitive to sampling noise. The Λ-ES models use $[0.90, 0.99]$ as an admissible confidence range. Their optimized loss levels and induced effective confidence levels are examined in Section 5.3.

The numerical implementation uses NumPy 2.4.6, SciPy 1.17.1, OSQP 1.1.3, Clarabel 0.11.1, and python-flint 0.9.0. The core implementation, tested with Python 3.14.5, is publicly available on GitHub.* All optimization problems were solved on a 64-bit Windows workstation with an Intel Core Ultra 5 225H processor and 32 GB of memory.

5.2. Algorithm verification and numerical diagnostics

We conduct algorithm verification using the January 2019–December 2021 data segment of each of the four evaluation datasets. For MΛES, we compare BB with BIS and GRID. For MΛES-ℓ_2, we compare BB with GRID because the monotone-crossing guarantee need not hold. We report objective differences, certified BB gaps, feasibility, and search cost.

Within each model-window, the compared methods use identical return scenarios, return constraints, and loss-level intervals. GRID uses 20 points per round and at most three rounds, with

*https://github.com/zkppp/lambda-es-verified-optimization. The repository includes a synthetic example and instructions for running the code.

grid-spacing tolerance 10^{-6}. The 10^{-7} objective-change threshold records stagnation but does not stop GRID. BIS uses bracket-width tolerance 10^{-6}, a tolerance of 10^{-10} when comparing the crossing value with zero, and at most 80 iterations. BB declares a certified gap only if $UB - LB \le \tau(UB)$, where $\tau(UB) = 10^{-6} + 10^{-8}\max\{1, |UB|\}$. Its minimum interval width is 10^{-8}, with limits of 200 nodes and 300 seconds. Reaching a width or resource limit does not itself certify the solution.

The HiGHS linear optimization software solves the linear subproblems. The operator splitting quadratic program (OSQP) solver handles the convex quadratic programs (QPs) with absolute and relative tolerances of 10^{-8}, at most 10^5 iterations, polishing, and warm starts. An OSQP result is accepted only with a solved or solved-inaccurate status, primal and dual residuals no greater than 10^{-7}, an absolute duality gap no greater than 10^{-7} times the objective scale, and independently checked constraint violation no greater than 10^{-7}. Initial weights are equal weights when return-feasible and otherwise are mixed with the highest-mean-return asset. Tail variables start from the empirical loss quantile. If OSQP fails these checks, sequential least squares programming (SLSQP) uses an available OSQP candidate or the original initialization, with `ftol` $= 10^{-9}$ and at most 1000 iterations. Failed or nonfinite SLSQP results are discarded. The BB lower bounds are verified separately, as described in Section 4.2.

Tables 3 and 4 report objective differences, certified gaps, feasibility, and search cost. In the "Certified / feasible" column, the first count gives the number of BB runs that meet the certified-gap criterion, and the second gives the number of feasible model-windows. The global lower bound includes every current leaf interval, even after removal from the active search set.

Table 3. Numerical agreement among BIS, GRID, and BB for MΛES.

Dataset	$\lvert V_{\mathrm{BIS}} - V_{\mathrm{BB}}^{\mathrm{inc}}\rvert$ mean / max (10^{-6})	$\lvert V_{\mathrm{GRID}} - V_{\mathrm{BB}}^{\mathrm{inc}}\rvert$ mean / max (10^{-6})	Max $UB - LB$ (10^{-6})	Certified / feasible	BB nodes median / max	BB time (s) median / max	Max comparator violation
FF49	0.343 / 0.960	25.027 / 97.123	0.981	25/25	29 / 33	1.20 / 1.38	1.11×10^{-16}
FTSE100	0.364 / 0.766	9.626 / 45.061	0.947	25/25	27 / 33	1.58 / 1.81	7.77×10^{-16}
SSE Industry	0.332 / 0.728	12.110 / 25.271	0.900	23/23	29 / 31	1.40 / 1.68	6.66×10^{-16}
100SBM	0.325 / 0.684	27.574 / 93.524	1.004	24/24	31 / 33	1.72 / 1.95	2.22×10^{-16}

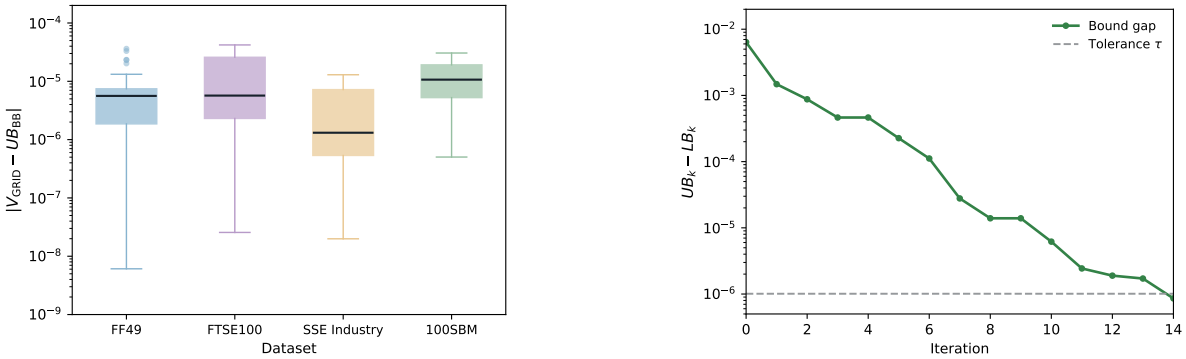
Table 4. Numerical agreement between GRID and BB for MΛES-ℓ_2.

Dataset	$\lvert V_{\mathrm{GRID}} - V_{\mathrm{BB}}^{\mathrm{inc}}\rvert$ mean / max (10^{-6})	Max $UB - LB$ (10^{-6})	Certified / feasible	BB nodes median / max	BB time (s) median / max	Max GRID violation
FF49	8.713 / 36.188	1.007	25/25	53 / 107	25.12 / 66.12	7.82×10^{-14}
FTSE100	12.341 / 42.058	0.998	25/25	29 / 49	16.10 / 54.12	2.48×10^{-13}
SSE Industry	3.699 / 12.955	1.009	23/23	47 / 79	27.81 / 61.63	7.16×10^{-14}
100SBM	11.832 / 30.553	1.007	24/24	59 / 101	38.64 / 126.64	7.28×10^{-14}

For MΛES, the mean BIS–BB objective differences in Table 3 range from 0.325×10^{-6} to 0.364×10^{-6} across datasets, and the largest window-level difference is 0.960×10^{-6}, versus 97.123×10^{-6} for GRID–BB. Because BIS uses the monotone-crossing structure, its agreement with BB provides an independent check on the incumbent. All feasible BB runs meet the certified-gap criterion, with a largest terminal gap of 1.004×10^{-6}. The verified gap bounds the incumbent's objective error relative to the global optimum.

For MΛES-ℓ_2, Table 4 shows a maximum GRID–BB difference of 42.058×10^{-6} and a maximum GRID constraint violation of 2.48×10^{-13}. All feasible BB runs satisfy the certified-gap criterion, with

a largest gap of 1.009×10^{-6}. The differences between GRID and BB describe finite-grid agreement. The verified BB gaps provide the objective error bounds. Both maximum gaps exceed 10^{-6} but satisfy $\tau(UB)$, which includes the scale term. Compared with MΛES, the regularized model requires more BB nodes and longer runtimes in every dataset.



(a) Absolute GRID–BB objective difference

(b) Certified BB gap over iterations

Figure 1. Numerical agreement and convergence diagnostics for MΛES-ℓ_2.

Figure 1(a) shows the distribution of absolute GRID–BB differences across algorithm-verification windows. The median difference on 100SBM exceeds that on SSE Industry, consistent with the means in Table 4. Panel (b) traces the verified BB gap in the first FF49 window. The gap declines to the stopping tolerance as the bounds are refined. At each iteration, LB_k is the minimum over all current leaf bounds, so the plotted gap is global for that window.

Section A.5 of the supplementary materials provides further computational checks. Table A1 reports point and interval evaluations and runtime. Table A2 audits the BB certificates, and Table A3 checks fixed-level QP solutions against Clarabel.

5.3. Out-of-sample performance and loss-level diagnostics

Under the rolling-window setting in Section 5.1, the out-of-sample evaluation covers January 2022 through June 2026. At each feasible estimation window, both MΛES and MΛES-ℓ_2 use the interval branch-and-bound method (BB) of Section 4.2 for the outer loss-level search, with the gap stopping criterion $UB - LB \leq 10^{-6} + 10^{-8} \max\{1, |UB|\}$. Following the algorithm verification in Section 5.2, the same solution and bound verification procedure is applied independently to each feasible window in the out-of-sample backtests, and all such windows satisfy this criterion. Table 5 summarizes the rebalancing-window counts and maximum certified objective gaps. Scheduled windows include the initial allocation. Feasible windows admit portfolios satisfying the mean-return constraint.

Table 5. Out-of-sample rebalancing windows and maximum certified objective gaps.

Dataset	Scheduled windows	Feasible windows	Skipped rebalances	MΛES (10^{-6})	MΛES-ℓ_2 (10^{-6})
FF49	51	51	0	0.991763	0.999958
FTSE100	53	53	0	0.961914	1.009662
SSE Industry	51	51	0	0.985020	1.008787
100SBM	52	52	0	1.001612	1.008053

Note. The last two columns report maximum certified gaps over all feasible windows.

Table 6. Out-of-sample performance of different portfolio models on four equity datasets.

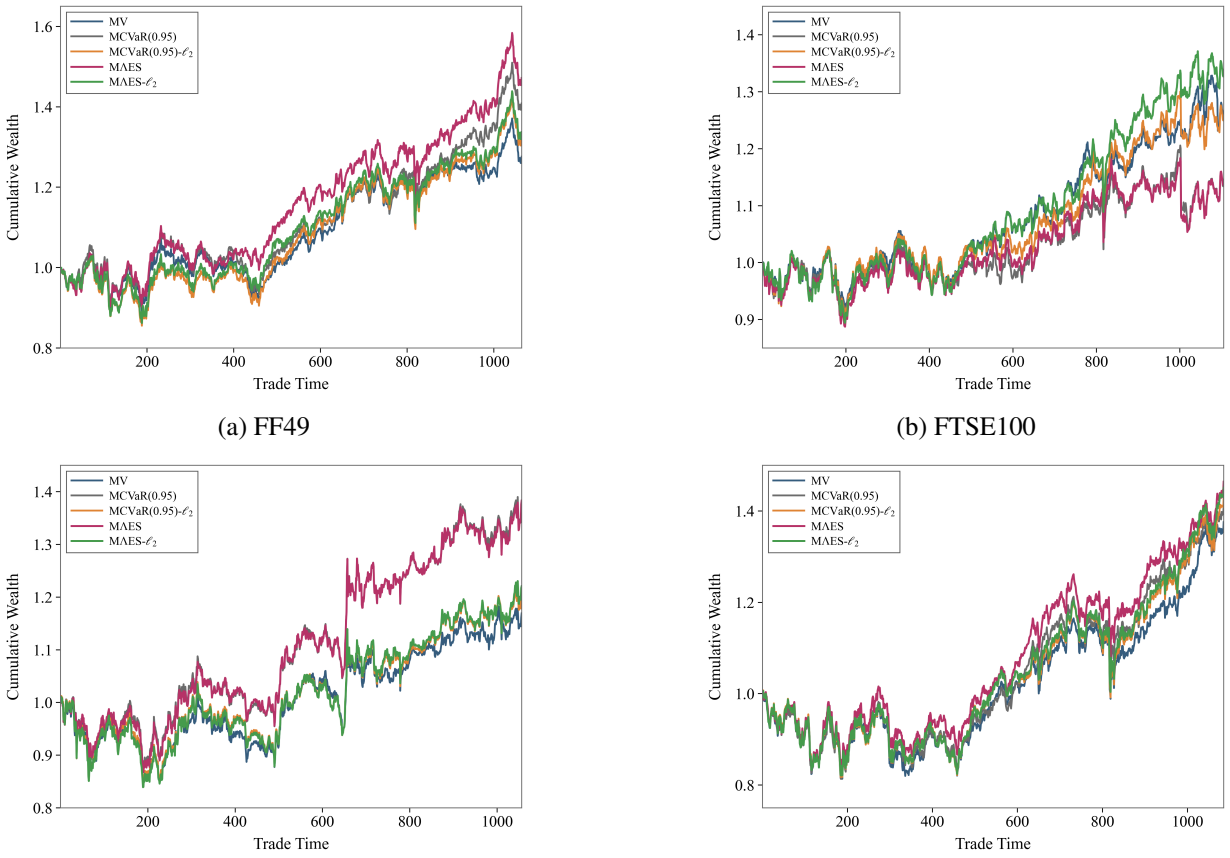
Dataset	Model	Cumulative wealth	Sharpe ratio	MaxDD (%)	$ES_{0.95}$ (%)	Turnover (%)
FF49	1/N	1.429	0.575	21.01	2.429	**1.97**
	GMV	1.228	0.459	15.98	1.779	12.60
	MV	1.276	0.535	14.94	**1.777**	15.80
	MCVaR(0.90)	1.334	0.612	14.19	1.808	27.04
	MCVaR(0.90)-ℓ_2	1.341	0.610	16.62	1.855	10.52
	MCVaR(0.95)	1.409	0.694	15.13	1.844	28.74
	MCVaR(0.95)-ℓ_2	1.322	0.581	16.69	1.864	11.73
	MΛES	**1.472**	**0.769**	**12.86**	1.875	30.87
	MΛES-ℓ_2	1.337	0.601	16.24	1.855	17.34
FTSE100	1/N	**1.370**	0.567	22.23	2.121	**2.68**
	GMV	1.250	0.525	12.07	1.654	15.65
	MV	1.261	0.545	12.09	1.653	15.72
	MCVaR(0.90)	1.307	0.589	12.12	1.681	28.12
	MCVaR(0.90)-ℓ_2	1.314	0.579	**11.53**	**1.618**	15.23
	MCVaR(0.95)	1.137	0.295	12.19	1.874	31.90
	MCVaR(0.95)-ℓ_2	1.252	0.491	12.30	1.740	24.97
	MΛES	1.140	0.304	12.94	1.836	33.39
	MΛES-ℓ_2	1.331	**0.635**	12.44	1.630	21.58
SSE Industry	1/N	1.184	0.328	21.35	2.413	**1.90**
	GMV	1.145	0.304	14.92	**2.039**	12.60
	MV	1.169	0.337	16.38	2.072	21.08
	MCVaR(0.90)	1.244	0.436	14.50	2.082	31.17
	MCVaR(0.90)-ℓ_2	1.212	0.384	15.65	2.177	18.01
	MCVaR(0.95)	1.376	0.590	**13.75**	2.126	28.80
	MCVaR(0.95)-ℓ_2	1.202	0.368	16.39	2.180	18.48
	MΛES	**1.383**	**0.599**	13.99	2.117	30.43
	MΛES-ℓ_2	1.221	0.393	16.95	2.182	21.56
100SBM	1/N	1.438	0.508	24.67	2.843	**1.43**
	GMV	1.421	0.642	16.91	**2.035**	12.54
	MV	1.376	0.568	18.63	2.227	24.30
	MCVaR(0.90)	1.379	0.577	18.94	2.195	29.99
	MCVaR(0.90)-ℓ_2	1.432	0.597	19.05	2.353	17.40
	MCVaR(0.95)	1.411	0.609	18.97	2.178	32.20
	MCVaR(0.95)-ℓ_2	1.430	0.588	18.85	2.372	18.79
	MΛES	**1.464**	**0.660**	**16.85**	2.192	33.60
	MΛES-ℓ_2	1.451	0.618	18.39	2.329	21.29

Note. Bold and underlined values denote the best and second-best results within each dataset and metric.

Table 6 compares different portfolio models in cumulative wealth, the Sharpe ratio, maximum drawdown, realized $ES_{0.95}$, and turnover. The 1/N and GMV portfolios have no return target, while the remaining models use the 10% annualized target. The regularization coefficients were selected separately on NSDX100 before this evaluation, as detailed in Section 5.4. They are $\gamma = 0.075$ for MΛES-ℓ_2 and $\gamma = 0.060$ for both MCVaR(0.90)-ℓ_2 and MCVaR(0.95)-ℓ_2.

Table 6 shows that MΛES attains the highest cumulative wealth and Sharpe ratio on FF49, SSE Industry, and 100SBM. On FF49, it improves wealth (1.472 versus 1.409), the Sharpe ratio (0.769 versus 0.694), and maximum drawdown (12.86% versus 15.13%) relative to MCVaR(0.95). The wealth and Sharpe advantages over MCVaR(0.95) are smaller on SSE Industry, at 1.383 versus 1.376 and 0.599 versus 0.590. On 100SBM, MΛES also has the lowest maximum drawdown at 16.85%. FTSE100 is an exception, with 1/N giving the highest wealth and MΛES-ℓ_2 the highest Sharpe ratio. The results are consistent with a benefit from loss-dependent tail emphasis in three datasets, mainly in wealth and risk-adjusted return. MΛES does not attain the lowest realized $ES_{0.95}$ in any dataset, so the improvement does not extend uniformly to that downside metric.

Relative to MΛES, MΛES-ℓ_2 lowers average rebalancing turnover by 8.87–13.53 percentage points across the four datasets. On FTSE100, turnover falls from 33.39% to 21.58%, while wealth rises from 1.140 to 1.331, the Sharpe ratio rises from 0.304 to 0.635, and realized $ES_{0.95}$ falls from 1.836% to 1.630%. On 100SBM, wealth declines slightly from 1.464 to 1.451 as turnover falls from 33.60% to 21.29%. Wealth falls more substantially on FF49 and SSE Industry. Thus, the penalty reduces trading across all four datasets, but its effect on realized performance depends on the dataset. The turnover reductions are consistent with the weight-stabilizing role of ℓ_2 regularization [45, 46].



(a) FF49

(b) FTSE100

(c) SSE Industry

(d) 100SBM

Figure 2. Cumulative wealth over trading time for five models on four datasets.

Figure 2 traces five selected strategies from Table 6: MV, the two MCVaR(0.95) models, and the two MΛES models. On FF49, MΛES gains a sustained lead around the middle of the evaluation period. On SSE Industry, it moves ahead later, but finishes only slightly above MCVaR(0.95). It also stays ahead of the plotted comparators through much of the latter half on 100SBM. On FTSE100, MΛES-ℓ_2 finishes ahead of the other plotted strategies after pulling away from MΛES. These paths show that the terminal wealth differences in Table 6 generally build over successive trading periods, with a smaller advantage on SSE Industry than on FF49 or 100SBM.

Figure 3 pairs the optimized loss level $\widehat{x}_k$ with the effective confidence level of MΛES-ℓ_2 at rebalancing date k:

$$\alpha_k^* = \Lambda_k(\widehat{x}_k).$$

The endpoint confidence levels remain fixed, but Λ_k is recalibrated to the feasible loss interval in each window. The optimized $\widehat{x}_k$ then selects a point on that window's confidence profile. The observed α_k^* ranges from about 0.916 to 0.982, corresponding to roughly 5 to 21 tail observations in a 252-day estimation window. It stays above 0.95 in most FF49 windows, varies on both sides of 0.95 on FTSE100, and remains closer to 0.95 on SSE Industry and 100SBM. In the fixed-level representation of Section 2.2, α_k^* determines the coefficient on positive tail excess. The figure therefore displays the changing tail emphasis used by the regularized model. Its variation reflects both the window-specific calibration of Λ_k and the optimized loss level.

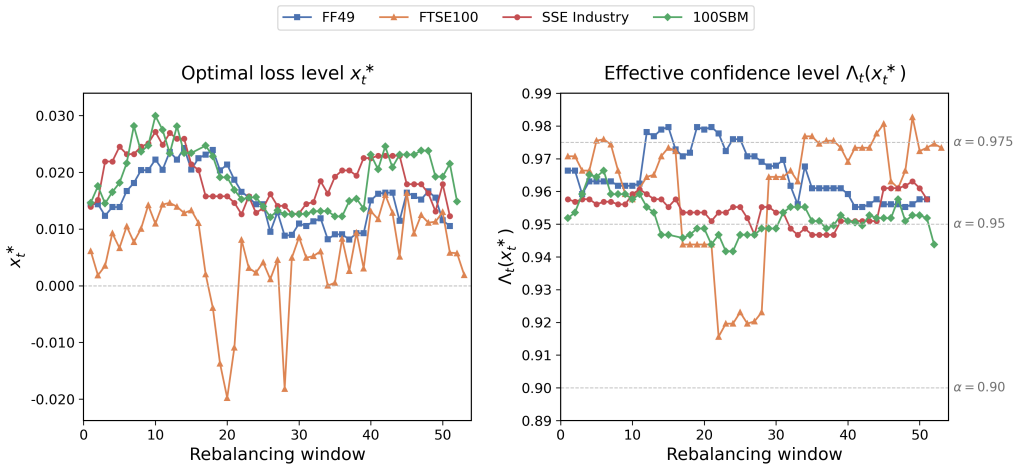
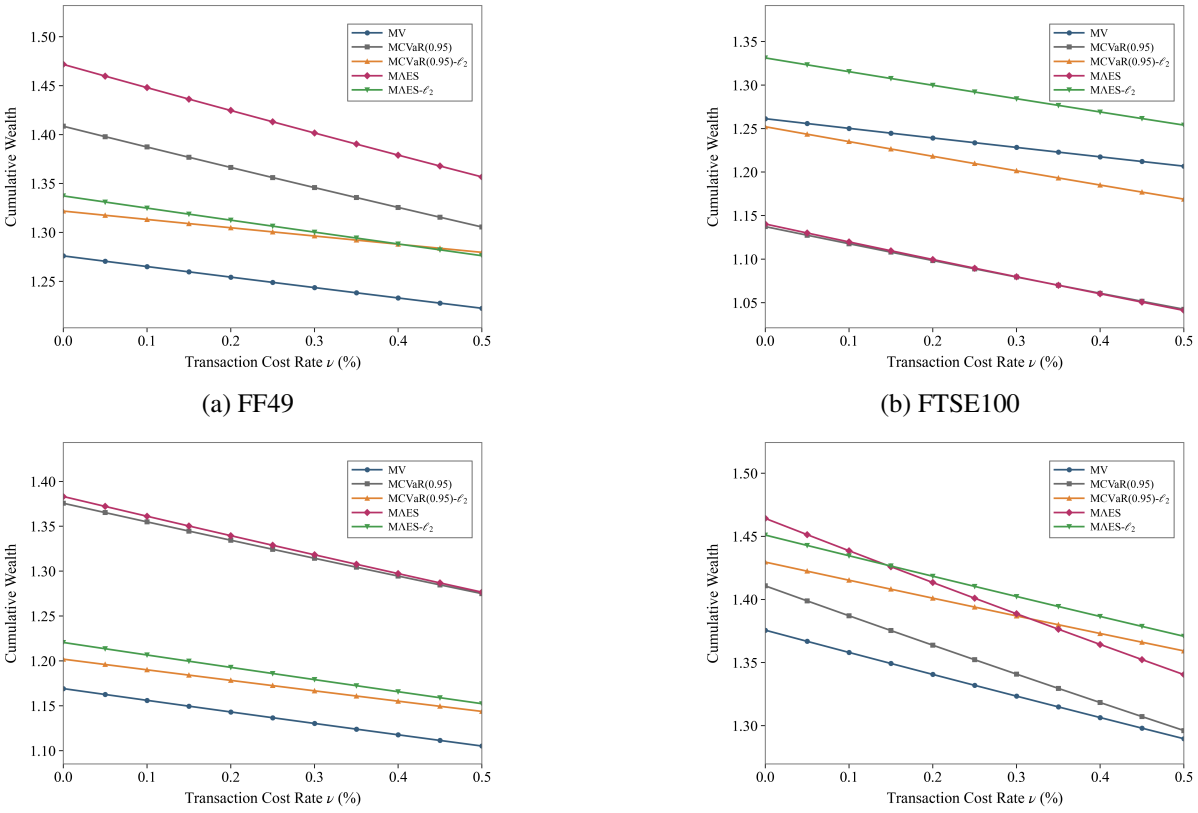


Figure 3. Optimized loss and effective confidence levels for MΛES-ℓ_2.

We assess transaction-cost sensitivity after the backtest using a proportional-cost convention for portfolio rebalancing [37, 47, 48] and the same holdings used in Figure 2. Let $\boldsymbol{w}^{(\tau_k,-)}$ be the weights after return-induced drift and before rebalancing, and let $\boldsymbol{w}^{(\tau_k)}$ be the new weights. The one-way traded fraction q_k and terminal wealth at the round-trip cost rate ν are

$$q_k = \frac{1}{2}\left\|\boldsymbol{w}^{(\tau_k)} - \boldsymbol{w}^{(\tau_k,-)}\right\|_1, \qquad S^{(T_o)}(\nu) = S^{(T_o)}(0)\prod_{k=1}^{K}(1-\nu q_k).$$

The initial purchase from cash has $q_1 = 0.5$ because pre-trade risky weights are zero. Under the rule in Section 5.1, a skipped rebalance has $q_k = 0$. Unlike the turnover reported in Table 6, the cost calculation charges for the initial purchase. If buying and selling each incur a rate c, this convention gives $\nu = 2c$. We vary ν from 0 to 0.5% in 0.05% increments, applying costs only when trades occur. Holdings and rebalancing dates remain those of the cost-free backtest, and no terminal liquidation charge is imposed.



(a) FF49 (b) FTSE100

(c) SSE Industry (d) 100SBM

Figure 4. Final cumulative wealth versus transaction cost rate across four datasets.

Figure 4 applies the cost convention to the same five strategies plotted in Figure 2. Among these strategies, MΛES retains the highest terminal wealth on FF49 throughout the cost range and reaches 1.357 at $\nu = 0.5\%$. MΛES-ℓ_2 remains highest on FTSE100 and ends at 1.254 at the same rate. On SSE Industry, the small zero-cost wealth gap between MΛES and MCVaR(0.95) narrows as costs rise. On 100SBM, MΛES-ℓ_2 overtakes MΛES from $\nu = 0.15\%$ and ends at 1.371, compared with 1.341 for MΛES at 0.5%. The turnover reduction in Table 6 helps explain why regularization reverses the zero-cost ranking on 100SBM.

5.4. Regularization parameter calibration

We calibrate the ℓ_2 regularization coefficient separately for MΛES-ℓ_2, MCVaR(0.90)-ℓ_2, and MCVaR(0.95)-ℓ_2 on the NSDX100 dataset over the period between 2015 and 2019. We choose γ from historical portfolio performance, as in performance-based portfolio validation [49]. The calibration uses a 10% annualized return target, a 252-trading-day estimation window, and rebalancing every 21 trading days without transaction costs. Each candidate is evaluated over 48 rolling windows.

For each model, we evaluate the same candidate grid:

$$\Gamma = \{0, 0.02, 0.05, 0.055, 0.06, 0.065, 0.07, 0.075, 0.08, 0.10, 0.15, 0.20\}.$$

The value $\gamma = 0$ provides an unregularized reference. The coefficient selected for each model is then applied to that model in the subsequent experiments on all datasets.

Let $S_{q,\gamma}^{(t)}$ be the calibration wealth path for model q and candidate γ, with $S_{q,\gamma}^{(0)} = 1$. Over T_c calibration trading days, the annualized return is

$$\mathrm{AR}_q(\gamma) = \left(S_{q,\gamma}^{(T_c)}\right)^{252/T_c} - 1.$$

Let $\hat{w}_{i,q,\gamma}^{(t)}$ denote the weight of asset i held on calibration day t, after any scheduled rebalance and before that day's return. The effective number of assets is

$$\mathrm{EN}_q(\gamma) = \frac{1}{T_c}\sum_{t=1}^{T_c}\left(\sum_{i=1}^{N}(\hat{w}_{i,q,\gamma}^{(t)})^2\right)^{-1}.$$

The five scoring metrics are the annualized return (AR), Sharpe ratio (SR), maximum drawdown (MaxDD), empirical expected shortfall at the 95% level ($\mathrm{ES}_{0.95}$), and effective number of assets (EN). Higher AR, SR, and EN are preferred, whereas lower MaxDD and $\mathrm{ES}_{0.95}$ are preferred.

We aggregate these metrics through equal-weight ranks, drawing on the broader use of multiple criteria in fund evaluation [50,51]. For each model q and metric j, we rank the $K = 12$ candidates from worst to best, assigning average ranks to ties. If $r_{q,j}(\gamma)$ is the resulting rank, the normalized score and the equally weighted joint score are

$$s_{q,j}(\gamma) = \frac{r_{q,j}(\gamma) - 1}{K - 1}, \qquad J_q(\gamma) = \frac{1}{5}\sum_{j\in\{\mathrm{AR},\mathrm{SR},\mathrm{MaxDD},\mathrm{ES}_{0.95},\mathrm{EN}\}} s_{q,j}(\gamma).$$

Ranks are computed from unrounded metric values. If candidates have the same joint score, we choose the one with the higher Sharpe ratio, then the lower $\mathrm{ES}_{0.95}$, and finally the smaller γ. Turnover is examined as a diagnostic and does not enter the score.

Table 7 reports the selected coefficients and their calibration-period outcomes. The joint scores are highest among the tested candidates at $\gamma = 0.075$ for MΛES-ℓ_2 and at $\gamma = 0.060$ for both regularized MCVaR models. Figure 5 displays the joint score over the candidate grid for each model.

Table 7. Model-specific regularization parameter calibration on NSDX100.

Model	Selected γ	Joint score	AR (%)	SR	MaxDD (%)	$\mathrm{ES}_{0.95}$ (%)	EN
MΛES-ℓ_2	0.075	0.800	15.87	1.417	9.26	1.654	17.61
MCVaR(0.90)-ℓ_2	0.060	0.709	16.21	1.430	10.46	1.708	19.96
MCVaR(0.95)-ℓ_2	0.060	0.727	15.81	1.411	9.20	1.667	16.79

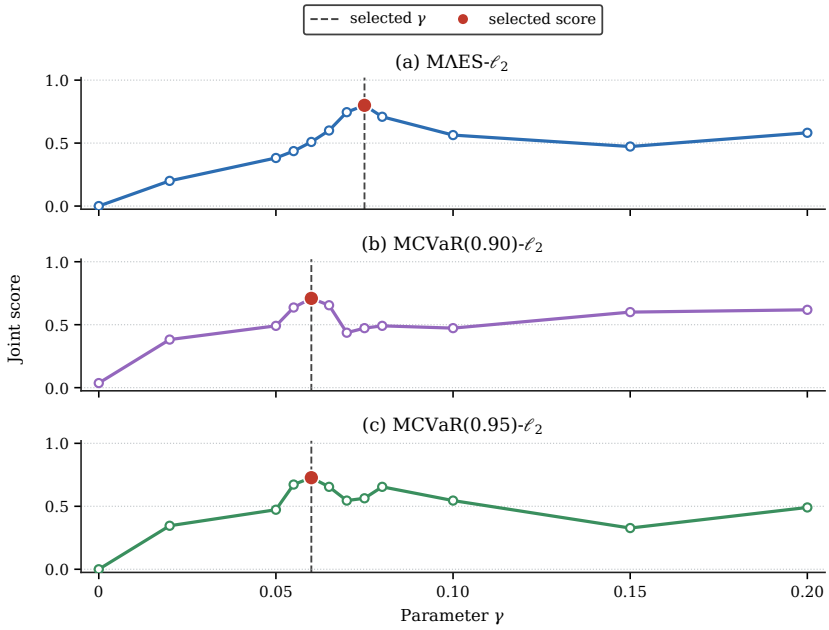


Figure 5. Joint calibration scores for the three regularized models on NSDX100.

The selected coefficients are interior points of the tested grid. For each model, the joint score at $\gamma = 0$ is below the selected score, while larger coefficients do not consistently improve it.

6. Conclusion

This paper develops a mean-constrained lambda expected shortfall (MΛES) portfolio model and its regularized extension. The loss-level decomposition reduces each joint empirical formulation to a scalar outer problem with convex fixed-confidence ES subproblems. The optimized loss level induces the effective confidence level through Λ. The monotone-crossing structure yields a bisection procedure for MΛES. Since the regularized model need not retain this structure, we derive a problem-specific interval relaxation, establish its lower-bound validity and interval-width tightness, and use it in a relaxation-guided interval branch-and-bound procedure. In the algorithm-verification experiments, every feasible BB run satisfies the certified-gap criterion using a feasible incumbent upper bound and a verified global lower bound. Comparisons with BIS and GRID, and independent QP checks, further support the numerical implementation.

Numerical experiments show that the optimized loss levels and effective confidence levels vary across rebalancing periods. MΛES attains the highest cumulative wealth and Sharpe ratio in most datasets. The regularized model reduces turnover in every dataset. The transaction-cost analysis shows that this reduction can reverse the ranking of terminal wealth as trading costs rise. The decomposition thus makes the loss-dependent confidence specification amenable to certified search, while the backtests show how the resulting models perform under rebalancing and trading costs.

Author contributions

Kunpeng Zhu: Conceptualization, methodology, formal analysis, data curation, investigation, software, visualization, validation, writing–original draft, writing–review and editing; Yanjing Guo: Methodology, formal analysis, validation, 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.

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. H. Markowitz, Portfolio selection, *J. Financ.*, **7** (1952), 77–91. https://doi.org/10.1111/j.1540-6261.1952.tb01525.x
2. P. Artzner, F. Delbaen, J. M. Eber, D. Heath, Coherent measures of risk, *Math. Financ.*, **9** (1999), 203–228. https://doi.org/10.1111/1467-9965.00068
3. C. Acerbi, D. Tasche, On the coherence of expected shortfall, *J. Bank. Financ.*, **26** (2002), 1487–1503. https://doi.org/10.1016/S0378-4266(02)00283-2
4. W. Liu, Y. Hu, On optimal reinsurance in the presence of premium budget constraint and reinsurer's risk limit, *Adv. Appl. Probab.*, **56** (2024), 1251–1278. https://doi.org/10.1017/apr.2024.2
5. W. Jammernegg, P. Kischka, L. Silbermayr, Risk preferences, newsvendor orders and supply chain coordination using the mean-CVaR model, *Int. J. Prod. Econ.*, **270** (2024), 109171. https://doi.org/10.1016/j.ijpe.2024.109171
6. R. T. Rockafellar, S. Uryasev, Optimization of conditional value-at-risk, *J. Risk*, **2** (2000), 21–41. https://doi.org/10.21314/JOR.2000.038
7. R. T. Rockafellar, S. Uryasev, Conditional value-at-risk for general loss distributions, *J. Bank. Financ.*, **26** (2002), 1443–1471. https://doi.org/10.1016/S0378-4266(02)00271-6
8. P. Krokhmal, J. Palmquist, S. Uryasev, Portfolio optimization with conditional value-at-risk objective and constraints, *J. Risk*, **4** (2002), 43–68. https://doi.org/10.21314/JOR.2002.057
9. F. Caccioli, I. Kondor, G. Papp, Portfolio optimization under expected shortfall: contour maps of estimation error, *Quant. Finance*, **18** (2018), 1295–1313. https://doi.org/10.1080/14697688.2017.1390245

10. G. Papp, I. Kondor, F. Caccioli, Optimizing expected shortfall under an ℓ_1 constraint—an analytic approach, *Entropy*, **23** (2021), 523. https://doi.org/10.3390/e23050523

11. K. Nakagawa, M. Abe, S. Kuroki, Doubly robust mean-CVaR portfolio, *arXiv preprint arXiv:2309.11693*, 2023. https://doi.org/10.48550/arXiv.2309.11693

12. E. Anderson, H. Xu, D. Zhang, Varying confidence levels for CVaR risk measures and minimax limits, *Math. Program.*, **180** (2020), 327–370. https://doi.org/10.1007/s10107-018-01359-8

13. C. S. Pun, T. Wang, Z. Yan, Data-driven distributionally robust CVaR portfolio optimization under a regime-switching ambiguity set, *Manuf. Serv. Oper. Manag.*, **25** (2023), 1779–1795. https://doi.org/10.1287/msom.2023.1229

14. M. Frittelli, M. Maggis, I. Peri, Risk measures on $\mathcal{P}(\mathbb{R})$ and value at risk with probability/loss function, *Math. Financ.*, **24** (2014), 442–463. https://doi.org/10.1111/mafi.12028

15. A. Hitaj, C. Mateus, I. Peri, Lambda value at risk and regulatory capital: A dynamic approach to tail risk, *Risks*, **6** (2018), 17. https://doi.org/10.3390/risks6010017

16. J. Corbetta, I. Peri, Backtesting Lambda value at risk, *Eur. J. Financ.*, **24** (2018), 1075–1087. https://doi.org/10.1080/1351847X.2017.1339105

17. F. Bellini, M. Huang, Q. Wang, R. Wang, Lambda expected shortfall, *arXiv preprint arXiv:2512.23139*, 2025. https://doi.org/10.48550/arXiv.2512.23139

18. T. J. Boonen, Y. Chen, X. Han, Q. Wang, Optimal insurance design with lambda-value-at-risk, *Eur. J. Oper. Res.*, **327** (2025), 232–246. https://doi.org/10.1016/j.ejor.2025.04.038

19. I. Peri, L. Wunderlich, Numerical methods for lambda quantiles: Robust evaluation and portfolio optimization, *SIAM J. Financial Math.*, **17** (2026), 1308–1332. https://doi.org/10.1137/25M1819983

20. V. DeMiguel, L. Garlappi, F. J. Nogales, R. Uppal, A generalized approach to portfolio optimization: Improving performance by constraining portfolio norms, *Manag. Sci.*, **55** (2009), 798–812. https://doi.org/10.1287/mnsc.1080.0986

21. S. Still, I. Kondor, Regularizing portfolio optimization, *New J. Phys.*, **12** (2010), 075034. https://doi.org/10.1088/1367-2630/12/7/075034

22. G. Papp, F. Caccioli, I. Kondor, Bias–variance tradeoff in portfolio optimization under expected shortfall with ℓ_2 regularization, *J. Stat. Mech. Theory Exp.*, **2019** (2019), 013402. https://doi.org/10.1088/1742-5468/aaf108

23. J. Fernández, B. G. Tóth, Interval tools in branch-and-bound methods for global optimization, in *The Palgrave Handbook of Operations Research*, S. Salhi, J. Boylan eds., Palgrave Macmillan, Cham, 2022, 237–267. https://doi.org/10.1007/978-3-030-96935-6_8

24. S. Elloumi, A. Lambert, B. Neveu, G. Trombettoni, Global solution of quadratic problems using interval methods and convex relaxations, *J. Glob. Optim.*, **91** (2025), 331–353. https://doi.org/10.1007/s10898-024-01370-8

25. M. Alda, Importance of portfolio optimization in SRI and conventional pension funds, *Financ. Innov.*, **11** (2025), 79. https://doi.org/10.1186/s40854-025-00761-4

26. K. Tanaka, R. Yamamoto, Cardinality-constrained maximal predictability portfolios with an ℓ_2 regularization, *J. Ind. Manag. Optim.*, **22** (2026), 2755–2783. https://doi.org/10.3934/jimo.2026101
27. T. Zhang, S. Lai, Portfolio selection balancing concentration and diversification, *J. Ind. Manag. Optim.*, **21** (2025), 3618–3647. https://doi.org/10.3934/jimo.2025025
28. S. Boyd, L. Vandenberghe, *Convex Optimization*, Cambridge University Press, Cambridge, 2004. https://doi.org/10.1017/CBO9780511804441
29. I. D. Coope, C. J. Price, On the convergence of grid-based methods for unconstrained optimization, *SIAM J. Optim.*, **11** (2001), 859–869. https://doi.org/10.1137/S1052623499354989
30. J. Larson, M. Menickelly, S. M. Wild, Derivative-free optimization methods, *Acta Numer.*, **28** (2019), 287–404. https://doi.org/10.1017/S0962492919000060
31. R. Horst, H. Tuy, *Global Optimization: Deterministic Approaches*, 3rd edition, Springer, Berlin, Heidelberg, 1996. https://doi.org/10.1007/978-3-662-03199-5
32. A. Neumaier, Complete search in continuous global optimization and constraint satisfaction, *Acta Numer.*, **13** (2004), 271–369. https://doi.org/10.1017/S0962492904000194
33. O. Stein, *Basic Concepts of Global Optimization*, Springer, Berlin, Heidelberg, 2024. https://doi.org/10.1007/978-3-662-66240-3
34. F. Johansson, Arb: Efficient arbitrary-precision midpoint-radius interval arithmetic, *IEEE Trans. Comput.*, **66** (2017), 1281–1292. https://doi.org/10.1109/TC.2017.2690633
35. C. Bongiorno, D. Challet, Non-linear shrinkage of the price return covariance matrix is far from optimal for portfolio optimization, *Financ. Res. Lett.*, **52** (2023), 103383. https://doi.org/10.1016/j.frl.2022.103383
36. B. Li, D. Sahoo, S. C. H. Hoi, OLPS: A toolbox for on-line portfolio selection, *J. Mach. Learn. Res.*, **17** (2016), 1–5. Available from: https://www.jmlr.org/papers/v17/15-317.html.
37. Z. R. Lai, H. Yang, A survey on gaps between mean-variance approach and exponential growth rate approach for portfolio optimization, *ACM Comput. Surv.*, **55** (2023), 1–36. https://doi.org/10.1145/3485274
38. S. Goto, Y. Xu, Improving mean variance optimization through sparse hedging restrictions, *J. Financ. Quant. Anal.*, **50** (2015), 1415–1441. https://doi.org/10.1017/S0022109015000526
39. M. Ao, Y. Li, X. Zheng, Approaching mean-variance efficiency for large portfolios, *Rev. Financ. Stud.*, **32** (2019), 2890–2919. https://doi.org/10.1093/rfs/hhy105
40. D. Neděla, S. Ortobelli Lozza, T. Tichý, Dynamic return scenario generation approach for large-scale portfolio optimisation framework, *Comput. Econ.*, **65** (2025), 819–843. https://doi.org/10.1007/s10614-023-10541-w
41. J. Yu, K. C. Chang, A hierarchical signal-to-policy learning framework for risk-aware portfolio optimization, *Int. J. Financ. Stud.*, **14** (2026), 75. https://doi.org/10.3390/ijfs14030075
42. A. A. P. Santos, F. J. Nogales, E. Ruiz, D. J. C. van Dijk, Optimal portfolios with minimum capital requirements, *J. Bank. Financ.*, **36** (2012), 1928–1942. https://doi.org/10.1016/j.jbankfin.2012.03.001

43. L. Callot, M. Caner, A. Ö. Önder, E. Ulaşan, A nodewise regression approach to estimating large portfolios, *J. Bus. Econ. Stat.*, **39** (2021), 520–531. https://doi.org/10.1080/07350015.2019.1683018
44. P. Jorion, *Value at Risk: The New Benchmark for Managing Financial Risk*, 3rd edition, McGraw-Hill, New York, 2007.
45. H. Zhao, L. Kong, Penalty method for the sparse portfolio optimization problem, *J. Ind. Manag. Optim.*, **20** (2024), 2864–2884. https://doi.org/10.3934/jimo.2024031
46. Q. Wang, K. Zhu, Y. Guo, L. Yang, Z. Wu, An efficient ADMM for multi-period sparse behavioral portfolio optimization based on cumulative prospect theory, *J. Ind. Manag. Optim.*, **22** (2026), 1726–1757. https://doi.org/10.3934/jimo.2026064
47. A. Blum, A. Kalai, Universal portfolios with and without transaction costs, *Mach. Learn.*, **35** (1999), 193–205. https://doi.org/10.1023/A:1007530728748
48. Y. Lin, Y. Zhang, Z. R. Lai, C. Li, Autonomous sparse mean-CVaR portfolio optimization, in *Proceedings of the 41st International Conference on Machine Learning*, vol. 235 of Proceedings of Machine Learning Research. PMLR, Vienna, Austria, 2024, 30440–30456. Available from: https://proceedings.mlr.press/v235/lin24w.html.
49. G. Y. Ban, N. El Karoui, A. E. B. Lim, Machine learning and portfolio optimization, *Manag. Sci.*, **64** (2018), 1136–1154. https://doi.org/10.1287/mnsc.2016.2644
50. L. C. Dias, P. Xidonas, A. Samitas, A novel Sigma-Mu multiple criteria decision aiding approach for mutual funds portfolio selection, *Eur. J. Oper. Res.*, **322** (2025), 589–598. https://doi.org/10.1016/j.ejor.2024.11.003
51. E. F. Aguirre González, P. C. Manyoma Velásquez, D. F. Manotas-Duque, Prioritization of CIFs taking into account multiple criteria for the construction of efficient portfolios, *Cogent Bus. Manag.*, **11** (2024), 2371072. https://doi.org/10.1080/23311975.2024.2371072

AIMS Press

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