Applied Soft Computing 113 (2021) 107948

Contents lists available at ScienceDirect

Applied Soft Computing

journal homepage: www.elsevier.com/locate/asoc

Robust mean-risk portfolio optimization using machine learning-based trade-off parameter

Liangyu Min, Jiawei Dong, Jiangwei Liu, Xiaomin Gong *

School of Information Management and Engineering, Shanghai University of Finance and Economics, Shanghai 200433, China

ARTICLE INFO

Article history:
Received 30 May 2021
Received in revised form 20 August 2021
Accepted 24 September 2021
Available online 8 October 2021

Keywords:
Portfolio selection
Robust optimization
Hybrid models
Machine learning
Risk measures

ABSTRACT

Conservatism is the notorious problem of the worst-case robust portfolio optimization, and this issue has raised broad discussion in academia. To this end, we propose the hybrid robust portfolio models under ellipsoidal uncertainty sets in this paper, where both the best-case and the worst-case counterparts are involved. In the suggested models, we introduce a trade-off parameter to adjust the portfolio optimism level. Machine learning algorithms including Long Short-Term Memory (LSTM) and eXtreme Gradient Boosting (XGBoost) are used to evaluate the potential market movements and provide forecasting information to generate the hyperparameter for modeling. Additionally, we develop a clustering-based algorithm for properly constructing joint ellipsoidal uncertainty sets to reduce conservatism further. In the modeling phase, we design the hybrid portfolios based on variance (HRMV) and value at risk (VaR) and prove the equivalent relationship between the hybrid robust mean-VaR model (HRMVaR) and the hybrid robust mean-CVaR (conditional value at risk) according to the existing research. The US 12 industry portfolio data set retrieved from Kenneth R. French is employed for the in-sample and out-of-sample numerical experiments. The experimental results demonstrate the effectiveness and robustness of the proposed portfolios, where HRMV models have better Sharpe ratios and Calmar ratios than the corresponding nominal mean–variance model, and HRMVaR models outperform the baseline VaR-based portfolios in terms of returns. Sensitivity analysis supports the superiority of the joint ellipsoidal uncertainty set U_δ^2, where the proposed portfolios constrained with U_δ^2 show stable risk characteristics.

1. Introduction

The seminal mean–variance (MV) portfolio model proposed by Markowitz [1] provided the essential foundation for modern portfolio theory. Since then, numerous quantitative methods based on MV have been developed for making portfolio decisions [2–5]. With the rapid development of computer science, however, further studies have found several problems within this classical theory: MV model is sensitive to perturbations of the observations [6,7], that is, when the estimation of market parameters, especially for expected returns [6], is error-prone, the model is severely handicapped. Also, poor out-of-sample performance is a notorious issue that restricts the widespread use of the MV model in practice. Besides, the financial crisis in 2008 again reminds investors of the importance of risk diversification. Unfortunately, the resulting optimized allocations of MV models are often highly concentrated, which may cause significant losses when extreme events occur. As a result, developing robust portfolio models with satisfactory returns has been one of the hot topics discussed by scholars.

1.1. Research motivation

To make up for the deficiencies of the classical MV model, some sophisticated risk measures have been applied in portfolio formation. VaR is one of the intuitive and robust downside risk measures that have been written into the Basel II Accord. Although VaR has been widely used in industry and academia, there still exists a good deal of criticism for being sub-additive, non-convex, and computationally intractable to optimize [8,9]. CVaR is also a downside risk measure but satisfies the axiomatic properties of coherent risk measures proposed by Artzner et al. (1999) [10]. Due to the coherence and computational tractability [9], Basel III shifted from VaR to CVaR to capture tail risk. However, it has been verified that CVaR is not a robust risk measure, which performs unstable in out-of-sample data sets [11]. The requirement of risk management and the dynamic, complex, and uncertain nature of the financial market motivates more and more researchers dedicated to improving the robustness of portfolio out-of-sample performance.

* Corresponding author.
E-mail addresses: minux@163.sufe.edu.cn (L. Min), hightempler@126.com (J. Dong), majorliujw@gmail.com (J. Liu), gongxiaomin961@163.com (X. Gong).

https://doi.org/10.1016/j.asoc.2021.107948
1568-4946/

Robust optimization provides a feasible solution to construct portfolio models with robust out-of-sample performance. Even when the input data is uncertain and ambiguous, the constraints in robust models are still required to be satisfied, and the objective value to be insensitive to data ambiguity. Although robust optimization is a powerful technique for stabilizing the out-of-sample performance, its solutions could be over-conservatism for most investors [12]. The main reason for the conservative nature of robust solutions is that all constraints should be strictly satisfied in different scenarios, which is not the case in practice. Theoretically, the essence of the modern portfolio theory is to balance return and risk, which is a bi-criteria optimization. Practically, many investors with high risk appetites prefer to pursue higher returns while assuming more risk, which constitutes the motivation to design less conservative robust portfolio models. Accordingly, we aim to develop the hybrid robust portfolio models for variance and VaR under general ellipsoidal uncertainty sets, where the best-case of robust optimization is considered to reduce the conservatism. Meanwhile, forecasting information provided by machine learning algorithms is also utilized to improve the proposed models' performance further.

1.2. Literature review

Our work touches on the previous works reviewed blow, and some existing results are extended in this paper. Literature research reveals that Ben-Tal & Nemirovski (1998) [13] proposed a robust convex optimization model considering an ellipsoidal uncertainty set for data ambiguity. This original work also provides insights and inspirations for the birth of robust portfolios with multiple risk measures. Goldfarb & Iyengar (2003) [14] developed robust portfolios based on factor models, which present stable out-of-sample performance. Ghaoui et al. (2003) [15] proposed a tractable robust VaR portfolio model in the manner of conic programming. Zhu et al. (2009) [16] investigated robust optimization methods for MV and mean-CVaR models, where the estimation risk for expected returns are considered and the corresponding min–max robust models are constructed. On the basis of [14], Zhu & Fukushima (2009) [17] proposed the robust CVaR portfolio where partial information on the underlying distribution is known. As [14,15], the worst-case scenario is only considered to keep the robustness of models.

Despite the relatively stable performance of the robust portfolio models, over-conservatism has raised heated discussions in academia. Bertsimas & Sim (2004) [18] firstly explored the potential conservatism of robust optimization from a theoretical point of view and guided the direction to fix this issue. Lotfi et al. (2017) [19] proposed a hybrid strategy that considering both worst-case and best-case simultaneously to reduce the conservatism of the corresponding robust portfolio model. This is also one of the works extended in this paper. Chen et al. (2020) [20] constructed the best-case robust portfolio for energy market, which again verified the effectiveness of best-case counterpart. Constructing feasible and compact uncertainty set for parameters is also a valid method to reduce conservatism. Chen et al. (2007) [21] and Natarajan et al. (2008) [22] proposed and applied the asymmetric ellipsoidal uncertainty set for portfolio formation. Lotfi & Zenios (2018) [7] developed analytical and heuristic algorithms for constructing ellipsoidal uncertainty sets, which laid the foundation for the proposed algorithms in Section 3.

With the significant progress of artificial intelligence, more accurate predictions could be obtained by machine learning algorithms, which are beneficial to building less conservative portfolios. Compared to the traditional approaches such as statistical methods, time series analysis, and econometrics models, the state-of-art artificial intelligent techniques, machine learning, and deep learning models have distinct advantages on nonlinear modeling [23]. These learning algorithms are being gradually used in portfolio construction, trading strategy design, and other engineering fields [24–29]. For example, Paiva et al. (2019) [30] used a fusion approach of support vector machine (SVM) and MV for portfolio decision-making, and the empirical research on the Brazilian market illustrates the usefulness of the novel method. Altan & Karasu (2019) [31] presented the reliable and accurate forecasting performance of SVM with different kernels on financial exchange rates. Chen et al. (2020) [32] proposed a hybrid approach to solve the portfolio with higher-order moments, where the two-stage clustering, neural network, and genetic algorithm (GA) are included. Sarmento & Horta (2020) [33] employed an unsupervised learning algorithm and LSTM to construct and manage trading pairs, which achieved a higher Sharpe ratio than standard approaches. Borges & Neves (2020) [34] fed technical indicators of cryptocurrencies into four machine learning algorithms: Logistic Regression (LR), Random Forest (RF), SVM, and Gradient Boosting Tree, to generate trading signals. Wang et al. (2020) [3] proposed an MV portfolio model in conjunction with LSTM as the asset preselection. Their model outperforms the selected baseline strategies by a clear margin. Chen et al. (2021) [4] also implemented an MV portfolio model with their optimized XGBoost as preselection. Ma et al. (2021) [5] constructed MV and Omega portfolios based on the prediction results generating from multiple deep learning models. These studies detailed investigate the superiority of artificial intelligence techniques on forecasting but ignoring to overcome the conservatism of robust portfolio models through machine learning algorithms and deep learning algorithms. As a result, to fill this gap, LSTM and XGBoost are employed to compute the trade-off parameter between the worst-case counterpart and the best-case counterpart of the proposed hybrid portfolio models.

1.3. Necessity of research

Existing literature shows that most studies concentrate solely on improve the robustness of the classical models or constructing traditional models based on forecasting information. However, few studies are committed to incorporating the forecasting information into the structure of the robust models, that is, adjusting the structure of the portfolio models according to the expectations of the future market conditions. The main purpose of this study is to overcome the conservatism of the classical robust portfolio models while maintaining the inherent robustness to large extent. Artificial intelligence exhibits reliable ability of pattern recognition, which reminds us of learning from historical samples and generating predictions for tackling the issue of over-conservatism. Although many scholars have done remarkable works on reducing the conservatism from the perspective of operational research, it is still necessary to combine the portfolio formation with the forecasting information provided by machine learning algorithms.

1.4. Novelty and main contribution

This paper contributes to develop the hybrid robust portfolio models that considering both the best-case counterpart and the worst-case counterpart. As Schöttle & Werner (2009) [35], and Lotfi & Zenios (2018) [7], the joint ellipsoidal uncertainty set is involved in this paper, and we derive the corresponding tractable optimization models as the second-order cone programming (SOCP). Table 1 compares the proposed models with current studies on portfolio construction in general. The key innovations of our work are three-fold:

Table 1
Comparison with existing related work.

Attribute	[19]	[7]	[30]	[32]	[3]	[4]	[5]	Proposed model
MV	✗	✗	✓	✓	✓	✓	✓	✓
mean-VaR	✓	✓	✗	✗	✗	✗	✗	✓
mean-CVaR	✗	✓	✗	✗	✗	✗	✗	✓
Worst-case	✓	✓	✗	✗	✗	✗	✗	✓
Best-case	✓	✗	✗	✗	✗	✗	✗	✓
LSTM	✗	✗	✗	✗	✓	✗	✓	✓
XGBoost	✗	✗	✗	✗	✗	✓	✗	✓
Sliding window	✓	✓	✓	✗	✓	✗	✓	✓
Data	Sampling	Sovereign CDS	Ibovespa	SSE 50	Sampling	SSE 50	CS100	US industry

Acronyms:- SSE: Shanghai stock exchange, CS: China Securities.

Table 2
Involved techniques in this study.

Tech.	Details
SOCP	Construct the fundamental and the derived hybrid mean-VaR, mean-CVaR models.
VaR	Value at risk, a robust risk measure describing tail risk.
CVaR	Conditional value at risk, a coherent risk measure describing tail risk.
Clustering	Develop the algorithm for describing the ellipsoidal uncertainty set.
LSTM & XGBoost	Evaluate the future market condition and compute the trade-off parameter.

- Both the hybrid robust mean–variance model and the hybrid mean-VaR model are developed. We also consider the feasibility of developing the hybrid robust mean-CVaR model, the relevant theoretical proofs of model equivalence according to existing research are supplemented.
- The method for tuning the trade-off parameter between the best-case counterpart and worst-case counterpart in the proposed models is explored in-depth. Instead of the direct linear search or grid search, we design a fusion approach based on the predictions provided by LSTM and XGBoost. In addition, a clustering algorithm is developed to construct a feasible and compact ellipsoidal uncertainty set.
- Detailed in-sample and out-of-sample experiments are implemented, and multiple indicators of risk and return are provided for further intuitive comparison, which provides practical guidance for individual investors in making reasonable decisions.

To the best of our knowledge, this contribution is novel in the literature. Some optimization methods of operational research, financial risk measures, and machine learning algorithms are involved in this work. Table 2 summarizes the required techniques to conduct this study and Fig. 1 outlines the flowchart of this work.

1.5. Organization of the paper

The remainder of our paper is organized as follows. We construct the hybrid robust models with necessary proofs in Section 2. The clustering algorithm for describing ellipsoidal uncertainty sets is presented in Section 3. In Section 4, LSTM and XGBoost are introduced for learning the trade-off parameter in the proposed hybrid robust models. In-sample and out-of-sample experiments are designed and implemented in Section 5. Also, related indicators and detailed analyses are provided. Finally, we summarize our main findings and outline future works in Section 6.

2. Hybrid robust models

Considering the worst-case of the robust optimization problem is a widely used method to obtain feasible and robust solutions [14,15,17,18,36–38]. Existing studies indicate that the worst-case optimization presents the most conservative solution, whereas the best-case optimization on the opposite shows the least conservative one [19,20,37,39]. One of the feasible approaches implemented by Lotfi et al. (2017) [19] is to construct an acceptable combination of the corresponding objective functions. Following the work of Lotfi et al. [7,19], we consider the convex combination of the worst-case objective function and the best-case objective function by introducing a trade-off parameter β. The hybrid robust models developed in this paper seek to keep the solution and structure robust while reducing the conservatism.

2.1. Ellipsoidal uncertainty

The ellipsoidal uncertainty is an intuitive set that has been frequently used in robust programming and stochastic programming [7,14,17,19,21,22,35,40–42]. Particularly, Chen et al. (2007) [21] and Natarajan et al. (2008) [22] elaborated on the asymmetric ellipsoidal uncertainty set, which generalizes the form of confidence ellipsoids. Two conventional ellipsoidal uncertainty sets, U_δ^1 and U_δ^2, are taken into account and are presented as follows.

2.1.1. Confidence ellipsoid U_δ^1

Suppose that asset returns have a joint normal distribution, and the mean estimate r can be calculated based on i.i.d samples of size S for n risky assets. Given covariance matrix Σ, we have

$$\frac{S(S-n)}{(S-1)n}(r-\bar{r})^T\Sigma^{-1}(r-\bar{r}) \sim \chi_n^2 \tag{1}$$

Define the associated ellipsoidal uncertainty, U_δ^1, around the point estimator $\hat{r}$ (with Σ fixed) as follows:

$$U_\delta^1 = \{r \in \mathbb{R}^n \mid (r-\hat{r})^T\Sigma^{-1}(r-\hat{r}) \le \delta^2\} \tag{2}$$

where $\chi_n^2(\delta^2) = \theta$, and $\theta \in (0, 1)$ is the desired confidence, that is, $\mathbb{P}(r \in U_\delta^1) = \theta$.

2.1.2. Confidence ellipsoid U_δ^2

U_δ^2 extends U_δ^1 to a joint uncertainty set for (r, Σ). Assuming that the sample estimators $\hat{r}$ and $\hat{\Sigma}$ are independent and their

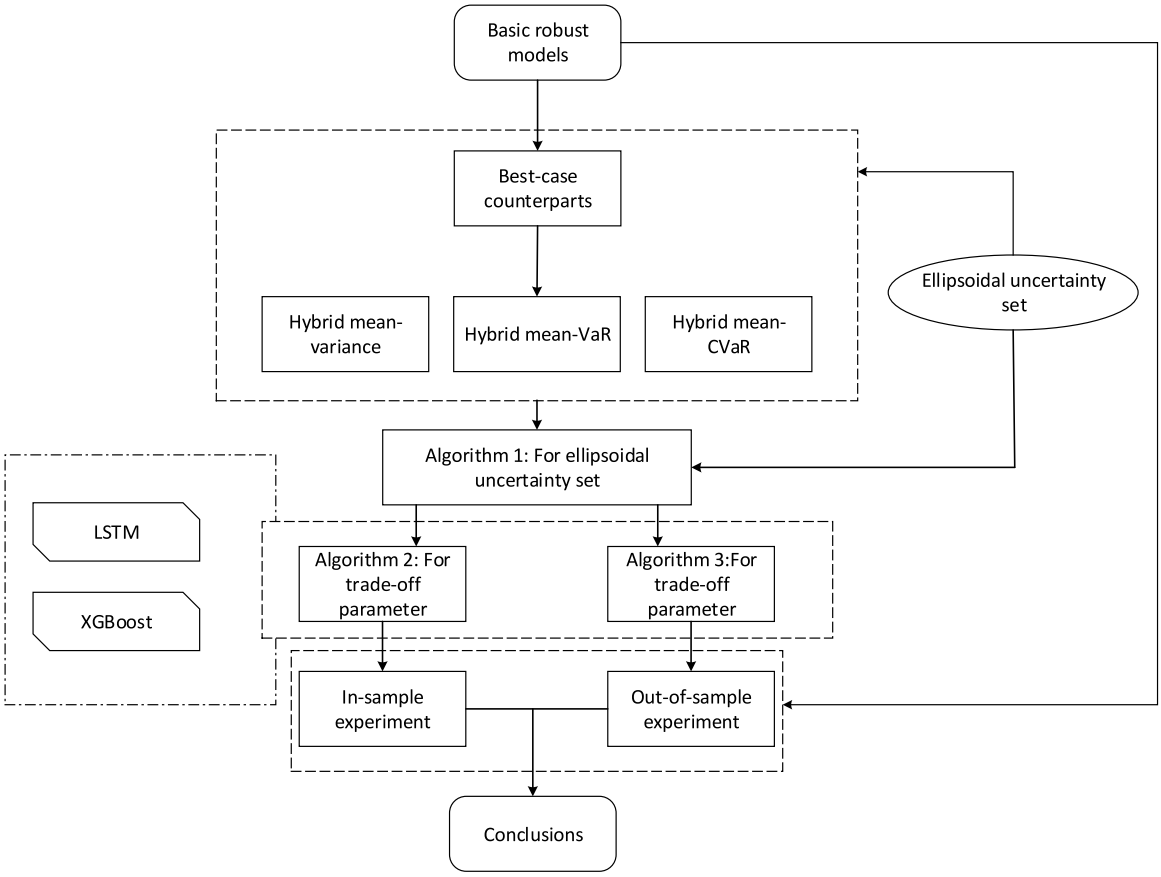


Fig. 1. Flowchart of this research.

distributions are as follows:

$$\begin{cases} \hat{r} \sim \mathcal{N}(r, \dfrac{\Sigma}{S}) \\ \hat{\Sigma} \sim \mathcal{W}(\dfrac{\Sigma}{S-1}, S-1) \end{cases}$$

where

$$\begin{cases} \hat{r} = \dfrac{1}{S}\sum_{i=1}^{S} r_i \\ \hat{\Sigma} = \dfrac{1}{S-1}\sum_{i=1}^{S}(r_i - \hat{r})(r_i - \hat{r})^T \end{cases} \tag{3}$$

where $\mathcal{N}(\mu, \sigma^2)$ is the Gaussian distribution, $\mathcal{W}(G, \nu)$ is the Wishart distribution with scale matrix G and ν degrees of freedom. Schöttle & Werner (2009) [35] derived an appropriate joint confidence ellipsoid U_δ^2 for estimated pair $(\hat{r}, \hat{\Sigma})$ as follows:

$$U_\delta^2 = \{(r, \Sigma) \in \mathbb{R}^n \times \mathbb{S}^n \mid S(r - \hat{r})^T \hat{\Sigma}^{-1}(r - \hat{r}) + \frac{S-1}{2}\| \hat{\Sigma}^{-1/2}(\Sigma - \hat{\Sigma})\hat{\Sigma}^{-1/2}\|_F^2 \le \delta^2\} \tag{4}$$

where $\|A\|_F^2 = \mathrm{tr}(AA^T) = \mathrm{tr}(A^TA)$. For simplicity, Lotfi et al. (2017) defined $K = \frac{\delta}{\sqrt{S}}$ representing the parameter of the ellipsoidal uncertainty.

2.2. Hybrid robust mean-VaR model

VaR optimization problem was firstly proposed by Basak & Shapiro (2001) [43]. By now, several methodological approaches have been developed for solutions. Parametric methods, Historical simulations, and Monte Carlo are three well-known approaches summarized in Table 3. In this paper, we would build the hybrid robust models based on the parametric methods. Also, we would implement the following mixed integer linear programming (MILP) formulation as one of the benchmarks for comparative analysis:

$$\min_{x\in\mathbb{X}, z\in\mathbb{R}, y\in\{0,1\}^S} z \quad s.t. \begin{cases} r^T x \ge \mu \\ -Rx - My - \mathbf{1}z \le 0 \\ \mathbf{1}^T y \le S(1-\alpha) \end{cases} \tag{5}$$

where R is a $S \times n$ matrix representing the return observations, M is a sufficiently large positive number, $\mathbf{1}$ is a vector of all ones, $\mathbb{X}$ denotes the set $\{x \in \mathbb{R}^n \mid x \ge 0, \|x\|_1 = 1\}$, r is the sample mean of portfolio returns, μ is the target return, $\alpha \in (0, 1)$ is a confidence level, e.g., $\alpha = 0.95$, and y denotes an auxiliary binary vector.

2.2.1. Uncertainty set $U = U_\delta^1$

Under the normality assumption, the SOCP formulation of mean-VaR model is as follows:

$$\min_{x\in\mathbb{X}} -r^T x + (1-\lambda)F^{-1}(\alpha)\|\Sigma^{1/2}x\| \quad s.t. \quad r^T x \ge \mu \tag{6}$$

where $F^{-1}(\alpha)$ is the normal inverse cumulative distribution function, $\lambda \in [0, 1]$ is the trade-off parameter between risk and return. According to the robust counterparts given by [19,35], we

Table 3
Methods for calculating VaR.

Type	Method	Attributes	Formulations	References
Analytical	Parametric	Normality assumption; Ignore higher-order moments; Easy to implement	SOCP	[7,14,19,35]
Simulation	Historical	No strong assumption; Depend heavily on historical data	MILP	[44,45]
	Monte Carlo	No distributional assumption; Large number of scenarios; Computationally challenging	MILP	[46–48]

can derive the best-case optimization as follows:

$$\min_{x\in\mathbb{X}} \min_{r\in U_\delta^1} -r^T x + (1-\lambda)F^{-1}(\alpha)\|\Sigma^{1/2}x\| \tag{7}$$

and the worst-case optimization as follows:

$$\min_{x\in\mathbb{X}} \max_{r\in U_\delta^1} -r^T x + (1-\lambda)F^{-1}(\alpha)\|\Sigma^{1/2}x\| \tag{8}$$

Hence, the proposed hybrid robust mean-VaR in U_δ^1 can be reformulated by introducing another trade-off parameter β as follows:

$$\min_{x\in\mathbb{X}} \Bigg[\beta\Big(\min_{r\in U_\delta^1} -r^T x + (1-\lambda)F^{-1}(\alpha)\|\Sigma^{1/2}x\|\Big) + (1-\beta)\Big(\max_{r\in U_\delta^1} -r^T x + (1-\lambda)F^{-1}(\alpha)\|\Sigma^{1/2}x\|\Big)\Bigg] \tag{9}$$

Lemma 1. *Let $U = U_\delta^1$ defined above, the best-case counterpart (7) can be reformulated as follows:*

$$\min_{x\in\mathbb{X}} -\hat{r}^T x + \left((1-\lambda)F^{-1}(\alpha) - \frac{\delta}{\sqrt{S}}\right)\|\Sigma^{1/2}x\| \tag{10}$$

Lemma 2 (*Lotfi et al. 2017*)**.** *Let $U = U_\delta^1$ defined above, the worst-case counterpart (8) can be reformulated as follows:*

$$\min_{x\in\mathbb{X}} -\hat{r}^T x + \left((1-\lambda)F^{-1}(\alpha) + \frac{\delta}{\sqrt{S}}\right)\|\Sigma^{1/2}x\| \tag{11}$$

Proof. From the ellipsoidal uncertainty U_δ^1, we have:

$$(r-\hat{r})^T\Sigma^{-1}(r-\hat{r}) \le (\frac{\delta}{\sqrt{S}})^2 \tag{12}$$

Due to Σ is a positive symmetric matrix, that is, $\Sigma^{-1} \succ 0$ and Σ^{-1} is a positive symmetric matrix. The Cholesky decomposition can be used, then we have:

$$\begin{aligned}
&\|(r-\hat{r})\Sigma^{1/2}\|^2 \le (\frac{\delta}{\sqrt{S}})^2\\
\Rightarrow &-\frac{\delta}{\sqrt{S}} \le \|(r-\hat{r})\Sigma^{-1/2}\| \le \frac{\delta}{\sqrt{S}}\\
\Rightarrow &-\frac{\delta}{\sqrt{S}}\|\Sigma^{1/2}\| \le \|(r-\hat{r})\| \le \|(r-\hat{r})\Sigma^{-1/2}\|\|\Sigma^{1/2}\|\\
&\le\frac{\delta}{\sqrt{S}}\|\Sigma^{1/2}\| \quad \text{(Cauchy–Schwarz)}\\
\Rightarrow &-\frac{\delta}{\sqrt{S}}\|\Sigma^{1/2}x\| \le (r-\hat{r})^T x \le \|(r-\hat{r})^T x\| \le \frac{\delta}{\sqrt{S}}\|\Sigma^{1/2}x\|\\
\Rightarrow &\hat{r}^T x - \frac{\delta}{\sqrt{S}}\|\Sigma^{1/2}x\| \le r^T x \le \hat{r}^T x + \frac{\delta}{\sqrt{S}}\|\Sigma^{1/2}x\| \quad \square
\end{aligned}$$

According to Lemmas 1 and 2 proved above, we substitute $r^T x$ with the two endpoints representing optimistic (best-case) and pessimistic (worst-case) respectively, then obtain the formulation for the optimization problem (9) as follows:

$$\min_{x\in\mathbb{X}} \underbrace{-\hat{r}^T x + (1-\lambda)F^{-1}(\alpha)\|\Sigma^{1/2}x\|}_{\text{Nominal mean-VaR}} + \overbrace{\left((1-2\beta)\frac{\delta}{\sqrt{S}}\right)\|\Sigma^{1/2}x\|}^{\text{Adjusted part}} \tag{13}$$

which is convex for $0 \le \beta \le \frac{1}{2} + \frac{\sqrt{S}}{2\delta}F^{-1}(\alpha)(1-\lambda)$. By varying the trade-off parameter β, the different hybrid efficient frontiers can be obtained.

2.2.2. Uncertainty set $U = U_\delta^2$

To derive the hybrid robust models in U_δ^2, we first present and prove the following lemmas:

Lemma 3. *Let $U = U_\delta^2$, the worst-case counterpart of mean-VaR portfolio optimization problem can be reformulated as:*

$$\begin{aligned}
&\min_{x\in\mathbb{X}} \max_{(r,\Sigma)\in U_\delta^2} -r^T x + (1-\lambda)F^{-1}(\alpha)\|\Sigma^{1/2}x\|\\
&= \min_{x\in\mathbb{X}} -\hat{r}^T x + \left(\max_{k\in[0,1]} g_1(k)\right)\|\hat{\Sigma}^{1/2}x\|
\end{aligned} \tag{14}$$

$$\text{where } g_1(k) = \delta\sqrt{\tfrac{k}{S}} + (1-\lambda)F^{-1}(\alpha)\sqrt{\left(1+\delta\sqrt{\tfrac{2(1-k)}{S-1}}\right)}$$

Proof. Following the procedure of Schöttle & Werner, we define the following two uncertainty set with additional parameter $k \in [0, 1]$:

$$\mathcal{U}_{\sqrt{k}\delta} := \left\{ r \in \mathbb{R}^n \mid S(r-\hat{r})\hat{\Sigma}(r-\hat{r}) \le k\delta^2 \right\}$$

$$\mathcal{U}_{\sqrt{1-k}\delta} := \left\{ \Sigma \in \mathbb{S}^n \mid \frac{S-1}{2}\|\hat{\Sigma}^{-1/2}(\Sigma-\hat{\Sigma})\hat{\Sigma}^{-1/2}\|_F^2 \le (1-k)\delta^2 \right\}$$

the inner maximization of (14) is equivalent to

$$\max_{k\in[0,1]} \max_{\Sigma\in\mathcal{U}_{\sqrt{1-k}\delta}} \max_{r\in\mathcal{U}_{\sqrt{k}\delta}} -r^T x + (1-\lambda)F^{-1}(\alpha)\|\Sigma^{1/2}x\| \tag{15}$$

and the innermost maximization of (15) can be obtained at endpoint:

$$r^* = \hat{r} - \frac{\delta\sqrt{k}}{\sqrt{S}}\frac{\hat{\Sigma}x}{\sqrt{x^T\hat{\Sigma}x}}$$

Plugging in r^*, the next innermost maximization in (15) is as follows:

$$\max_{\Sigma\in\mathcal{U}_{\sqrt{1-k}\delta}} -\hat{r}^T x + (1-\lambda)F^{-1}(\alpha)\sqrt{x^T\Sigma x} + \frac{\delta\sqrt{k}}{\sqrt{S}}\sqrt{x^T\hat{\Sigma}x} \tag{16}$$

Simplifying the problem (16), we can obtain the $\tilde{\Sigma}^*$ by solving the following optimization problem:

$$\begin{aligned}
&\max_{\tilde{\Sigma}\in\mathbb{S}^n} y^T\tilde{\Sigma}y\\
&s.t. \quad \|\tilde{\Sigma}\|_F^2 \le b
\end{aligned} \tag{17}$$

where $y = \hat{\Sigma}^{1/2}x$, $b = \frac{2}{S-1}(1-k)\delta^2$, $\tilde{\Sigma} = \hat{\Sigma}^{-1/2}(\Sigma - \hat{\Sigma})\hat{\Sigma}^{-1/2}$. Using the cyclic property of trace, we have:

$$y^T\tilde{\Sigma}y = \mathrm{tr}(y^T\tilde{\Sigma}y) = \mathrm{tr}(y^Ty\tilde{\Sigma}) = \langle y^Ty, \tilde{\Sigma}\rangle \tag{18}$$

According to the Cauchy–Schwarz inequality, we have:

$$y^T\tilde{\Sigma}y = \langle y^Ty, \tilde{\Sigma}\rangle \le \|y^Ty\|_F\|\tilde{\Sigma}\|_F \le \sqrt{b}\|y^Ty\| = \sqrt{b}\|y\|^2$$

It is obvious that the maximal value is at the boundary, that is:

$$y^T\tilde{\Sigma}^*y = \sqrt{b}\|y\|^2$$

$$\tilde{\Sigma}^* = \sqrt{b}\frac{y}{\|y\|}\frac{y^T}{\|y\|}$$

$$\Sigma^* = \hat{\Sigma} + \sqrt{b}\hat{\Sigma}^{1/2}\frac{y}{\|y\|}\frac{y^T}{\|y\|}\hat{\Sigma}^{1/2}$$

Plugging Σ^* into (16), the optimization problem (14) can be reformulated as follows:

$$\begin{aligned}
&\min_{x\in\mathbb{X}}\max_{k\in[0,1]} -\hat{r}^Tx + (1-\lambda)F^{-1}(\alpha)\sqrt{(1+\sqrt{b})x^T\hat{\Sigma}x} \\
&+ \frac{\delta\sqrt{k}}{\sqrt{S}}\sqrt{x^T\hat{\Sigma}x} \\
&= \min_{x\in\mathbb{X}}\max_{k\in[0,1]} -\hat{r}^Tx + \underbrace{\left((1-\lambda)F^{-1}(\alpha)\sqrt{1+\sqrt{b}} + \frac{\delta\sqrt{k}}{\sqrt{S}}\right)}_{g_1(k)}\|\hat{\Sigma}^{1/2}x\| \\
&= \min_{x\in\mathbb{X}} -\hat{r}^Tx + \left(\max_{k\in[0,1]} g_1(k)\right)\|\hat{\Sigma}^{1/2}x\| \quad \square
\end{aligned} \tag{19}$$

Lemma 4 (*Lotfi et al. 2017*)**.** *Let $U = U_\delta^2$, the best-case counterpart of the mean-VaR portfolio optimization problem can be reformulated as:*

$$\begin{aligned}
&\min_{x\in\mathbb{X}}\min_{(r,\Sigma)\in U_\delta^2} -r^Tx + (1-\lambda)F^{-1}(\alpha)\|\Sigma^{1/2}x\| \\
&= \min_{x\in\mathbb{X}} -\hat{r}^Tx + \left(\min_{k\in[\max(0,1-\frac{S-1}{2\delta^2}),1]} g_2(k)\right)\|\hat{\Sigma}^{1/2}x\|
\end{aligned} \tag{20}$$

where $g_2(k) = -\delta\sqrt{\frac{k}{S}} + (1-\lambda)F^{-1}(\alpha)\sqrt{\left(1-\delta\sqrt{\frac{2(1-k)}{S-1}}\right)}$

Proof. The inner minimization problem of (20) is equivalent to

$$\min_{k\in[0,1]}\min_{\Sigma\in\mathcal{U}_{\sqrt{1-k}\delta}}\min_{r\in\mathcal{U}_{\sqrt{k}\delta}} -r^Tx + (1-\lambda)F^{-1}(\alpha)\|\Sigma^{1/2}x\| \tag{21}$$

The innermost minimization problem of (21) can be attained at endpoint

$$r^* = \hat{r} + \frac{\delta\sqrt{k}}{\sqrt{S}}\frac{\hat{\Sigma}x}{\sqrt{x^T\hat{\Sigma}x}}$$

Plugging in r^*, the next innermost minimization problem in (21) is:

$$\min_{\Sigma\in\mathcal{U}_{\sqrt{1-k}\delta}} -\hat{r}^Tx + (1-\lambda)F^{-1}(\alpha)\sqrt{x^T\Sigma x} - \frac{\delta\sqrt{k}}{\sqrt{S}}\sqrt{x^T\hat{\Sigma}x} \tag{22}$$

Taking the identical variable transformation as the proof of Lemma 3, we have:

$$\begin{aligned}
&\min_{\tilde{\Sigma}\in\mathbb{S}^n} y^T\tilde{\Sigma}y \\
&s.t. \quad \|\tilde{\Sigma}\|_F^2 \le b
\end{aligned} \tag{23}$$

Using the property of symmetry, the minimal value can be obtained as follows:

$$y^T\tilde{\Sigma}^*y = -\sqrt{b}\|y\|^2$$

$$\tilde{\Sigma}^* = -\sqrt{b}\frac{y}{\|y\|}\frac{y^T}{\|y\|}$$

$$\Sigma^* = \hat{\Sigma} - \sqrt{b}\hat{\Sigma}^{1/2}\frac{y}{\|y\|}\frac{y^T}{\|y\|}\hat{\Sigma}^{1/2}$$

Plugging Σ^* into (21), the optimization problem (20) can be reformulated as follows:

$$\begin{aligned}
&\min_{x\in\mathbb{X}}\min_{k\in[0,1]} -\hat{r}^Tx + (1-\lambda)F^{-1}(\alpha)\sqrt{(1-\sqrt{b})x^T\hat{\Sigma}x} \\
&- \frac{\delta\sqrt{k}}{\sqrt{S}}\sqrt{x^T\hat{\Sigma}x} \\
&= \min_{x\in\mathbb{X}}\min_{k\in[0,1]} -\hat{r}^Tx + \underbrace{\left((1-\lambda)F^{-1}(\alpha)\sqrt{1-\sqrt{b}} - \frac{\delta\sqrt{k}}{\sqrt{S}}\right)}_{g_2(k)}\|\hat{\Sigma}^{1/2}x\| \\
&= \min_{x\in\mathbb{X}} -\hat{r}^Tx + \left(\min_{k\in[\max(0,1-\frac{S-1}{2\delta^2}),1]} g_2(k)\right)\|\hat{\Sigma}^{1/2}x\| \quad \square
\end{aligned} \tag{24}$$

According to Lemmas 3 and 4 proved above, the hybrid robust counterpart for joint uncertainty set U_δ^2 is given by:

$$\begin{aligned}
&\min_{x\in\mathbb{X}}\Bigg[\beta\left(\min_{(r,\Sigma)\in U_\delta^2} -r^Tx + (1-\lambda)F^{-1}(\alpha)\|\Sigma^{1/2}x\|\right) \\
&+ (1-\beta)\left(\max_{(r,\Sigma)\in U_\delta^2} -r^Tx + (1-\lambda)F^{-1}(\alpha)\|\Sigma^{1/2}x\|\right)\Bigg] \\
&= \min_{x\in\mathbb{X}} -\hat{r}^Tx + \left(\beta\min_{k\in[\max(0,1-\frac{S-1}{2\delta^2}),1]} g_2(k) + (1-\beta)\max_{k\in[0,1]} g_1(k)\right) \\
&\times \|\hat{\Sigma}^{1/2}x\|
\end{aligned} \tag{25}$$

For values of β for which the coefficient of $\|\hat{\Sigma}^{1/2}x\|$ is non-negative, the problem (25) is convex and can be solved efficiently via several polynomial time algorithms.

2.3. Hybrid robust mean–variance model

The canonical mean–variance model was initially proposed by Markowitz (1952) [1], which has been widely and deeply studied by numerous scholars. Although this theoretical model has many well-known shortcomings, it still lays the theoretical foundation for our hybrid robust mean–variance model. In this section, we develop two hybrid robust mean–variance models in the ellipsoidal uncertainty U_δ^1 and U_δ^2 respectively. Because the general idea is similar to the hybrid mean-VaR models, some detailed calculations and proofs are omitted here.

2.3.1. Uncertainty set $U = U_\delta^1$

In the ellipsoidal uncertainty U_δ^1, the best-case counterpart is given by:

$$\min_{x\in\mathbb{X}}\min_{r\in U_\delta^1} -r^Tx + \lambda\|\Sigma^{1/2}x\| \tag{26}$$

and the worst-case optimization counterpart is given by:

$$\min_{x\in\mathbb{X}}\max_{r\in U_\delta^1} -r^Tx + \lambda\|\Sigma^{1/2}x\| \tag{27}$$

Using Lemmas 1 and 2, the hybrid robust mean–variance model constrained with U_δ^1 is given by:

$$\begin{aligned}&\min_{x\in\mathbb{X}}\Bigg[\beta\left(-\hat{r}^Tx-\frac{\delta}{\sqrt{S}}\|\hat{\Sigma}^{1/2}x\|+\lambda\|\hat{\Sigma}^{1/2}x\|\right)\\&+(1-\beta)\left(-\hat{r}^Tx+\frac{\delta}{\sqrt{S}}\|\hat{\Sigma}^{1/2}x\|+\lambda\|\hat{\Sigma}^{1/2}x\|\right)\Bigg]\\&=\min_{x\in\mathbb{X}}\underbrace{-\hat{r}^Tx+\lambda\|\hat{\Sigma}^{1/2}x\|}_{\text{Nominal mean–variance}}+\overbrace{\left((1-2\beta)\frac{\delta}{\sqrt{S}}\right)}^{\text{Adjusted part}}\|\hat{\Sigma}^{1/2}x\|\end{aligned}\tag{28}$$

2.3.2. Uncertainty set $U=U_\delta^2$

Considering the joint ellipsoidal uncertainty U_δ^2, the best-case counterpart is given by:

$$\min_{x\in\mathbb{X}}\min_{(r,\Sigma)\in U_\delta^2}-r^Tx+\lambda\|\Sigma^{1/2}x\|\tag{29}$$

and the worst-case counterpart is given by:

$$\min_{x\in\mathbb{X}}\max_{(r,\Sigma)\in U_\delta^2}-r^Tx+\lambda\|\Sigma^{1/2}x\|\tag{30}$$

Using Lemmas 3 and 4, the hybrid robust mean–variance model constrained with U_δ^2 is reformulated as follows:

$$\begin{aligned}&\min_{x\in\mathbb{X}}\Bigg[\beta\left(\min_{(r,\Sigma)\in U_\delta^2}-r^Tx+\lambda\|\Sigma^{1/2}x\|\right)\\&+(1-\beta)\left(\max_{(r,\Sigma)\in U_\delta^2}-r^Tx+\lambda\|\Sigma^{1/2}x\|\right)\Bigg]\\&=\min_{x\in\mathbb{X}}-\hat{r}^Tx+\left(\beta\min_{k\in[\max(0,1-\frac{S-1}{2\delta^2}),1]}g_4(k)+(1-\beta)\max_{k\in[0,1]}g_3(k)\right)\\&\times\|\hat{\Sigma}^{1/2}x\|\end{aligned}\tag{31}$$

where

$$g_3(k)=\delta\sqrt{\frac{k}{S}}+\lambda\sqrt{\left(1+\delta\sqrt{\frac{2(1-k)}{S-1}}\right)}$$

$$g_4(k)=-\delta\sqrt{\frac{k}{S}}+\lambda\sqrt{\left(1-\delta\sqrt{\frac{2(1-k)}{S-1}}\right)}$$

2.4. Hybrid robust mean-CVaR model

Compared with variance, CVaR is a more sophisticated risk metric that satisfies the axioms of coherent risk measures [10]. Related literature has shown that CVaR could effectively describe tail risk, which is full of significance especially when the financial market is under stress. In this section, we present the similar corollary as Lotfi et al. (2018) [7] under the assumption of normality.

Assuming that the random variable ξ represents the return rate of a certain risky asset following the Gaussian distribution, with the mean estimate and covariance matrix estimate are as follows:

$$\mathbb{D}=\{\pi\mid\mathbb{E}_\pi[\xi]=\mu,\mathrm{Cov}_\pi[\xi]=\Sigma\}$$

Corollary 1 (*Lotfi et al. 2018*)**.** *The robust counterparts of VaR (RVaR for short) and CVaR (RCVaR for short) constrained with ellipsoidal uncertainty U_δ^2 are identical optimization models.*

Proof. The RCVaR constrained with U_δ^2 is as follows [7,9]:

$$\min_{x\in\mathbb{X},\gamma\in\mathbb{R}}\max_{(r,\Sigma)\in U_\delta^2}F_\alpha(x,\gamma)\tag{32}$$

where

$$\begin{cases}F_\alpha(x,\gamma)=\gamma+\dfrac{1}{1-\alpha}\mathbb{E}\{[f(x,\xi)-\gamma]^+\}\\f(x,\xi)=-x^T\xi\end{cases}$$

The RVaR constrained with U_δ^2 can be expressed as the following general equation:

$$\begin{aligned}&\min_{\gamma\in\mathbb{R},x\in\mathbb{X}}\gamma\\&s.t.\quad\max_{(r,\Sigma)\in U_\delta^2}\mathrm{Prob}\{\gamma\le f(x,\xi)\}\le1-\alpha\end{aligned}\tag{33}$$

According to THEOREM 1 in Ghaoui et al. (2003) [15], the following equivalent relationship holds:

$$\sup\mathrm{Prob}\{\gamma\le f(x,\gamma)\}\le1-\alpha\tag{34}$$

is equivalent with

$$-r^Tx+\sqrt{\frac{\alpha}{1-\alpha}}\le\gamma\tag{35}$$

Further, suppose that the maximum exists, the following equation holds according to the Eq. (10) in [40]:

$$\min_{\gamma\in\mathbb{R}}\max_{(r,\Sigma)\in U_\delta^2}F_\alpha(x,\gamma)=-r^Tx+\frac{\sqrt{\alpha}}{\sqrt{1-\alpha}}\sqrt{x^T\Sigma x}\tag{36}$$

Hence, the right-hand side of Eq. (36) can be reformulated as the following optimization problem:

$$\begin{aligned}&\min_{\gamma\in\mathbb{R}}\gamma\\&s.t.\quad-r^Tx+\frac{\sqrt{\alpha}}{\sqrt{1-\alpha}}\sqrt{x^T\Sigma x}\le\gamma\end{aligned}\tag{37}$$

It can be observed that the left-hand side of Eq. (36) is RCVaR and the right-hand side of Eq. (36) is RVaR. As a result, the robust counterparts of VaR and CVaR under joint ellipsoidal uncertainty set U_δ^2 are identical. □

Corollary 2. *If random variable ξ has a distribution from the set $\mathbb{D}$, then the hybrid robust counterparts of VaR and CVaR are identical in the ellipsoidal uncertainty set U_δ^2.*

Proof. According to Corollary 1 and Lemma 4, we can derive Corollary 2 easily. □

Remark. From the proof of Lemma 3, we can observe that U_δ^1 is a special structure of U_δ^2, for which we omit the discussion constrained with U_δ^1 in this section. For more details about Corollary 1, we recommend [7] for referring. Since the equivalence between robust mean-VaR and robust mean-CVaR holds, we omit the hybrid mean-CVaR portfolio models in this paper.

3. Constructing ellipsoidal sets

The ellipsoidal uncertainty set can be defined as the confidence regions of the statistical estimators associated with the parameters [7,14,35]. Lotfi et al. (2018) proposed two pioneering and insightful algorithms to consider multiple opinions from experts and describe a joint ellipsoidal set. The first one is to solve the following nonlinear semi-definite programming (SDP) for obtaining the center, $(\hat{\mu},\hat{\Sigma})$, of a joint ellipsoidal set:

$$\min_{\hat{\mu}\in\mathbb{R}^n,\hat{\Sigma}\in\mathbb{S}_{++}^n}\sqrt{\sum_{i=1}^K S(\bar{\mu}_i-\hat{\mu})^T\hat{\Sigma}^{-1}(\bar{\mu}_i-\hat{\mu})+\frac{S-1}{2}\|\hat{\Sigma}^{-1/2}(\bar{\Sigma}_i-\hat{\Sigma})\hat{\Sigma}^{-1/2}\|_F^2}\tag{38}$$

where $(\bar{\mu}_1, \bar{\Sigma}_1), \ldots, (\bar{\mu}_K, \bar{\Sigma}_K)$ are K estimate pairs for mean return and covariance matrix generating from K experts. And the second one is to construct the ellipsoidal set heuristically. Essentially, the heuristic algorithm's goal is to find a center in the K estimate pairs that minimizes the total distance of the other pairs to the center. The two proposed algorithms indeed bring some valuable inspiration for constructing the associated uncertainty set. However, solving an SDP with lots of observations is a time-consuming task, and the desired center may not certainly be one of the estimate pairs. The two issues motivate us to develop a novel algorithm based on the existing methodologies, where appropriate center and radius would be efficiently reached at the expense of some accuracy.

Clustering is one of the machine learning techniques that have been applied in portfolio selection problems (see [49,50]). To obtain the desired center $(\hat{\mu}, \hat{\Sigma})$ and radius δ, we design Algorithm 1 based on clustering. The proposed approach combines the characteristics of the analytical and heuristic algorithms mentioned above. A relatively reasonable center and the associated ellipsoidal uncertainty set can be obtained via the proposed algorithm.

Algorithm 1 Constructing ellipsoidal uncertainty sets.

Input:
K estimates for mean returns and covariance matrices $\vec{x}_i = (\bar{r}_i, \bar{\Sigma}_i)$
m initialized centroids
The parameter of tolerance ϵ
Output:
Obtain the center $(\hat{r}, \hat{\Sigma})$ and radius δ
Define $C_i = \varnothing, i = \{i, 2, \ldots, m\}$
Choose random n centroids from K estimates, $\vec{c}_1, \ldots, \vec{c}_m$
while all of C_i are $\varnothing$ or $\vec{c}_i, i = 1, 2, \ldots, m$ is changed **do**
  for $i = 1, 2, \ldots, K$ **do**
    Compute the distance d_{ij} between estimate $(\vec{x}_i)$ and each centroid $\vec{c}_j, j = \{1, 2, \ldots, m\}$
    Label $\vec{x}_i$ according to the minimum distance $\lambda_i = \arg \min_{j \in \{1,2,\ldots,m\}} d_{ij}$
    Classify estimate $\vec{x}_i$: $C_{\lambda_i} = C_{\lambda_i} \cup \{\vec{x}_j\}$
  end for
  Check $C_i, i = \{1, 2, \ldots, m\}$
  for $j = 1, 2, \ldots, m$ **do**
    Compute a new centroid $\vec{c}_i' = \frac{1}{|C_i|} \sum_{\vec{x}_i \in C_i} \vec{x}$
    if $|\vec{c}_i' - \vec{c}_i| > \epsilon$ **then**
      Change $\vec{c}_i \to \vec{c}_i'$
    else
      Keep $\vec{c}_i$
    end if
  end for
end while
Compute the center of gravity of $\vec{c}_i, i = 1, 2, \ldots, m$ as the center of ellipsoid set, $(\hat{r}, \hat{\Sigma})$
Choose δ such that the resulting ellipsoid set can inscribe all of $(\bar{r}_i, \bar{\Sigma}_i)$. Compute the distance of each estimate from the center, $\delta_1, \ldots, \delta_K$, and let δ be the maximum value.

Remark. The clustering algorithm designed requires K observations, m centroids, and a $n \times n$ matrix with the inverse operation, the computational complexity is $\mathcal{O}(K(n^3 + m) + Kn^3)$. However, if we regard $\bar{r}_i$ and $\bar{\Sigma}_i$ as features from the perspective of machine learning, and use Euclidean distance in Algorithm 1, the complexity is $\mathcal{O}(Km + Kn^3)$. Additionally, the K estimate pairs can be obtained by bootstrap sampling. Fig. 2 illustrates analytical, heuristic and our proposed algorithm ellipsoids in two dimensions. In the follow-up numerical experiments, we use δ_1 representing the δ of ellipsoidal uncertainty U_δ^1, δ_2 indicating the δ of confidence ellipsoid U_δ^2. As shown in Fig. 6, the hybrid robust portfolio models are constructed on the ellipsoidal uncertainty generated by Algorithm 1.

4. Computing adjusted parameter β

Line search, as well as grid search, is a widely used approach for finding appropriate hyperparameters. (e.g. [19]). However, with the development of forecasting theory, multiple algorithms and tools are developed for analyzing future information. Opposite to the brute methods as line search or grid search, machine learning provides several elegant algorithms for forecasting without consuming too much computational resource. In the proposed hybrid robust model, the trade-off parameter β, representing the level of optimism about the future, can be computed by analyzing the historical pattern. According to the assumption of Fama (1965) [51], the historical trend of the price in an individual asset is inclined to repeat in the future. Hence, it is promising to employ machine learning algorithms to infer the possible trend of risky assets and provide convincing evidence for determining the trade-off parameter. In this section, a fusion method involved LSTM and XGBoost is offered to compute the hyperparameter β.

4.1. LSTM

LSTM is an alternative neural network model devised by Hochreiter & Schmidhuber (1997) [52] to learn sequential patterns. In essence, LSTM is one of the classes of existing recurrent neural networks (RNNs), but the insightful design of memory block (see Fig. 3) makes LSTM could utilize the clues over a long period, while naive RNNs would ignore this stale information [3,53–55]. Dropout [56] is developed on the hidden layer to overcome overfitting, the inherent shorting of neural networks. To be specific, a certain fraction of neurons would be randomly selected and masked, as well as the corresponding connections between input layer and output layer. Additionally, due to the memory block structure in LSTM, the problem of gradient explosion/vanishing in back-propagation is surmounted to some extent.

Assuming at moment t, x_t represents the state of input layer, h_t represents the state of hidden layer, an LSTM memory block is composed of three gates: input gate i_t, output gate o_t, forget gate g_t, and a memory cell c_t. Table 4 reveals the rest of symbols we defined. The state of a memory block are calculated as Eq. (39). For more details and code snippets about LSTM, we recommend [54,55,57,58] for referring.

$$\begin{aligned} g_t &= \sigma(U_g x_t + W_g h_{t-1} + b_g) \\ i_t &= \sigma(U_i x_t + W_i h_{t-1} + b_i) \\ c_t^* &= \sigma(U_c x_t + W_c h_{t-1} + b_c) \\ c_t &= g_t \otimes c_{t-1} + i_t \otimes c_t^* \\ o_t &= \sigma(U_o x_t + W_o h_{t-1} + b_o) \\ h_t &= o_t \otimes \tanh(c_t) \end{aligned} \tag{39}$$

The excellent performance of LSTM has already attracted extensive attention from both academics and industries [55]. Kim & Won (2018) [55] built a hybrid model by combining LSTM with multiple econometric models to forecast the volatility of stock price; Fischer & Krauss (2018) [54] derived a trading strategy based on the large-scale financial market predictions coming from LSTM, and demonstrated the efficiency of LSTM in extracting

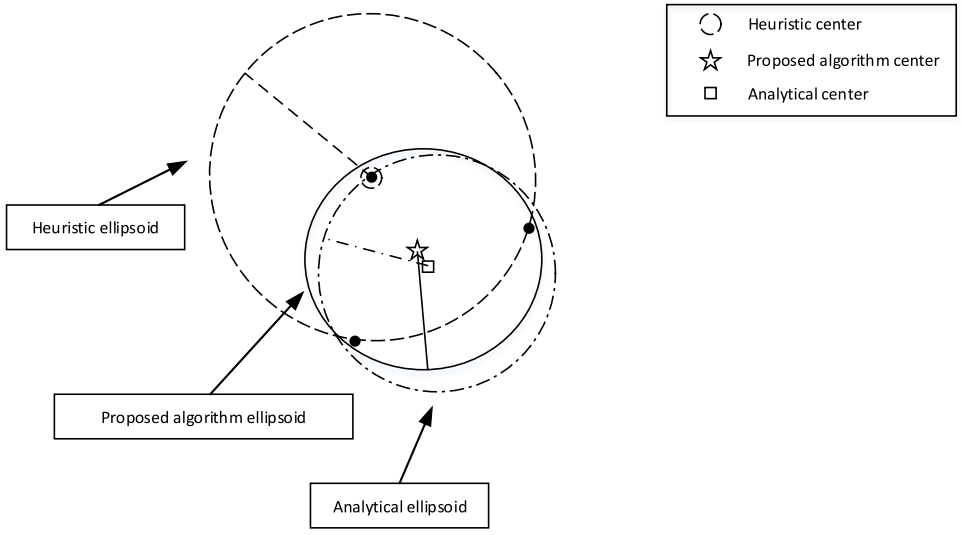


Fig. 2. The proposed algorithm ellipsoid, heuristic, and analytical ellipsoids proposed by Lotfi et al. (2018).

Table 4
Symbols of the LSTM we used in this work.

Symbols	Description
$\otimes$	Element-wise multiplication
U	Weight matrix
W	Weight matrix
b	Bias term
$\sigma(\cdot)$	Sigmoid activate function
$\tanh(\cdot)$	Hyperbolic tangent activate function
c_t^*	Input modulate gate
h_t	Output of memory block

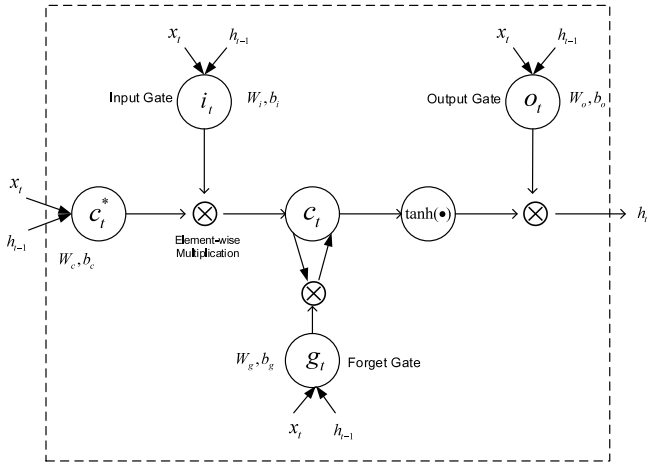


Fig. 3. Structure of an LSTM memory block.

meaningful information from noisy financial data; Wang et al. (2020) [3] employed LSTM preselecting potential stocks then built mean–variance portfolio model; Altan et al. (2021) [24] developed a non-linear hybrid model to provide reliable and accurate forecasting for wind speed, where LSTM is an indispensable component; Ma et al. (2021) [5] constructed omega model and mean–variance model based on forecasting results given by SVM regression (SVR), RF, LSTM, etc. Accordingly, LSTM is an essential component of the proposed fusion method to evaluate the optimism for the future market.

4.2. XGBoost

Boosting aims to convert a weak learning algorithm into one that achieves arbitrary high accuracy [59]. Several boosting-based machine learning models have been developed and applied in practice [60–62]. XGBoost [63] is one of the scalable tree systems optimized based on the gradient boosting decision tree (GBDT). Slightly different from GBDT, XGBoost merges the steps of the partition for J child nodes and optimization for each sub-partition into one procedure. The loss function of XGBoost is as follows:

$$L_t = \sum_{i=1}^{n} L(y_i, \hat{y}_i^{t-1} + f_t(x_i)) + \underbrace{\gamma J + \frac{\lambda}{2}\sum_{j=1}^{J} w_{tj}^2}_{\Omega(f_t)} \tag{40}$$

where the penalty term $\Omega(f_t)$ is added to curb overfitting, L is a differential convex function, γ is the l_1 regularization coefficient, λ is the l_2 regularization coefficient, $\hat{y}_i^{t-1}$ is the output of the ith instance at the $(t-1)$th iteration.

According to the second-order Taylor expansion of Eq. (40), we obtain the following loss function:

$$\begin{aligned} L_t &\approx \sum_{i=1}^{n}\left[L(y_t, \hat{y}_i^{t-1} + g_i f_t(x_i) + \frac{1}{2}h_i f_t^2(x_i))\right] + \Omega(f_t) \\ g_i &= \partial_{\hat{y}^{(t-1)}} L(y_i, \hat{y}_i^{t-1}) \\ h_i &= \partial^2_{\hat{y}^{(t-1)}} L(y_i, \hat{y}_i^{t-1}) \end{aligned} \tag{41}$$

The construction process of XGBoost is presented in Fig. 4. As shown, the prediction error in XGBoost is inversely proportional to the number of decision trees. XGBoost takes the greedy strategy to obtain the best split points in the growth of each decision tree, and the issue of local minima could be avoided to a large extent in practice.

Currently, XGBoost has been paid much attention from financial fields as well as other industries. Dey et al. (2016) [64] used XGBoost to predict the trend of stock market; Chen et al. (2021) [4] constructed a mean–variance portfolio model based on the forecasted results provided by XGBoost where hyperparameters are optimized by an improved firefly algorithm; Lu et al. (2020) [65] employed XGBoost to develop mill vibration

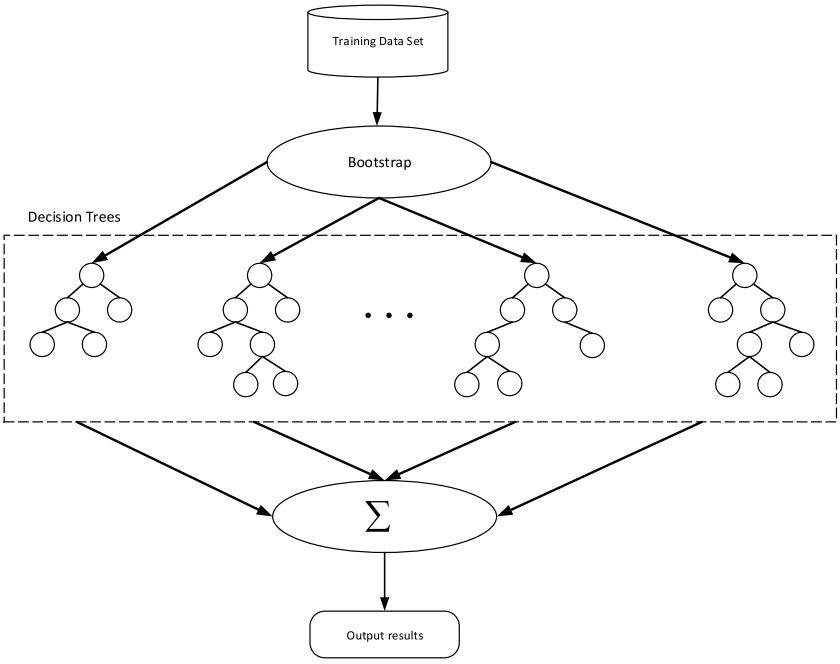


Fig. 4. Schematic diagram of XGBoost.

monitoring scheme; Nobre & Neves (2019) [66] presented a trading strategy where the XGBoost optimized by a multi-objective Genetic algorithm is utilized to detect trading signals.

4.3. Training & computing

Here we introduce the procedure for training and computing the trade-off parameter β. In the proposed hybrid robust model, β is defined to reflect the level of optimism of the market. To this end, we retrieve the S&P 500 data for representing the whole market. If there exists three consecutive return rates of S&P 500 (calculated by adjusted close prices), r_{t-2}, r_{t-1}, r_t, satisfy that $r_t > r_{t-1} > r_{t-2}$, then we can label the state of day t as $+1$ (otherwise, 0), which indicates a signal of optimism. Such a method refers to the idea of [67,68], estimating the state of market from the perspective of the stock price pattern. Forecasting can be implemented by shifting the label sequence, that is, we can match the indicators of day t with the label of day $t + k$. Existing studies demonstrate that some technical indicators could provide insightful information for predictions. For example, exponential moving average (EMA) [69–71], average true range (ATR) [72,73], and on balance volume (OBV) [74–76], etc. In this work, we select 24 indicators as the inputted features, including return rates of 12 industries from Kenneth R. French database, 6 indicators of price and volume, 3 technical indicators, and 3 lagged return observations, the detailed information of selected indicators is shown in Table 5. In order to improve the accuracy of prediction, all inputted features are standardized at the preprocessing phase.

Particularly, we design two different schemes for in-sample and out-of-sample experiments respectively. For the in-sample numerical experiment, the observations can be directly fed into the machine learning models to obtain predicted labels, then calculate the value of β statistically. Algorithm 2 reveals the details of this scheme. However, for the out-of-sample experiment, a dynamical sliding window strategy (see Fig. 5) is applied, that is, in each period beginning (except the first one), we would drop the oldest data points and add the most recently up-to-date observations to train the machine learning models, then calculate β for this period based on the predictions given by the well-trained models. Algorithm 3 illustrates the details of this dynamic scheme.

Algorithm 2 Calculating β: in-sample.

Input:
$N \times K$ selected indicators representing state of the market
$(N + k) \times 1$ labels representing signals of optimism
Output:
The estimated value of β
Preprocessing: standardize all selected indicators
Shift labels k units to make indicators of day t match up with label of day $t + k$
Feed processed data set into LSTM/XGBoost
Retrieve predicted labels from trained LSTM/XGBoost
Estimate the value of β as follows:

$$\beta = \frac{\sum \text{labels}}{N} \tag{42}$$

Algorithm 3 Calculating β: out-of-sample.

Input:
Historical data set D representing state of the market
$\beta = \emptyset$
Output:
The estimated sequence of β: $\{\beta_1, \beta_2, \ldots, \beta_T\}$
while $t <= T$ **do**
  if $t = 1$ **then**
    Call Algorithm 2 with inputted data set D, obtain the β_1
  else
    Construct new training data set in period t
    Add the test data set of the last period: $D = D \bigcup D_{t-1}$
    Drop the oldest data: $D = D \backslash D_{\text{oldest}}$
    Call Algorithm 2 with inputted data set D, obtain the β_t
  end if
  $\beta = \beta \bigcup \beta_t$
  $t = t + 1$
end while

Table 5
Summary of input indicators.

Attribute	Details	Attribute	Details	Attribute	Details
1	Return rates of NoDur[a]	9	Return rates of Shops	17	Adjusted close price
2	Return rates of Durbl	10	Return rates of Hlth	18	Volume
3	Return rates of Manuf	11	Return rates of Money	19	$\frac{(\text{adj.close}_t - \text{adj.close}_{t-1})}{\text{adj.close}_{t-1}}$
4	Return rates of Enrgy	12	Return rates of Other	20	$\frac{(\text{adj.close}_t - \text{adj.close}_{t-2})}{\text{adj.close}_{t-2}}$
5	Return rates of Chems	13	High price	21	$\frac{(\text{adj.close}_t - \text{adj.close}_{t-3})}{\text{adj.close}_{t-3}}$
6	Return rates of BusEq	14	Low price	22	EMA (time period=5)
7	Return rates of Telcm	15	Open price	23	OBV
8	Return rates of Utils	16	Close price	24	ATR (time period=5)

[a]The abbreviation of industry in this paper can be referred to the website of KennethR.French.

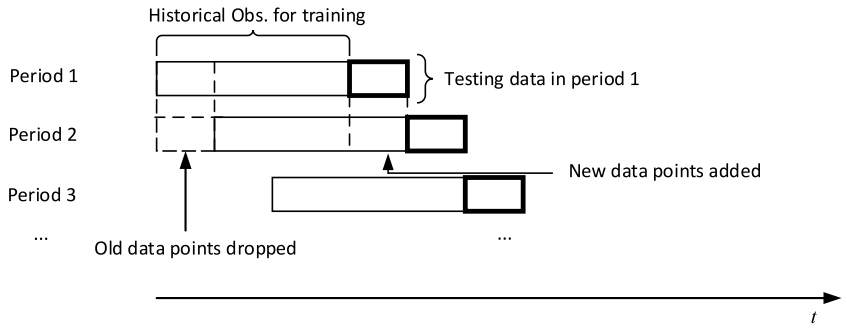


Fig. 5. Sliding window strategy for out-of-sample experiment.

5. Experimental analysis

We present both in-sample and out-of-sample experimental results in this section, acronyms of the proposed models and the benchmarks are shown in Table 6. Additionally, comprehensive portfolio evaluation metrics and associated detailed analysis are also provided. The involved software and toolboxes are MATLAB 2019b, MOSEK 9, YALMIP, CVX, and scikit-learn package. Fig. 6 outlines the overall flowchart of the numerical experiments.

5.1. Data description

The data set is the daily returns of 12 industries provided by the website of Kenneth R. French from Jan. 4th, 2010 to May 31st, 2019 (total 2369 observations). S&P 500 index of the same period fetched from Yahoo finance representing the market is selected as one of the benchmarks. To verify the effectiveness and robustness of the proposed models, we split our data set into two parts: the observations from Jan. 4th 2010 to Dec. 20th, 2013 (1000 observations) constitute the in-sample data set, where the static models are constructed and tested; the out-of-sample data set is composed of data points from Dec. 23th, 2013 to May 6th, 2019 (1350 observations), where the multi-period models are built and verified. The descriptive statistics of the data set is shown in Table 7.

5.2. Evaluation metrics

The primary purpose of the proposed portfolios is to maximize possible returns while minimizing potential risks. With this goal in mind, we calculate the following auxiliary metrics for comprehensively evaluating the obtained experimental results:

5.2.1. Return on investment (ROI)

ROI is a crucial performance indicator used to evaluate the efficiency and profitability of an investment. The standard formula of ROI is as follows:

$$\text{ROI} = \frac{\text{FinalWealth} - \text{InitialWealth}}{\text{InitialWealth}} \times 100\% \tag{43}$$

where FinalWealth corresponds to the cumulative return obtained from the investment bought with InitialWealth.

5.2.2. Annual percentage yield (APY)

Correspondingly, APY can be calculated as follows:

$$\text{APY} = \sqrt[n]{1 + \text{ROI}} - 1 \tag{44}$$

where n is the number of investment horizontal years.

5.2.3. Maximum drawdown (MDD)

MDD measures the capital break, which is one of the most significant criteria to evaluate a portfolio. The capital break is equal to the upper bound decline from the peak to a through before a new peak is attained, thus MDD can be expressed as follows:

$$\text{MDD} = \max_{t\in[0,T]} \{ \max_{i\in[0,t]} \text{ROI}_i - \text{ROI}_t \} \tag{45}$$

where $\max_{i\in[0,t]} \text{ROI}_i$ is the highest peak from the starting point to the instant t. Customarily, MDD is quoted as a percentage of the peak value, that is:

$$\text{MDD} = \max_{t\in[0,T]} \left\{ \frac{\max_{i\in[0,t]} \text{ROI}_i - \text{ROI}_t}{\max_{i\in[0,t]} \text{ROI}_i} \right\} \times 100\% \tag{46}$$

5.2.4. Sharpe ratio (SR)

SR is a risk-adjusted indicator that measures the excess return obtained while bearing a specific risk, which can be calculated as follows:

$$\text{SR} = \frac{\text{APY} - R_f}{\sigma_p} \tag{47}$$

where R_f is the risk-free rate, we set its value to 3% per year in this paper, and σ_p is the annual standard deviation (STD) of the excess return.

Table 6
Acronyms of proposed models and benchmarks.

Acronyms	Portfolio models	Ellipsoidal uncertainty	Algorithms for computing β
HRMV-u1-G	Hybrid robust mean–variance	U_δ^1	XGBoost
HRMV-u1-L	Hybrid robust mean–variance	U_δ^1	LSTM
HRMV-u2-G	Hybrid robust mean–variance	U_δ^2	XGBoost
HRMV-u2-L	Hybrid robust mean–variance	U_δ^2	LSTM
HRMVaR-u1-G	Hybrid robust mean-VaR	U_δ^1	XGBoost
HRMVaR-u1-L	Hybrid robust mean-VaR	U_δ^1	LSTM
HRMVaR-u2-G	Hybrid robust mean-VaR	U_δ^2	XGBoost
HRMVaR-u2-L	Hybrid robust mean-VaR	U_δ^2	LSTM
EQ	1/N [77]	✗	✗
MV	Mean–variance	✗	✗
SP500	S&P 500 index	✗	✗
MVaR-mip	Mean-VaR (MILP form)	✗	✗
MVaR-socp	Mean-VaR (SOCP form)	✗	✗

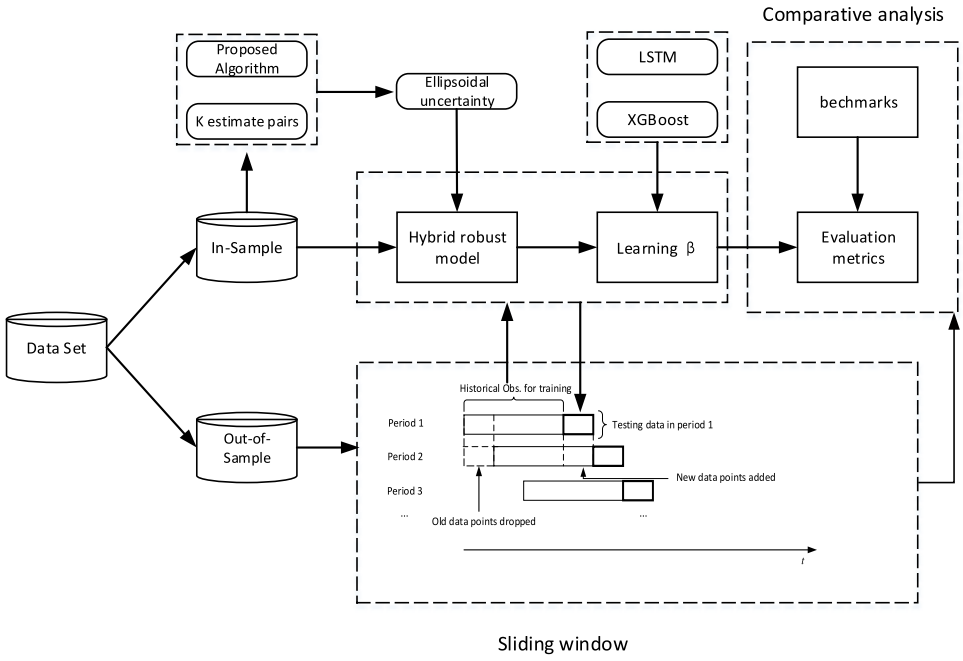


Fig. 6. Overall flowchart of experiments.

Table 7
Descriptive statistics of the data set.

Industry	Mean (%)	Stdev. (%)	25% Quantile (%)	50% Quantile (%)	75% Quantile (%)	Range
NoDur	0.0461	0.7863	−0.3650	0.0600	0.5100	0.0809
Durbl	0.0411	1.1430	−0.6750	0.0950	0.8200	0.1564
Manuf	0.0527	1.1655	−0.4900	0.0900	0.6500	0.1376
Energy	0.0171	1.3975	−0.6900	0.0200	0.7700	0.1510
Chems	0.0447	0.9222	−0.4400	0.0600	0.5600	0.1052
BusEq	0.0620	1.1243	−0.4400	0.1050	0.6500	0.1212
Telcm	0.0540	0.9413	−0.4250	0.0900	0.5900	0.1120
Utils	0.0436	0.8691	−0.4350	0.0700	0.5500	0.1060
Shops	0.0609	0.9204	−0.3700	0.0900	0.5800	0.1201
Hlth	0.0538	0.9867	−0.4200	0.0900	0.6100	0.1013
Money	0.0538	1.2350	−0.5150	0.0700	0.7100	0.1716
Other	0.0491	1.0815	−0.4600	0.0900	0.6300	0.1283
SP500	0.0426	0.9396	−0.3352	0.0556	0.5016	0.1162

5.2.5. Calmar ratio (CR)

CR also measures the balance between the risk and the return of a portfolio:

$$\mathrm{CR} = \frac{\mathrm{APY}}{\mathrm{MDD}} \tag{48}$$

5.3. In-sample performance

5.3.1. Cardinality constraints

Several studies pointed that holding too many assets is inefficient and hard to manage [3,4,30], that is, some cardinality constraints should be set from the perspective of theory and

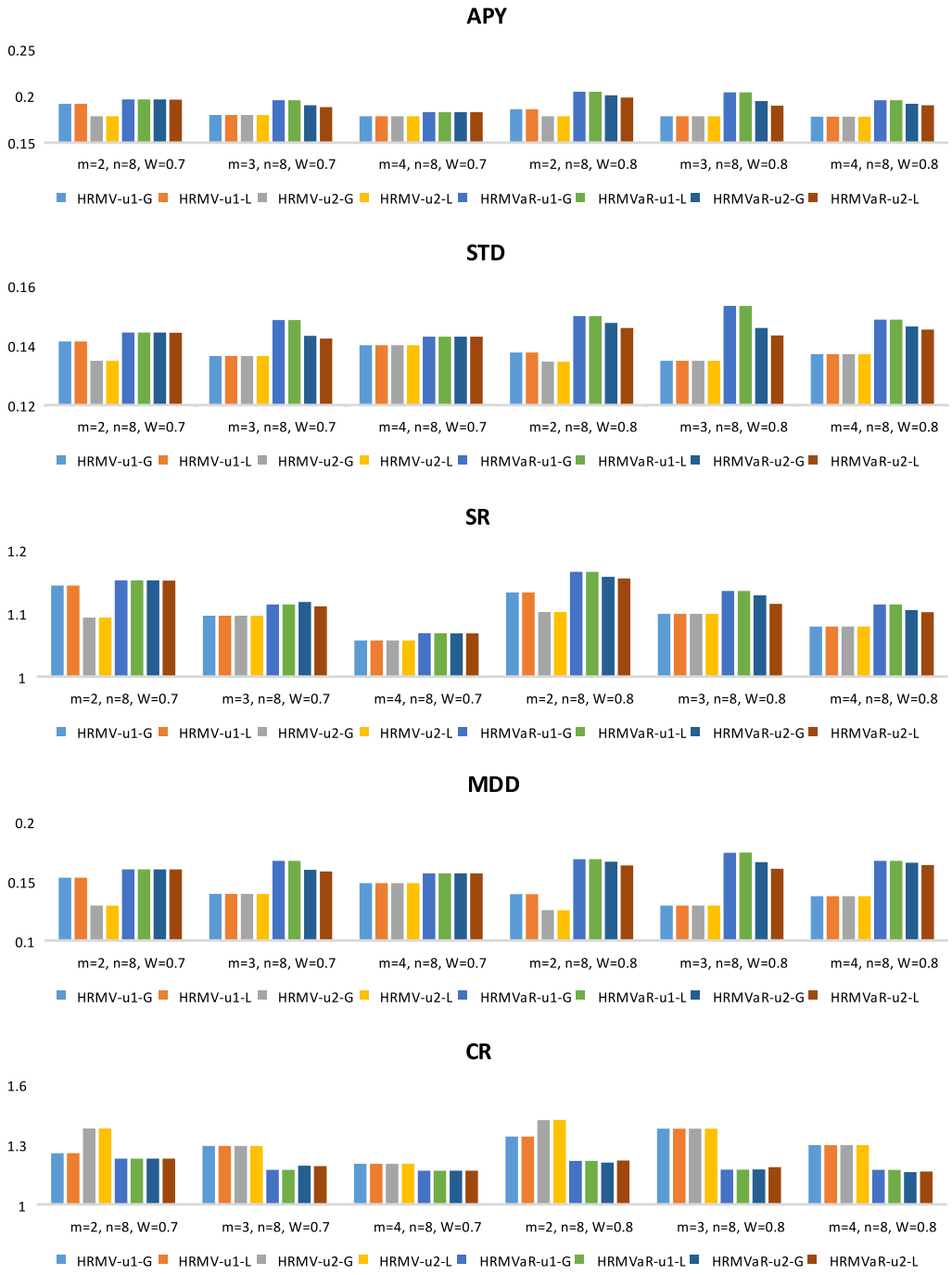


Fig. 7. Annualized performance of different cardinality constraints.

practice. Thus, the following constraints are taken into account in this work:

$$\begin{cases} m \le \|x\|_0 \le n \\ \sum_{[m]} x \le W \\ x \in \mathbb{X} \\ r^T x \ge \mu \end{cases} \tag{49}$$

where m is the minimal number of assets should be selected, n is the maximal number of assets could be selected, $\sum_{[m]} x \le W$ means the sum of the top m weights should not exceed W. To find the appropriate parameters of the additional cardinality constraints, we examine some different combinations of m, n, W to form the proposed portfolios. Fig. 7 illustrates the annualized performance of the tested models. For APY, the hybrid mean-VaR (HRMVaR) strategies obtain better performance than the hybrid mean–variance (HRMV) approaches. In terms of risk, HRMV series portfolios have lower STD than the HRMVaR series ones. With regard to the risk-adjusted indicators, when $m = 2, n = 8, W = 0.8$, all of the proposed models have relatively higher SR and CR than in other scenarios. Specifically, under this cardinality constraint, SR of the HRMVaR portfolios constrained with U_δ^1 uncertainty set is 1.1652, followed by 1.1021 for the HRMVaR portfolios constrained with U_δ^2 uncertainty set. CR of the HRMV-u2-L is 1.4204, compared to 1.4199 for the HRMV-u2-G, 1.3378 for the HRMV-u1-G and HRMV-u1-L, 1.2162 for HRMVaR-u1-G and HRMVaR-u1-L, 1.2081 for HRMVaR-u2-G, 1.2182 for

Table 8
Sources of the associated parameters.

Parameter	Source
δ_1	$\delta_1 = \sqrt{S * (\hat{r} * \hat{\Sigma}^{-1} * \hat{r}')}$, radius of ellipsoidal uncertainty set U_δ^1
δ_2	Algorithm 1, radius of ellipsoidal uncertainty set U_δ^2
β	Algorithm 2 or Algorithm 3, represents the trade-off between the worst-case counterpart and the best-case counterpart
λ	An empirical value ranging from 0 to 1, represents the trade-off between return and risk
μ	Mean value of all historical samples, lower bound of target return
M	An arbitrage large number

HRMVaR-u2-L. As for MDD, the HRMV-u2-G and HRMV-u2-L obtain the best overall performance in all parameter combinations. Additionally, LSTM and XGBoost give similar estimated values of β, which are 0.38 and 0.39, respectively. It is the reason for the similar performance between the hybrid robust models with β generating from LSTM and XGBoost. Other parameters in the tested models are as follows: $\lambda = 0.5, \delta_1 = 3.768, \delta_2 = 9, \mu = 6.9248 \times 10^{-4}, M = 1024$. Table 8 summarizes the sources of the parameters in the proposed models.

5.3.2. Comparison with benchmark models

Table 9 summarizes the in-sample performance of the proposed models and baseline strategies. Four panels in Table 9 present the return characteristics, risk characteristics, risk-adjusted indicators, and statistical tests for daily returns respectively:

(1) Panel A shows return characteristics including ROI, APY, Skewness, and Kurtosis. It can be observed that all of the hybrid robust models have higher ROI and APY than EQ, MVaR-mip, SP500. MVaR-socp obtains the highest ROI and APY: 1.1780 and 0.2148, HRMVaR-u1-G and HRMVaR-u1-L follow, with the ROI of 1.1059 and the APY of 0.2046.

(2) Panel B reveals risk characteristics, where STD, MDD, VaR(5%), and CVaR(5%) are provided. MV has the lowest STD, followed by HRMV-u2-G and HRMV-u2-L, but their values are almost identical: 0.1345 and 0.1346 respectively. With regard to MDD, MVaR-mip obtains the lowest MDD, 0.1237, the hybrid models, HRMV-u2-L and HRMV-u2-G follow, with 0.1255 and 0.1256 respectively. In terms of tail risk, HRMV-u2-G and HRMV-u2-L have the same lowest VaR(5%), 0.0131, and the same lowest CVaR(5%), 0.0202. Moreover, HRMVaR models exhibit better performance than MVaR-socp on all of the risk characteristics.

(3) Risk-adjusted indicators are shown in Panel C, where SR and CR are reported. HRMVaR-u1-G, HRMVaR-u1-L have the highest SR, 1.1652, compared to 1.1029 for MV, 1.0997 for MVaR-mip, 1.1189 for MVaR-socp. Our hybrid mean–variance models constrained with U_δ^2 ellipsoidal uncertainty have the SR of 1.1021, which is slightly higher than MVaR-mip. However, MVaR-mip has the highest CR, 1.4407, followed by HRMV-u2-L and HRMV-u2-G, which is 1.4204 and 1.4199 respectively.

(4) Panel D proves the effectiveness of the proposed models from a statistical point of view. Considering that the industry data is relatively stable (see Table 7), the daily risk-free rate is set as the benchmark for one-sided t-tests, where the null hypothesis is that the difference between the tested model and the benchmark rate is equal to zero, against the alternative that the difference is greater than zero. The statistical results further confirm the usefulness of the proposed models, as all of $p-$ values are less than the significance level of 5%.

Fig. 8 analyzes the portfolio weights, where the number of risky assets selected and the maximal weight of each portfolio are presented. Besides EQ, MV, HRMV-u2-G, and HRMV-u2-L select the largest number of risky assets. EQ has the lowest maximal weight, 0.0833, HRMVaR-u2-L ranks the second with 0.4478, HRMVaR-u2-G follows, with 0.5012. From the perspective of portfolio weight concentration, HRMVaR strategies tend to construct more diversified portfolios than HRMV, while MV seems to be more diversified than MVaR-mip and MVaR-socp, which demonstrates the effectiveness of the best-case counterpart in the proposed hybrid models.

Additionally, Figs. 9 and 10 visualize the results shown in Table 9. Fig. 9 shows the cumulative returns of the tested portfolios. Although MVaR-socp has the highest in-sample ROI and APY in the experiment, HRMVaR-u1-L gives the best performance from Q3-2011 to Q4-2012, covering approximately a third of the tested data points. From the distribution of daily returns presented in Fig. 10, the left tail of MVaR-socp is longer than the proposed hybrid models'. Such an characteristic is also reflected in Table 9, where the VaR(%5) and CVaR(%5) of MVaR-socp are 0.0161 and 0.0249, respectively, whereas the proposed hybrid models have lower VaR(%5) and CVaR(%5), that is, MVaR-socp assumes more tail risk than the hybrid models. Compared with the nominal MV, HRMV-u1 portfolios have higher in-sample ROI, APY, and SR, but this superiority vanishes on HRMV-u2 portfolios. Overall, the proposed models are beneficial to overcome the conservatism by bearing relatively more risk. However, further experiments should be carried out on the out-of-sample data set to obtain more convincing conclusions.

5.3.3. Sensitivity analysis

In the experiment above, the hyperparameter β is given by Algorithm 2, where the forecasting information is provided by LSTM and XGBoost. Essentially, β represents the optimism level about future market conditions, and we investigate the influence of different β for the proposed hybrid models. Considering the in-sample performance of the portfolios, the hybrid models constrained with ellipsoidal uncertainty U_δ^1 outperform the ones in U_δ^2. As a result, we present the results of risk analysis for HRMV-u1-G and HRMVaR-u1-G as follows.

Fig. 11 shows the in-sample APY, STD, SR, and CR of HRMV-u1-G with different β ranging from 0.1 to 0.5, and all of the indicators are relatively robust to βs in HRMV-u1-G. As shown in Fig. 12, HRMVaR-u1-G shows the better risk-adjusted performance on condition that $\beta < 0.25$ than other values, but the difference is not significant. Although the in-sample results of sensitivity analysis are stable, the performance of out-of-sample experiment deserves more attention. Additionally, We do not report the results of the other proposed hybrid models, as the conclusions are identical. Similar results are also shown in Tables 4 $\sim$ 6 of Lotfi et al. [19].

5.4. Out-of-sample performance

5.4.1. Generation of training and testing periods

As shown in Figs. 5 and 6, the sliding window method is applied in the out-of-sample experiment. Specifically, five overlapping training-testing periods are set, and each period contains a training stage of 995 days and a testing stage of 270 days. In each period, the hyperparameters and portfolio weights are obtained in the training stage, and the rest data fully out-of-sample in the testing stage is performed for testing based on the optimized portfolio weights. Then, the entire window will roll forward 270 days, with the most recent updated samples are added while the equal number of oldest observations are

Table 9
In-sample performance characteristics.

Model	EQ	MV	MVaR-mip	MVaR-socp	HRMV-u1-G	HRMV-u1-L	HRMV-u2-G	HRMV-u2-L	HRMVaR-u1-G	HRMVaR-u1-L	HRMVaR-u2-G	HRMVaR-u2-L	SP500
Panel A: Return characteristics													
ROI	0.8818	0.9276	0.9274	1.1780	0.9783	0.9783	0.9276	0.9276	1.1059	1.1059	1.0799	1.0633	0.6306
APY	0.1712	0.1783	0.1783	0.2148	0.1860	0.1860	0.1783	0.1783	0.2046	0.2046	0.2009	0.1985	0.1300
Skewness	−0.4003	−0.4765	−0.4674	−0.4887	−0.4997	−0.4997	−0.4732	−0.4731	−0.5023	−0.5023	−0.5071	−0.5099	−0.3687
Kurtosis	7.1336	6.7850	6.7892	6.8530	6.7456	6.7456	6.6972	6.6963	6.8894	6.8894	6.9472	6.9548	7.1028
Panel B: Risk characteristics													
STD	0.1738	0.1345	0.1348	0.1652	0.1377	0.1377	0.1346	0.1346	0.1499	0.1499	0.1476	0.1459	0.1702
MDD	0.2028	0.1272	0.1237	0.1994	0.1390	0.1390	0.1256	0.1255	0.1683	0.1683	0.1663	0.1630	0.1939
VaR(5%)	0.0168	0.0132	0.0129	0.0161	0.0137	0.0137	0.0131	0.0131	0.0147	0.0147	0.0147	0.0146	0.0168
CVaR(5%)	0.0264	0.0203	0.0203	0.0249	0.0208	0.0208	0.0202	0.0202	0.0226	0.0226	0.0223	0.0221	0.0260
Panel C: Risk-adjusted indicators													
SR	0.8125	1.1029	1.0997	1.1189	1.1328	1.1328	1.1021	1.1021	1.1652	1.1652	1.1575	1.1550	0.5879
CR	0.8443	1.4019	1.4407	1.0775	1.3378	1.3378	1.4199	1.4204	1.2162	1.2162	1.2081	1.2182	0.6707
Panel D: Statistical tests for daily returns													
t-statistic	1.6560	2.1406	2.1351	2.1696	2.1916	2.1916	2.1392	2.1393	2.2460	2.2460	2.2333	2.2291	1.2614
p-value	0.0490	0.0163	0.0165	0.0151	0.0143	0.0143	0.0163	0.0163	0.0125	0.0125	0.0129	0.0130	0.1037

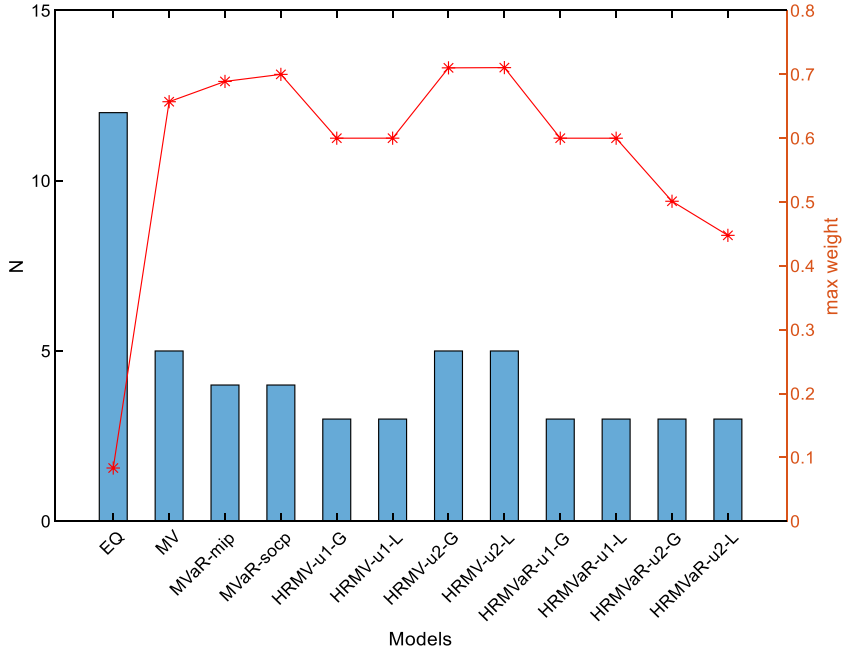


Fig. 8. In-sample performance: Portfolio weight.

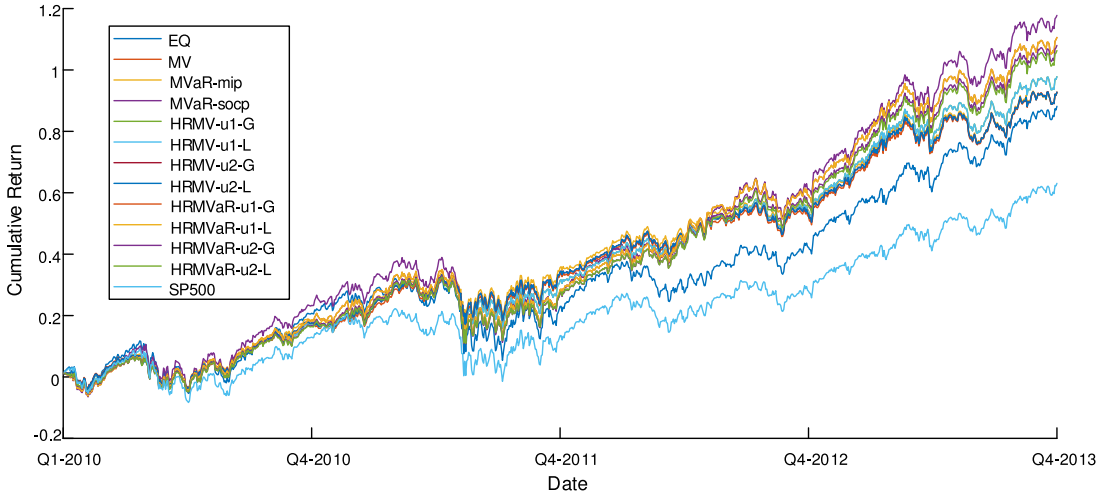


Fig. 9. In-sample performance: Cumulative returns.

dropped. Fig. 13 presents the hyperparameters, δ and β, in the five overlapping training-testing periods.

5.4.2. Comparison with benchmark models

In order to maintain the consistency of conditions, the cardinality constraints (49) still hold in the out-of-sample experiment.

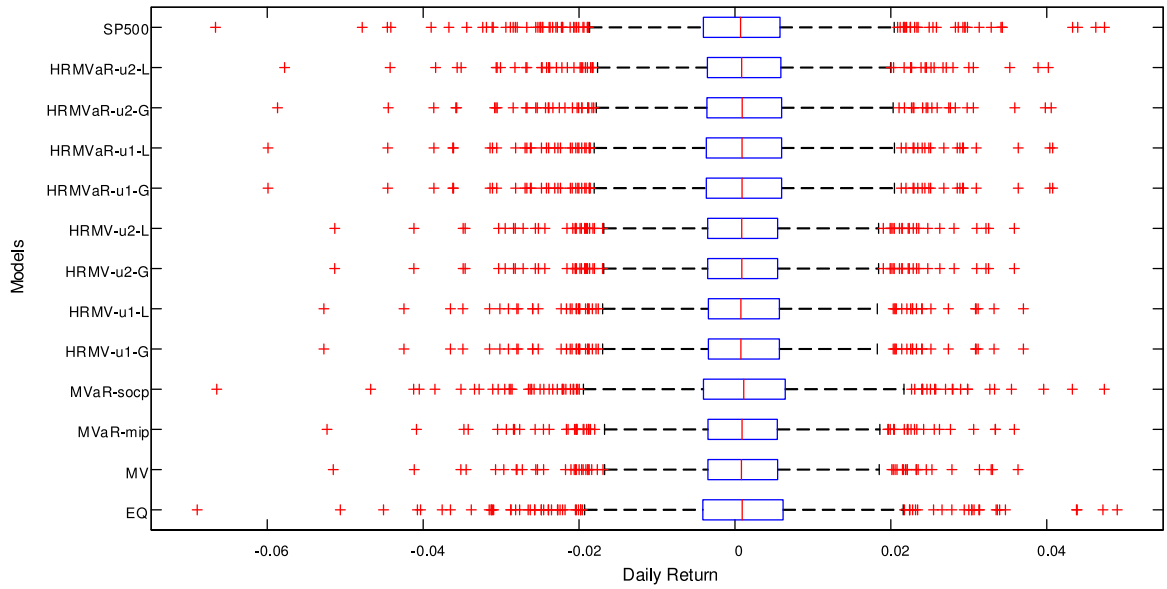
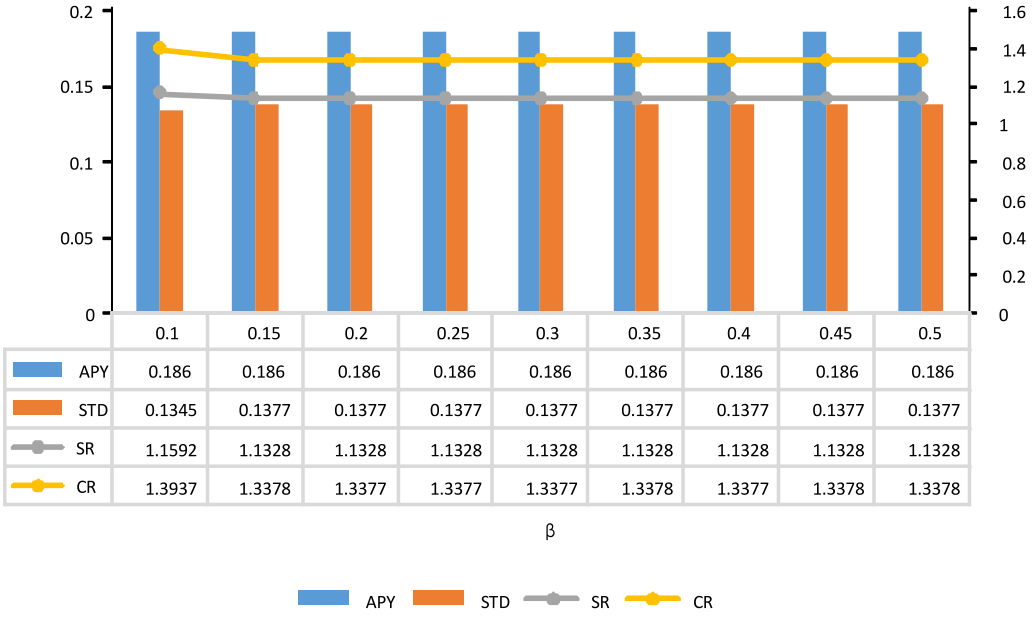


Fig. 10. In-sample performance: Daily returns.

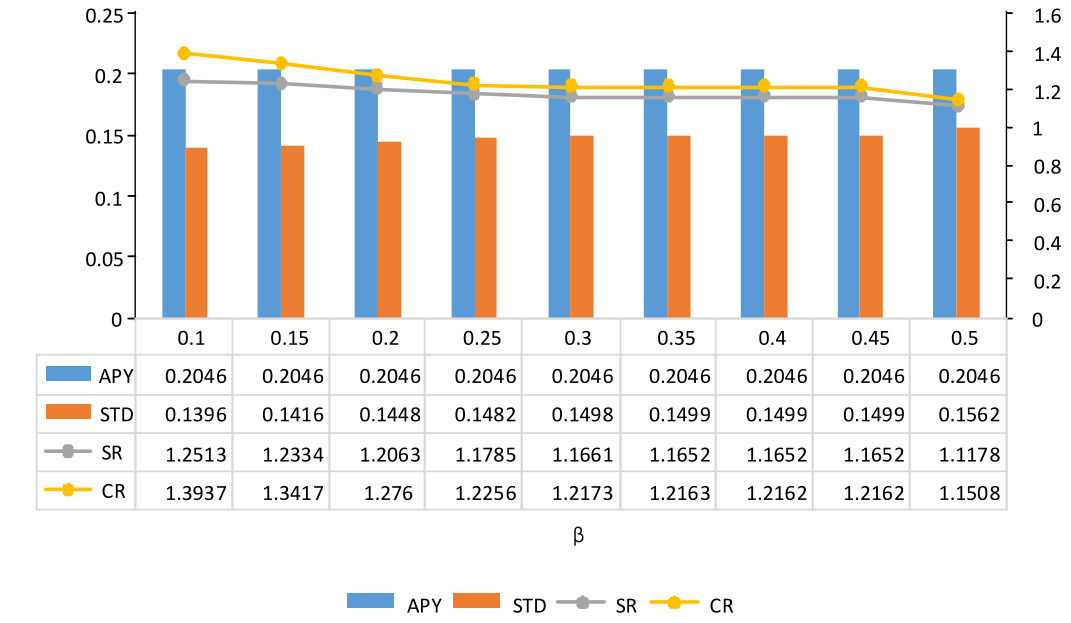


	0.1	0.15	0.2	0.25	0.3	0.35	0.4	0.45	0.5
APY	0.186	0.186	0.186	0.186	0.186	0.186	0.186	0.186	0.186
STD	0.1345	0.1377	0.1377	0.1377	0.1377	0.1377	0.1377	0.1377	0.1377
SR	1.1592	1.1328	1.1328	1.1328	1.1328	1.1328	1.1328	1.1328	1.1328
CR	1.3937	1.3378	1.3377	1.3377	1.3377	1.3378	1.3377	1.3378	1.3378

Fig. 11. In-sample performance: Different β for HRMV-u1-G.

Table 10
Out-of-sample performance characteristics.

Model	EQ	MV	MVaR-mip	MVaR-socp	HRMV-u1-G	HRMV-u1-L	HRMV-u2-G	HRMV-u2-L	HRMVaR-u1-G	HRMVaR-u1-L	HRMVaR-u2-G	HRMVaR-u2-L	SP500
Panel A: Return characteristics													
ROI	0.6015	0.6238	0.5926	0.5346	0.7818	0.7820	0.7054	0.7057	0.6630	0.6859	0.6679	0.6678	0.6127
APY	0.0918	0.0947	0.0907	0.0832	0.1138	0.1138	0.1047	0.1048	0.0995	0.1023	0.1001	0.1001	0.0933
Skewness	−0.4191	−0.3901	−0.3936	−0.4730	−0.3203	−0.3201	−0.3974	−0.3971	−0.4000	−0.3983	−0.3869	−0.3873	−0.4005
Kurtosis	6.2395	5.4623	5.2660	5.5371	6.4559	6.4557	5.6330	5.6319	6.6816	6.7182	6.7821	6.7854	6.8066
Panel B: Risk characteristics													
STD	0.1277	0.1126	0.1112	0.1307	0.1234	0.1234	0.1109	0.1109	0.1441	0.1438	0.1422	0.1422	0.1314
MDD	0.1885	0.1632	0.1622	0.1473	0.1819	0.1819	0.1475	0.1475	0.2242	0.2242	0.2226	0.2226	0.1978
VaR(5%)	0.0135	0.0113	0.0117	0.0135	0.0126	0.0126	0.0115	0.0115	0.0156	0.0153	0.0152	0.0152	0.0142
CVaR(5%)	0.0202	0.0172	0.0168	0.0203	0.0189	0.0189	0.0169	0.0169	0.0230	0.0229	0.0227	0.0227	0.0210
Panel C: Risk-adjusted indicators													
SR	0.4842	0.5744	0.5461	0.4070	0.6788	0.6791	0.6737	0.6741	0.4826	0.5030	0.4932	0.4932	0.4814
CR	0.4817	0.5799	0.5592	0.5646	0.6257	0.6258	0.7101	0.7103	0.4440	0.4565	0.4499	0.4499	0.4751
Panel D: Statistical tests for daily returns													
t-statistic	1.1978	1.3745	1.3132	1.0363	1.6027	1.6032	1.5823	1.5831	1.2104	1.2531	1.2310	1.2310	1.1955
p-value	0.1156	0.0848	0.0947	0.1501	0.0546	0.0546	0.0569	0.0568	0.1132	0.1052	0.1093	0.1093	0.1161



	0.1	0.15	0.2	0.25	0.3	0.35	0.4	0.45	0.5
APY	0.2046	0.2046	0.2046	0.2046	0.2046	0.2046	0.2046	0.2046	0.2046
STD	0.1396	0.1416	0.1448	0.1482	0.1498	0.1499	0.1499	0.1499	0.1562
SR	1.2513	1.2334	1.2063	1.1785	1.1661	1.1652	1.1652	1.1652	1.1178
CR	1.3937	1.3417	1.276	1.2256	1.2173	1.2163	1.2162	1.2162	1.1508

Fig. 12. In-sample performance: Different β for HRMVaR-u1-G.

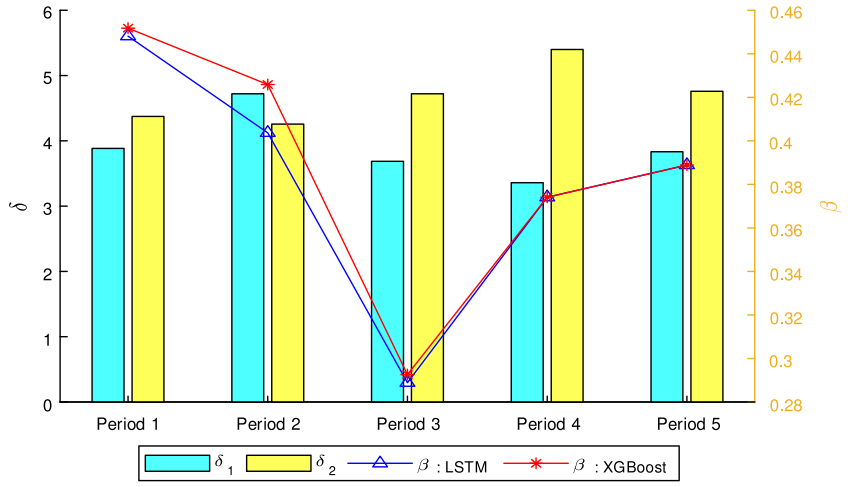


Fig. 13. Out-of-sample: Hyperparameters.

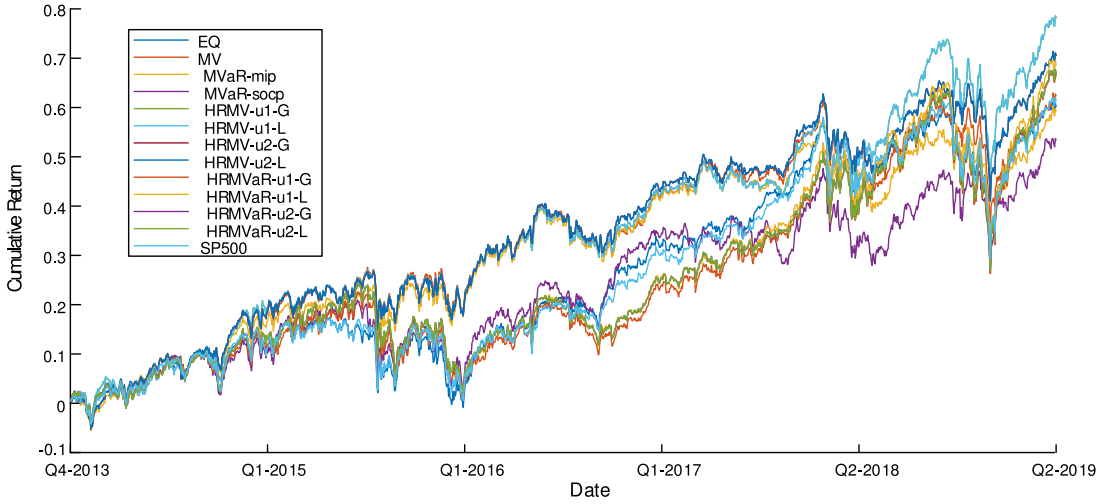


Fig. 14. Out-of-sample performance: Cumulative returns.

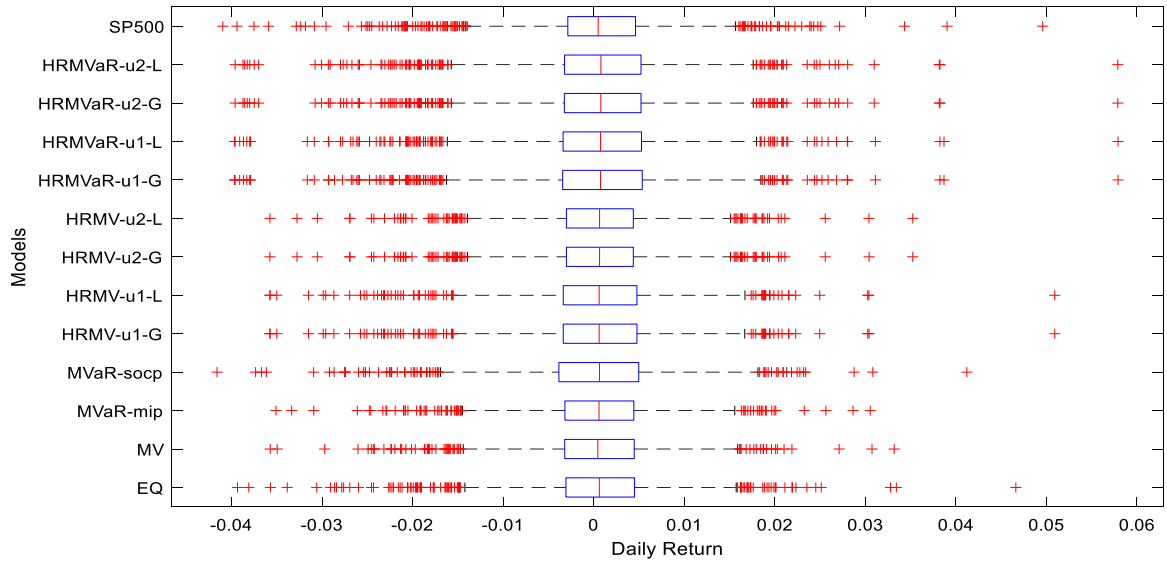


Fig. 15. Out-of-sample performance: Daily returns.

Table 10 illustrates the performance of the proposed models as well as the baseline strategies. Four panels included in Table 10 are as follows:

(1) Panel A. As for return characteristics, the hybrid robust models have better performance than benchmarks, HRMV-u1-L obtain the highest ROI and APY in the out-of-sample experiment, 0.7820 and 0.1138, respectively. In a comparison, MVaR-socp, which has the highest ROI and APY in the in-sample experiment, only obtains ROI of 0.5346 and APY of 0.0832, which are even lower than SP500's performance. Also, it is easy to observe that HRMVaR models gain higher ROI and APY than MVaR-mip and MVaR-socp, which demonstrate the profitability of the best-case counterpart. Return characteristics illustrate that HRMV models outperform HRMVaR strategy in the out-of-sample experiment, which is the opposite of the in-sample circumstance.

(2) Panel B. With regard to risk characteristics, HRMV-u2-G and HRMV-u2-L reach the lowest STD of 0.1109, compared to 0.1112 for MVaR-mip, 0.1126 for MV. As for MDD, MVaR-socp have the lowest MDD, 0.1473, HRMV-u2-G and HRMV-u2-L rank the second with 0.1475. However, HRMVaR models have larger out-of-sample STD, MDD, VaR(5%), CVaR(5%) than benchmarks. MV has the lowest VaR(5%), 0.0113, HRMV-u2-G and HRMV-u2-L have the second lowest VaR(5%), 0.0115. MVaR-mip shows the lowest CVaR(5%) of 0.0168, HRMV-u2-G and HRMV-u2-L achieve CVaR(5%) of 0.0169, which is approximate to MVaR-mip. Overall, HRMV-u2 models have the lowest out-of-sample STD, and the second lowest MDD, VaR(5%), CVaR(5%). According to the in-sample and out-of-sample performances, HRMV-u2 models show the best risk characteristics. Meanwhile, the effectiveness of ellipsoidal uncertainty U_δ^2 as well as Algorithm 1 is reflected in the out-of-sample experiment, where models constrained with U_δ^2 show more robust performance than the ones in U_δ^1.

(3) Panel C. In terms of risk-adjusted indicators, HRMV-u1-L reaches the highest SR as 0.6791, and HMRV-u1-G ranks the second with 0.6788. As for CR, HRMV-u2-L achieves the highest level as 0.7103, followed by HRMV-u2-G, 0.7101. HRMVaR models have higher SR than MVaR-socp, but lower than MVaR-mip. Compared to the SR and CR values from the in-sample experiment, all models have lower SR and CR in the out-of-sample results even though the rolling window scheme is taken. However, instead of the subtle differences shown in the in-sample results, HRMV beat HRMVaR and other models by a large margin concerning SR and CR.

(4) Panel D. All HRMV portfolio models obtain higher daily return rates than the daily risk-free rate at a significance level of 10%. Hence, the out-of-sample experimental results further verify the efficiency and effectiveness of the proposed hybrid models from statistics.

Additionally, Figs. 14 to 15 visualize the results of the out-of-sample experiment. From Fig. 14, HRMV-u2-L has a slight advantage over HRMV-u1-L most of the testing period, but HRMV-u1-L obtains evident higher cumulative return than other models since Q2-2018. The excellent out-of-sample performance of HRMV models can also observed in Fig. 15, where HRMV strategies have obvious shorter left tails than HRMVaR models. The skewness, kurtosis, and CVaR(%5) in Table 10 also indicate this conclusion, where HRMV-u1-L has the highest skewness of -0.3201, HRMV-u2 models have the almost lowest CVaR(%5) of 0.0169 (only higher than MVaR-mip, 0.0168). HRMVaR models have clearly high kurtosis, which corresponds to the fat tails in Fig. 15. When Comparing the in-sample and out-of-sample performances of several VaR-based portfolios, MVaR-mip has shorter left tails than MVaR-socp, and the out-of-sample data points exacerbate the gap, which means MILP is a more accurate technique for capturing tail risk than SOCP modeling. A reasonable explanation for this phenomenon is that the MILP method abandons the assumption of normality of distribution and constructs models based on sampling. Fortunately, HRMVaR portfolios narrow the gap to some extent (especially in-sample) with the help of forecasting information and the elaborately designed ellipsoidal uncertainty set. However, due to the unknown, random and uncertainty of the out-of-sample data points, HRMVaR still show significant tail risk characteristic. Intuitively, the proposed hybrid models constrained with U_δ^2 have better risk indicators than those in U_δ^1, which is consistent with the basic assumption.

Fig. 16 presents the proposed portfolio weights in different periods, it can be found that the proposed models prefer to select $3 \sim 5$ assets to invest, and the max weights of different periods are between $35\% \sim 75\%$. Accordingly, the proposed hybrid models constrained with U_δ^2 are more diversified than those consider U_δ^1, which is also an appealing property of the joint ellipsoidal uncertainty set.

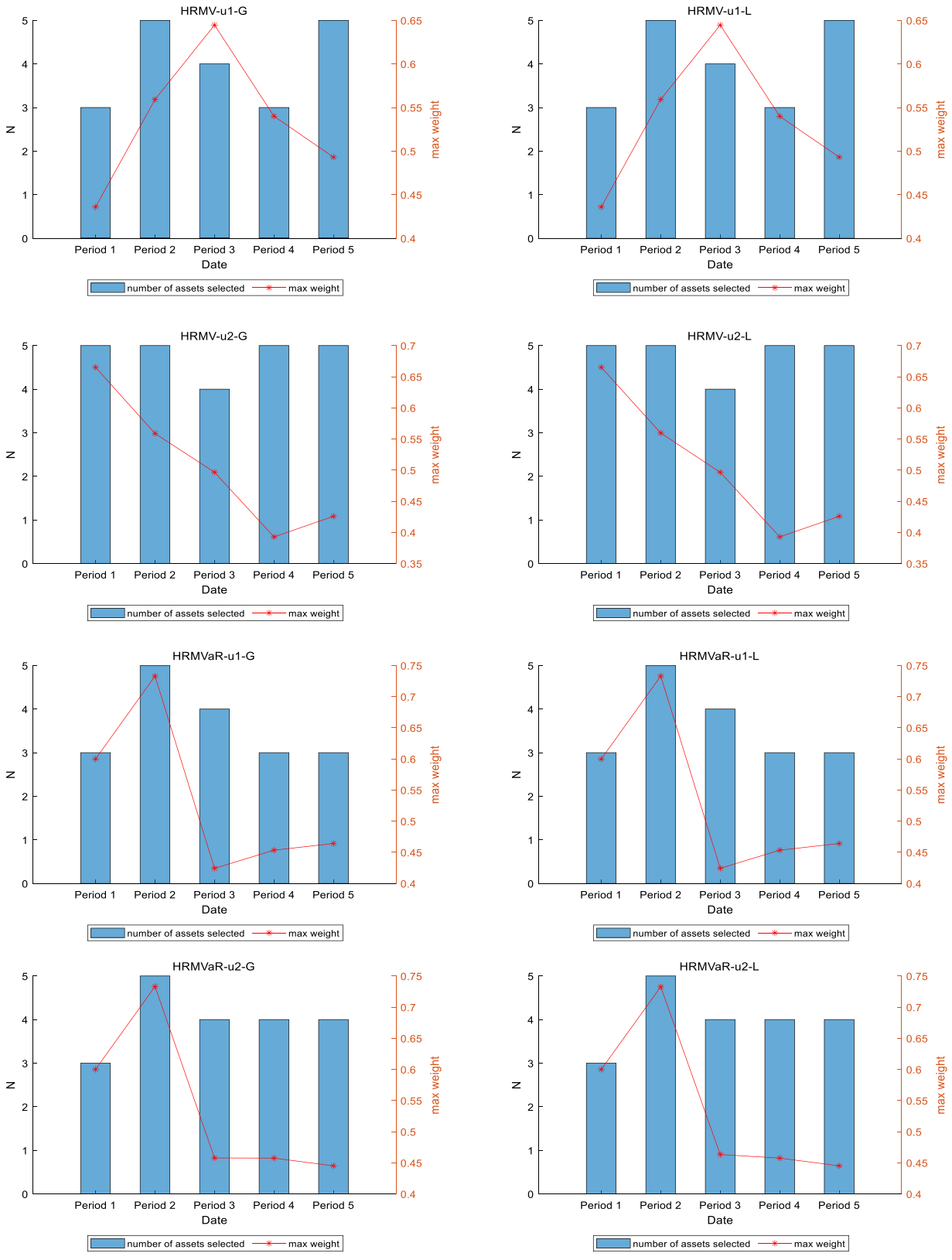


Fig. 16. Out-of-sample performance: Weights analysis.

5.4.3. Sensitivity analysis

Table 11 shows the results of out-of-sample sensitivity analysis, where the APY, STD, SR, CR of HRMV and HRMVaR under different combinations of β are presented. Apparently, STD is more sensitive to the change of β than APY, which is identical to the case shown in Figs. 11 and 12. HRMV models are less sensitive to out-of-sample perturbances than HRMVaR portfolios, and gain better overall performance. With regard to risk-related characteristics, models constrained with U_δ^2 exhibit more stable STD, SR, and CR trends than those in U_δ^1, which again verify the

Table 11
Out-of-sample performance: Sensitivity analysis.

Model	HRMV-u1	HRMV-u2	HRMVaR-u1	HRMVaR-u2
$\beta:(0-0-0-0-0)$				
APY	0.1138	0.1047	0.0995	0.1001
STD	0.1110	0.1106	0.1213	0.1187
SR	0.7548	0.6754	0.5734	0.5911
CR	0.7763	0.7176	0.5201	0.5541
$\beta:(0.1-0.1-0.1-0.1-0.1)$				
APY	0.1138	0.1047	0.0995	0.1001
STD	0.1125	0.1107	0.1246	0.1227
SR	0.7449	0.6750	0.5582	0.5718
CR	0.6985	0.7159	0.5079	0.5225
$\beta:(0.2-0.2-0.2-0.2-0.2)$				
APY	0.1138	0.1047	0.0995	0.1001
STD	0.1128	0.1108	0.1306	0.1273
SR	0.7426	0.6746	0.5323	0.5508
CR	0.7081	0.7141	0.4875	0.5023
$\beta:(0.3-0.3-0.3-0.3-0.3)$				
APY	0.1138	0.1047	0.0995	0.1001
STD	0.1150	0.1108	0.1384	0.1342
SR	0.7288	0.6742	0.5024	0.5228
CR	0.6011	0.7119	0.4709	0.4893
$\beta:(0.4-0.4-0.4-0.4-0.4)$				
APY	0.1138	0.1047	0.0995	0.1001
STD	0.1150	0.1109	0.1467	0.1460
SR	0.7284	0.6736	0.4741	0.4804
CR	0.6028	0.7098	0.4440	0.4509
$\beta:(0.5-0.5-0.5-0.5-0.5)$				
APY	0.1138	0.1047	0.0995	0.1001
STD	0.1156	0.1110	0.1473	0.1473
SR	0.7248	0.6730	0.4723	0.4763
CR	0.6311	0.7072	0.4440	0.4467

effectiveness of Algorithm 1. Overall, the improved robustness of the proposed hybrid portfolios generated by U_δ^2 is noticeable in out-of-sample experiment.

5.5. Comparison with existing researches

To the best of our knowledge, the related hybrid mean-VaR model was also mentioned in [19], and the worst-case mean–variance or mean-VaR portfolio is the special case of the hybrid model where $\beta = 0$. The proposed HRMV models have non-inferior returns from the numerical experiments above without sacrificing portfolio robustness to the existing models. Regarding computational complexity, the SOCP-based hybrid models are computationally tractable, while the MVaR-mip is based on MILP, which is NP-hard and resorts to the branch-and-bound or cutting-plane algorithm for solutions. Using Gurobi optimizer, the time consumption of solving a single hybrid model is about 0.17 s $\sim$ 0.30 s, whereas 9.59 s for solving an MVaR-mip model.

6. Conclusions & discussions

6.1. Discussions for key findings

This paper proposes the hybrid robust portfolio models based on multiple risk measures. Specifically, HRMV and HRMVaR portfolios are developed with detailed steps. Meanwhile, the equivalent relationship between robust VaR and robust CVaR under the specific assumptions is also proved according to the existing literature. Current studies show that the best-case counterpart is helpful to construct less conservative portfolios while the worst-case counterpart is beneficial to robust out-of-sample performance. Hence, both of the two counterparts deserve attention. In order to adjust the optimism level of the hybrid models according to real market movements, we introduce the trade-off parameter β between the best-case and the worst-case counterparts. Machine algorithms, LSTM and XGBoost, are employed to provide forecasting information about future market conditions, and a fusion approach is developed to synthesize the predictions and provide the reference points for β. We also suggest an clustering-based algorithm with polynomial time complexity to construct joint ellipsoidal uncertainty sets, where the conservatism of the robust portfolios could be reduced by the generated compact and feasible confidence ellipsoid. However, portfolio performance may not be the best on the parameters directly generated by the designed methods, but the suggested algorithms could provide some insightful start points for the follow-up parameters searching. Model structure robustness is the primary goal of robust portfolio, where some degree of parameter uncertainty should not severely influence the final solutions. Namely, robust portfolios should be insensitive to the sample perturbances as well as estimation errors. Accordingly, the out-of-sample performance and the sensitivity analysis results verify the robustness of the proposed hybrid models. Numerical results suggest that the best-case counterpart is beneficial to improve portfolio performance and could affect the risk level of the hybrid models. Machine learning algorithms shed light on building the bridge between possible market conditions and the portfolio optimism level. A fine-tuned machine learning model could provide pinpoint predictions for establishing the rational coefficient of the best-case counterpart, which is also one of the future directions to perfect the proposed models.

6.2. Main conclusions

The in-sample and out-of-sample numerical results support the following conclusions:

- HRMV portfolios have higher ROI and APY than nominal MV, and HRMVaR models have higher ROI and APY than MVaR-mip in both in-sample and out-of-sample experiments. Namely, considering the best-case counterpart in robust portfolio formation has some positive impacts on overcoming conservatism.
- The designed clustering-based algorithm can generate the joint ellipsoidal uncertainty set for robust modeling. Unlike the existing analytical and heuristic methods, we neither get the analytical center by solving a nonlinear SDP problem nor iterative search the ideal center. The suggested algorithm can achieve a relatively compact confidence ellipsoid without consuming too much computing resources as an unsupervised learning algorithm.
- The out-of-sample performance illustrates that the proposed hybrid portfolios constrained with U_δ^2 have better risk characteristics. Similar results can also be obtained from the sensitivity analysis. Overall, the hybrid portfolios with U_δ^1 tend to pursue higher returns by assuming more risk, while those constrained with U_δ^2 have to seek a more robust approach to reach the goal.
- Incorporating forecasting information into portfolio modeling is a promising way to construct efficient investment strategies, and machine learning algorithms expand the arsenal of intelligent portfolio construction for scholars and practitioners.

6.3. Limitations & future works

Although this study achieves some encouraging conclusions, there are still several limitations to this work. Potential future

directions could deepen the existing research from the following aspects. First, we only simulate the ideal trading conditions in experiments, but transaction costs should be considered as an indispensable component in the subsequent work. Second, box uncertainty, asymmetric ellipsoidal uncertainty, and mixture distribution uncertainty are demonstrated to be scenarios in robust optimization conducive to developing versatile and practical hybrid robust portfolios. Besides, more sophisticated artificial intelligence techniques would be integrated into portfolio formation in future works.

CRediT authorship contribution statement

Liangyu Min: Conceptualization, Methodology, Format analysis, Software, Validation, Data curation, Writing – original draft, Writing – review & editing, Funding acquisition. **Jiawei Dong:** Data curation, Format analysis, Validation, Writing – review & editing. **Jiangwei Liu:** Data curation, Methodology, Format analysis. **Xiaomin Gong:** Conceptualization, Methodology, Format analysis, Validation, Writing – review & editing.

Declaration of competing 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.

Acknowledgments

This research was supported by the National Natural Science Foundation of China (No. 71971132), and the Fundamental Funds for the Central Universities of China(Nos. 2020110552, CXJJ-2018-401). The authors gratefully acknowledge the associated professor from the School of Information Management and Engineering, Shanghai University of Finance & Economics, Jianjun Gao, for his professional guidance. The authors also gratefully acknowledge the Shanghai Engineering Research Center of Finance Intelligence (No. 19DZ2254600) for the help provided.

References

[1] H.M. Markowitz, Portfolio selection, J. Finance 7 (1) (1952) 77.
[2] L. Plachel, A unified model for regularized and robust portfolio optimization, J. Econom. Dynam. Control 109 (2019) 103779.
[3] W. Wang, W. Li, N. Zhang, K. Liu, Portfolio formation with preselection using deep learning from long-term financial data, Expert Syst. Appl. 143 (2020) 113042.
[4] W. Chen, H. Zhang, M.K. Mehlawat, L. Jia, Mean–variance portfolio optimization using machine learning-based stock price prediction, Appl. Soft Comput. 100 (2021) 106943.
[5] Y. Ma, R. Han, W. Wang, Portfolio optimization with return prediction using deep learning and machine learning, Expert Syst. Appl. 165 (2021) 113973.
[6] V.K. Chopra, W.T. Ziemba, The effect of errors in means, variances, and covariances on optimal portfolio choice, J. Portf. Manag. 19 (2) (1993) 6.
[7] S. Lotfi, S.A. Zenios, Robust VaR and CVaR optimization under joint ambiguity in distributions, means, and covariances, European J. Oper. Res. 269 (2) (2018) 556–576.
[8] R.T. Rockafellar, S. Uryasev, et al., Optimization of conditional value-at-risk, J. Risk 2 (2000) 21–42.
[9] R.T. Rockafellar, S. Uryasev, Conditional value-at-risk for general loss distributions, J. Bank. Financ. 26 (7) (2002) 1443–1471.
[10] P. Artzner, F. Delbaen, J.-M. Eber, D. Heath, Coherent measures of risk, Math. Finance 9 (3) (1999) 203–228.
[11] S. Kou, X. Peng, C.C. Heyde, External risk measures and Basel accords, Math. Oper. Res. 38 (3) (2013) 393–417.
[12] E. Roos, D. den Hertog, Reducing conservatism in robust optimization, INFORMS J. Comput. 32 (4) (2020) 1109–1127.
[13] A. Ben-Tal, A. Nemirovski, Robust convex optimization, Math. Oper. Res. 23 (4) (1998) 769–805.
[14] D. Goldfarb, G. Iyengar, Robust portfolio selection problems, Math. Oper. Res. 28 (1) (2003) 1–38.
[15] L.E. Ghaoui, M. Oks, F. Oustry, Worst-case value-at-risk and robust portfolio optimization: A conic programming approach, Oper. Res. 51 (4) (2003) 543–556.
[16] L. Zhu, T.F. Coleman, Y. Li, Min-max robust CVaR robust mean-variance portfolios, J. Risk 11 (3) (2009) 55.
[17] S. Zhu, M. Fukushima, Worst-case conditional value-at-risk with application to robust portfolio management, Oper. Res. 57 (5) (2009) 1155–1168.
[18] D. Bertsimas, M. Sim, The price of robustness, Oper. Res. 52 (1) (2004) 35–53.
[19] S. Lotfi, M. Salahi, F. Mehrdoust, Adjusted robust mean-value-at-risk model: less conservative robust portfolios, Opt. Eng. 18 (2) (2017) 467–497.
[20] C. Chen, D. Liu, L. Xian, L. Pan, L. Wang, M. Yang, L. Quan, Best-case scenario robust portfolio for energy stock market, Energy 213 (2020) 118664.
[21] X. Chen, M. Sim, P. Sun, A robust optimization perspective on stochastic programming, Oper. Res. 55 (6) (2007) 1058–1071.
[22] K. Natarajan, D. Pachamanova, M. Sim, Incorporating asymmetric distributional information in robust value-at-risk optimization, Manage. Sci. 54 (3) (2008) 573–585.
[23] S. Gu, B. Kelly, D. Xiu, Empirical asset pricing via machine learning, Rev. Financ. Stud. 33 (5) (2020) 2223–2273.
[24] A. Altan, S. Karasu, E. Zio, A new hybrid model for wind speed forecasting combining long short-term memory neural network, decomposition methods and grey wolf optimizer, Appl. Soft Comput. 100 (2021) 106996.
[25] A. Altan, S. Karasu, S. Bekiros, Digital currency forecasting with chaotic meta-heuristic bio-inspired signal processing techniques, Chaos Solitons Fractals 126 (2019) 325–336.
[26] S. Karasu, A. Altan, S. Bekiros, W. Ahmad, A new forecasting model with wrapper-based feature selection approach using multi-objective optimization technique for chaotic crude oil time series, Energy 212 (2020) 118750.
[27] A. Altan, A. Parlak, Adaptive control of a 3D printer using whale optimization algorithm for bio-printing of artificial tissues and organs, in: 2020 Innovations in Intelligent Systems and Applications Conference (ASYU), IEEE, 2020, pp. 1–5.
[28] A. Altan, Performance of metaheuristic optimization algorithms based on swarm intelligence in attitude and altitude control of unmanned aerial vehicle for path following, in: 2020 4th International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT), IEEE, 2020, pp. 1–6.
[29] S. Karasu, A. Altan, Recognition model for solar radiation time series based on random forest with feature selection approach, in: 2019 11th International Conference on Electrical and Electronics Engineering (ELECO), IEEE, 2019, pp. 8–11.
[30] F.D. Paiva, R.T.N. Cardoso, G.P. Hanaoka, W.M. Duarte, Decision-making for financial trading: A fusion approach of machine learning and portfolio selection, Expert Syst. Appl. 115 (2019) 635–655.
[31] A. Altan, S. Karasu, The effect of kernel values in support vector machine to forecasting performance of financial time series, J. Cogn. Syst. 4 (1) (2019) 17–21.
[32] B. Chen, J. Zhong, Y. Chen, A hybrid approach for portfolio selection with higher-order moments: Empirical evidence from Shanghai Stock Exchange, Expert Syst. Appl. 145 (2020) 113104.
[33] S.M. Sarmento, N. Horta, Enhancing a pairs trading strategy with the application of machine learning, Expert Syst. Appl. 158 (2020) 113490.
[34] T.A. Borges, R.F. Neves, Ensemble of machine learning algorithms for cryptocurrency investment with different data resampling methods, Appl. Soft Comput. 90 (2020) 106187.
[35] K. Schöttle, R. Werner, Robustness properties of mean-variance portfolios, Optimization 58 (6) (2009) 641–663.
[36] A. Ben-Tal, L. El Ghaoui, A. Nemirovski, Robust Optimization, Princeton University Press, 2009.
[37] A.E. Lim, J.G. Shanthikumar, G.-Y. Vahn, Robust portfolio choice with learning in the framework of regret: Single-period case, Manage. Sci. 58 (9) (2012) 1732–1746.
[38] S. Zymler, D. Kuhn, B. Rustem, Worst-case value at risk of nonlinear portfolios, Manage. Sci. 59 (1) (2013) 172–188.
[39] A. Beck, A. Ben-Tal, Duality in robust optimization: primal worst equals dual best, Oper. Res. Lett. 37 (1) (2009) 1–6.
[40] A.B. Paç, M.Ç. Pınar, Robust portfolio choice with CVaR and VaR under distribution and mean return ambiguity, Top 22 (3) (2014) 875–891.
[41] N. Gülpınar, D. Pachamanova, E. Çanakoğlu, A robust asset–liability management framework for investment products with guarantees, OR Spectrum 38 (4) (2016) 1007–1041.
[42] A. Ling, J. Sun, M. Wang, Robust multi-period portfolio selection based on downside risk with asymmetrically distributed uncertainty set, European J. Oper. Res. 285 (1) (2020) 81–95.
[43] S. Basak, A. Shapiro, Value-at-risk-based risk management: optimal policies and asset prices, Rev. Financ. Stud. 14 (2) (2001) 371–405.
[44] A.A. Gaivoronski, G. Pflug, Value-at-risk in portfolio optimization: properties and computational approach, J. Risk 7 (2) (2005) 1–31.

[45] S. Benati, R. Rizzi, A mixed integer linear programming formulation of the optimal mean/value-at-risk portfolio problem, European J. Oper. Res. 176 (1) (2007) 423–434.
[46] P. Glasserman, P. Heidelberger, P. Shahabuddin, Efficient Monte Carlo Methods for Value-at-Risk, IBM Thomas J. Watson Research Division, 2000.
[47] P. Abad, S. Benito, C. López, A comprehensive review of value at risk methodologies, Span. Rev. Financ. Econ. 12 (1) (2014) 15–32.
[48] L.J. Hong, Z. Hu, G. Liu, Monte Carlo methods for value-at-risk and conditional value-at-risk: a review, ACM Trans. Model. Comput. Simul. (TOMACS) 24 (4) (2014) 1–37.
[49] J. Puerto, M. Rodríguez-Madrena, A. Scozzari, Clustering and portfolio selection problems: A unified framework, Comput. Oper. Res. 117 (2020) 104891.
[50] M. Khedmati, P. Azin, An online portfolio selection algorithm using clustering approaches and considering transaction costs, Expert Syst. Appl. 159 (2020) 113546.
[51] E.F. Fama, The behavior of stock-market prices, J. Bus. 38 (1) (1965) 34–105.
[52] S. Hochreiter, J. Schmidhuber, Long short-term memory, Neural Comput. 9 (8) (1997) 1735–1780.
[53] Y. LeCun, Y. Bengio, G. Hinton, Deep learning, Nature 521 (7553) (2015) 436–444.
[54] T. Fischer, C. Krauss, Deep learning with long short-term memory networks for financial market predictions, European J. Oper. Res. 270 (2) (2018) 654–669.
[55] H.Y. Kim, C.H. Won, Forecasting the volatility of stock price index: A hybrid model integrating LSTM with multiple GARCH-type models, Expert Syst. Appl. 103 (2018) 25–37.
[56] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, R. Salakhutdinov, Dropout: a simple way to prevent neural networks from overfitting, J. Mach. Learn. Res. 15 (1) (2014) 1929–1958.
[57] A. Karpathy, The Unreasonable Effectiveness of Recurrent Neural Networks, Vol. 21, Andrej Karpathy Blog, 2015, p. 23.
[58] A. Graves, Generating sequences with recurrent neural networks, 2013, arXiv preprint arXiv:1308.0850.
[59] R.E. Schapire, The strength of weak learnability, Mach. Learn. 5 (2) (1990) 197–227.
[60] Y. Freund, R.E. Schapire, A decision-theoretic generalization of on-line learning and an application to boosting, J. Comput. System Sci. 55 (1) (1997) 119–139.
[61] T. Hastie, R. Tibshirani, J. Friedman, The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Springer Science & Business Media, 2009.
[62] C. Krauss, X.A. Do, N. Huck, Deep neural networks, gradient-boosted trees, random forests: Statistical arbitrage on the S&P 500, European J. Oper. Res. 259 (2) (2017) 689–702.
[63] T. Chen, C. Guestrin, Xgboost: A scalable tree boosting system, in: Proceedings of the 22nd Acm Sigkdd International Conference on Knowledge Discovery and Data Mining, 2016, pp. 785–794.
[64] S. Dey, Y. Kumar, S. Saha, S. Basak, Forecasting to Classification: Predicting the Direction of Stock Market Price using Xtreme Gradient Boosting, Working Paper, 2016.
[65] X. Lu, J. Sun, Z. Song, G. Li, Z. Wang, Y. Hu, Q. Wang, D. Zhang, Prediction and analysis of cold rolling mill vibration based on a data-driven method, Appl. Soft Comput. 96 (2020) 106706.
[66] J. Nobre, R.F. Neves, Combining principal component analysis, discrete wavelet transform and XGBoost to trade in the financial markets, Expert Syst. Appl. 125 (2019) 181–194.
[67] T.C. Fu, F.L. Chung, R. Luk, C.M. Ng, Stock time series pattern matching: Template-based vs. rule-based approaches, Eng. Appl. Artif. Intell. 20 (3) (2007) 347–364.
[68] T.L.C. A, F.Y.C. B, An intelligent pattern recognition model for supporting investment decisions in stock market, Inform. Sci. s 346–347 (2016) 261–274.
[69] T.L. Chong, W.K. Ng, Technical analysis and the London stock exchange: testing the MACD and RSI rules using the FT30, Appl. Econ. Lett. 15 (14) (2008) 1111–1114.
[70] H. Seng, A new approach of moving average method in time series analysis, in: New Media Studies (CoNMedia), 2013 Conference on, 2013.
[71] M. Nakano, A. Takahashi, S. Takahashi, Generalized exponential moving average (EMA) model with particle filtering and anomaly detection, Expert Syst. Appl. 73 (MAY) (2016) 187–200.
[72] S. Basak, S. Kar, S. Saha, L. Khaidem, S.R. Dey, Predicting the direction of stock market prices using tree-based classifiers, N. Am. J. Econ. Finance 47 (JAN.) (2019) 552–567.
[73] M. Thakur, D. Kumar, A hybrid financial trading support system using multi-category classifiers and random forest, Appl. Soft Comput. 67 (2018) 337–349.
[74] A. Gorgulho, N. Rui, N. Horta, Applying a GA kernel on optimizing technical analysis rules for stock picking and portfolio composition, Expert Syst. Appl. 38 (11) (2011) 14072–14085.
[75] D.F. Gerritsen, E. Bouri, E. Ramezanifar, D. Roubaud, The profitability of technical trading rules in the Bitcoin market, Finance Res. Lett. 34 (2019).
[76] X. Li, C. Luo, An intelligent stock trading decision support system based on rough cognitive reasoning, Expert Syst. Appl. 160 (2020) 113763.
[77] V. DeMiguel, L. Garlappi, R. Uppal, Optimal versus naive diversification: How inefficient is the 1/N portfolio strategy? Rev. Financ. Stud. 22 (5) (2009) 1915–1953.