Worst-case Mean-VaR Portfolio Optimization with Higher-Order Moments

Liangyu Min, *Member*, *IAENG*, Dewen Liu*, Xiaohong Huang, Jiawei Dong

***Abstract*—Conservatism is a notorious shortcoming of the worst-case robust portfolio selection model. Numerous studies have been done to tackle this issue from the perspective of theoretical and practical. Based on the existing literature, this paper aims to develop less conservative portfolio models. When the assumption of normality for returns is not valid, higher-order moments have been demonstrated effective in improving portfolio performance. Hence, the worst-case mean-VaR optimization portfolio involving the higher-order moments is developed in this work. Additionally, the machine learning-based preselection is also designed and implemented for selecting risky assets to further overcome the potential conservatism. In the numerical experiments, the US 12 industry portfolio data set from Kenneth R. French is used to test and compare the proposed portfolio models and baseline strategies. The out-of-sample results show that the proposed portfolios have better comprehensive performance than benchmarks.**

***Index Terms*—Portfolio selection, Machine learning, VaR, Higher-order moments**

I. Introduction

QUANTITATIVE finance has gained impressive ground since the mean-variance (MV) portfolio model proposed by Markowitz [1]. On the basis of the classical bi-criteria optimization model and the factor model [2], [3], robust portfolio selection model [4] provides a feasible theoretical framework overcoming some well-known shortcomings such as parameter sensitivity of the traditional MV model. The main idea of robust portfolio selection is to consider the worst-case scenario for the feasible domain of model parameters. Goldfarb & Iyengar [4] estimate the covariance matrix resort to Fama-French factors and transform the primal optimization problem into a second-order cone programming, which can be solved efficiently. However, although robust portfolio models show relatively stable out-of-sample performance, some scholars pointed out that the inherent conservatism of such worst-case optimization makes this type of model not preferred by investors pursuing high returns [5], [6].

Existing literature reveals several feasible solutions to build less conservative portfolio models while maintaining the robust model structure. For example, [7] considered the best-case counterpart and worst-case counterpart simultaneously and presented the corresponding modeling flow. Value-at-risk (VaR) is the focused risk metric in their work, which reflects the tail risk of an investment instead of the symmetric risk shown by portfolio variance. The terrible sub-prime crisis in 2008 reminds scholars and practitioners of the importance of tail risk. As one of robust risk measures, VaR has been written into the Basel II accords. Following the precedent framework illustrated by Lotfi et al, we construct the basic worst-case mean-VaR portfolio constrained with the ellipsoidal uncertainty U_δ, whereas some techniques are applied to overcome the conservatism of the primal model. Based on [7], [8], [9] developed the hybrid portfolio models constrained with different ellipsoidal uncertainties, where some machine learning algorithms are used to predict the trade-off parameter.

With the development of computer science, portfolio models with higher-order moments are gradually adopted by academia. [10] show better performance would be obtained when the skewness is included in the optimization process. The superiority of mean-variance-skewness model over mean-variance model has been presented by [11]. Kurtosis is also noteworthy moment in portfolio formation. [12] compare the out-of-sample performance of mean-variance-skewness-kurtosis, mean-variance-skewness, mean-variance, where the mean-variance-skewness-kurtosis model obtains the best performance. A polynomial goal programming model for portfolio optimization considering kurtosis is implemented by [13], where a diversified portfolio with satisfactory performance is presented. Inspired by the existing research, higher-order moments shed light on a feasible approach to construct less conservative portfolio models, and constitute the indispensable parts of the proposed portfolio models.

Forecasting information could also be utilized reasonably to improve the performance of portfolio models. Several artificial intelligence techniques including machine learning and deep learning provide efficient and effective approaches to generate convincing predictions, which have been frequently used in financial modeling. [14] proposed a neural-based mean-variance-skewness model involving risk preferences, forecasts, and trading strategies, where the trained radial basis function (RBF) neural network is employed to solve the optimization problem. A hybrid approach based on two-stage clustering, RBF, and genetic algorithm (GA) is developed to construct port-

Manuscript received Sep. 13, 2021; revised Dec. 14, 2021. This work is supported by the National Natural Science Foundation of China (No. 71971132), the National Social Science Found of China (No. 21CGL019), the Graduate Innovation Fund in SUFE (Nos. CXJJ-2018-401, CXJJ-2019-325, CXJJ-2019-342), and the Fundamental Research Funds for the Central Universities (No. 2020110552).

Liangyu Min is a lecturer in the Faculty of Business Information, Shanghai Business School, Shanghai 201400, China (e-mail: minux@163.sufe.edu.cn).

Dewen Liu is a lecturer in the School of Management, Nanjing University of Posts and Telecommunications, Nanjing Jiangsu 210003, China (Corresponding author, e-mail: dewen0731@126.com).

Xiaohong Huang is a PhD candidate in the School of Information Management and Engineering, Shanghai University of Finance and Economics, Shanghai 200433, China (e-mail: huangxiaohong@163.sufe.edu.cn).

Jiawei Dong is a PhD candidate in the School of Information Management and Engineering, Shanghai University of Finance and Economics, Shanghai 200433, China (e-mail: hightempler@126.com).

folio model with higher-order moments by [12], where the two-stage clustering is used to preselect risky assets and RBF is employed to provide forecasting information. In their work, empirical evidence from Shanghai Stock Exchange is presented to verify the feasibility of the mean-variance-skewness-kurtosis portfolio model. Machine learning-based preselection is also used to improve the performance of portfolio models in [15], [16], [17]. As a result, we also design the process of preselection based on two ensemble learning algorithms, Random forest and LightGBM, to further overcome the conservatism of the worst-case mean-VaR model.

The main contribution of this paper is two-fold. Firstly, we present the worst-case mean-VaR optimization constrained with the ellipsoidal uncertainty U_δ with detailed proofs. Then, higher-order moments including skewness and kurtosis are integrated into the mean-VaR portfolio model in a manner of polynomial goal programming. Secondly, the preselection based on Random Forest and LightGBM is developed for selecting risky assets. Accordingly, we compare the proposed portfolio models with preselection and without preselection in the numerical experiments. Also, some baseline models such as MV and mean-VaR are evaluated in the empirical research to reveal the comparative result.

The rest of this paper is organized as follows. The ellipsoidal uncertainty set U_δ is introduced in Section II. Based on U_δ, the worst-case mean-VaR portfolio model is presented in Section III. In Section IV, skewness and kurtosis are considered, and the polynomial goal programming for portfolio formation is proposed. The machine learning-based preselection and the corresponding algorithms are presented and explained in Section V. Section VI presents the comparative result of the numerical experiments. Conclusions and related discussions are shown in Section VII.

II. Ellipsoidal uncertainty set

The ellipsoidal uncertainty set U_δ is introduced in this section. Suppose that risky assets returns follow a joint normal distribution, and the mean estimate r can be obtained based on a group of i.i.d samples of size S for n assets. Then, given covariance matrix Σ, the following relationship holds:

$$\frac{S(S-n)}{(S-1)n}(r-\bar{r})'\Sigma^{-1}(r-\bar{r}) \sim \chi^2_n \tag{1}$$

where χ^2_n is the chi-square distribution with degree of freedom n.

In this paper, we focus on the joint uncertainty set for the pair (r, Σ) as [7], [8]. Assuming that the two sample estimators, $\hat{r} = \frac{1}{S}\sum_{i=1}^{S} r_i, \hat{\Sigma} = \frac{1}{S}\sum_{i=1}^{S}(r_i - \hat{r})(r_i - \hat{r})'$, are independent, and the distributions are as follows:

$$\begin{cases} \hat{r} \sim \mathcal{N}(r, \dfrac{\Sigma}{S}) \\ \hat{\Sigma} \sim \mathcal{W}(\dfrac{\Sigma}{S-1}, S-1) \end{cases} \tag{2}$$

where $\mathcal{N}(\mu, \sigma^2)$ represents the Gaussian distribution, $\mathcal{W}(G, \nu)$ represents the Wishart distribution [18] with scale matrix G and degree of freedom ν. According to the procedure of [19], the joint ellipsoidal uncertainty set can be derived as follows:

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

where $\mathbb{S}^n$ is the set of the $n \times n$ symmetric matrices, $\|A\|_F^2 = tr(AA')$.

III. Worst-case mean-VaR optimization

Constrained with the ellipsoidal uncertainty set U_δ mentioned above, the worst-case counterpart of the mean-VaR portfolio optimization can be reformulated as follows:

$$\begin{aligned} &\min_{x\in\mathbb{X}} \max_{(r,\Sigma)\in U_\delta} -r'x + (1-\lambda)F^{-1}(\alpha)\|\Sigma^{1/2}x\| \\ =&\min_{x\in\mathbb{X}} -\hat{r}'x + \left(\max_{k\in[0,1]} g_1(k)\right)\|\hat{\Sigma}^{1/2}x\| \end{aligned} \tag{3}$$

where $\mathbb{X} = \{x \mid x_i \ge 0, \sum_{i=1}^{n} x_i = 1\}$ is the feasible domain of portfolio weight vector x, $g_1(k) = \delta\sqrt{\frac{k}{S}} + (1-\lambda)F^{-1}(\alpha)\sqrt{\left(1+\delta\sqrt{\frac{2(1-k)}{S-1}}\right)}$, $F^{-1}(\alpha)$ is the normal inverse cumulative distribution function, $\lambda \in [0,1]$ is the trade-off parameter between risk and return. The proof based on [19] and [7] is as follows. We first define the two uncertainty set with an auxiliary parameter $k \in [0,1]$:

$$\begin{aligned} &\mathcal{U}_{\sqrt{k}\delta} = \{r \in \mathbb{R}^n \mid S(r-\hat{r})\hat{\Sigma}(r-\hat{r}) \le k\delta^2\} \\ &\mathcal{U}_{\sqrt{1-k}\delta} = \{\Sigma \in \mathbb{S}^n \mid \frac{S-1}{2}\|\hat{\Sigma}^{-1/2}(\Sigma-\hat{\Sigma})\hat{\Sigma}^{-1/2}\|_F^2 \\ &\quad \le (1-k)\delta^2\} \end{aligned}$$

The inner maximization of (3) is equivalent to:

$$\max_{k\in[0,1]} \max_{\Sigma\in\mathcal{U}_{\sqrt{1-k}\delta}} \max_{r\in\mathcal{U}_{\sqrt{k}\delta}} -r'x + (1-\lambda)F^{-1}(\alpha)\|\Sigma^{1/2}x\| \tag{4}$$

and the optimal objective value of the innermost of (4) can be reached at the endpoint:

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

Plugging r^* into (4), the next innermost maximization in (4) is as follows:

$$\max_{U_{\sqrt{1-k}\delta}} -\hat{r}'x + (1-\lambda)F^{-1}(\alpha)\sqrt{x'\Sigma x} + \frac{\delta\sqrt{k}}{\sqrt{S}}\sqrt{x'\hat{\Sigma}x} \tag{5}$$

In essence, the problem (5) is an optimization about $\hat{\Sigma}$, and we can obtain the $\hat{\Sigma}^*$ by solving the following simplified optimization problem:

$$\begin{aligned} \max_{\tilde{\Sigma}\in\mathbb{S}^n} \quad & u'\tilde{\Sigma}u \\ s.t. \quad & \|\tilde{\Sigma}\|_F^2 \le b \end{aligned} \tag{6}$$

where $u = \hat{\Sigma}^{1/2}x, b = \frac{2}{S-1}(1-k)\delta^2, \tilde{\Sigma} = \hat{\Sigma}^{-1/2}(\Sigma - \hat{\Sigma})\hat{\Sigma}^{-1/2}$. According to the cyclic property of trace and the Cauchy-Schwarz inequality, we have the following equation:

$$\begin{aligned} u'\tilde{\Sigma}u &= tr(u'\tilde{\Sigma}u) = tr(u'u\tilde{\Sigma}) = < u'u, \tilde{\Sigma} > \\ &\le \|u'u\|_F\|\tilde{\Sigma}\|_F \\ &\le \sqrt{b}\|u'u\| \\ &= \sqrt{b}\|u\|^2 \end{aligned} \tag{7}$$

As a result, we have the optimal Σ^* at the boundary, that is, $\Sigma^* = \hat{\Sigma} + \sqrt{b}\hat{\Sigma}^{1/2}\frac{y}{\|y\|}\frac{y'}{\|y\|}\Sigma^{1/2}$. Also, the optimization problem (3) can be reformulated as follows:

$$\begin{aligned}
&\min_{x\in\mathbb{X}}\max_{k\in[0,1]} -\hat{r}'x + (1-\lambda)F^{-1}(\alpha)\sqrt{(1+\sqrt{b})x'\hat{\Sigma}x +}\\
&\frac{\delta\sqrt{k}}{\sqrt{S}}\sqrt{x'\hat{\Sigma}x}\\
=&\min_{x\in\mathbb{X}}\max_{k\in[0,1]} -\hat{r}'x +\\
&\underbrace{\left((1-\lambda)F^{-1}(\alpha)\sqrt{1+\sqrt{b}} + \frac{\delta\sqrt{k}}{\sqrt{S}}\right)}_{g_1(k)}\|\hat{\Sigma}^{1/2}x\|\\
=&\min_{x\in\mathbb{X}} -\hat{r}'x + \left(\max_{k\in[0,1]} g_1(k)\right)\|\hat{\Sigma}^{1/2}x\|
\end{aligned} \tag{8}$$

IV. Portfolio considering higher-order moments

A. Skewness & kurtosis

When the normality assumption of risky assets returns is not valid, the efficiency of the Markowitz portfolio model can still be sustained by adding some higher-order moments. Existing literature shows that more and more scholars consider portfolio models with higher-order moments [20], [21], [22], [13], [23], [24], [25]. Mean-variance-skewness is one of the mainly focused models, and skewness has been demonstrated to improve the performance of a portfolio [10]. The effectiveness of kurtosis in portfolio formation also be discussed by some researchers [26], [27], [13], [28]. A portfolio with excessive kurtosis indicates that the fatter tails than a normal distribution the portfolio would have, that is, a higher probability of extreme returns would be expected. To overcome the notorious conservatism in robust optimization, and obtain a stable financial model with satisfactory performance, we design the portfolio model as a non-linear, multi-functional approach to the issue of the optimization comprised of multiple conflicting objectives.

The skewness and kurtosis of a portfolio can be calculated as follows:

$$\begin{aligned}
&Skew(x) = \mathbb{E}(x'(\tilde{r}-\mathbb{E}(\tilde{r}))^3) = x'M_3(x\otimes x)\\
&Kurt(x) = \mathbb{E}(x'(\tilde{r}-\mathbb{E}(\tilde{r}))^4) = x'M_4(x\otimes x\otimes x)
\end{aligned}$$

where $\tilde{r}$ represents the returns of risky assets, M_3 is the co-skewness matrix, M_4 is the co-kurtosis matrix, and $\otimes$ denotes the kronecker product. Accordingly, the designed multi-objective portfolio optimization is as follows:

$$\begin{cases}
\min -\hat{r}'x + \left(\max_{k\in[0,1]} g_1(k)\right)\|\hat{\Sigma}^{1/2}x\|\\
\max x'M_3(x\otimes x)\\
\min x'M_4(x\otimes x\otimes x)\\
x\in\mathbb{X}
\end{cases} \tag{9}$$

B. Polynomial goal programming algorithm

Polynomial goal programming (PGP) is one of frequent-used approaches to solve multi-objective models [29], [28], [13], [25]. Hence, based on the current research results, we develop the PGP algorithm to solve Eq. (9). In the PGP process, we first solve each sub-problem to achieve the aspired levels, R^*, V^*, K^*. Each aspired level means the ideal scenario for a single objective while neglecting other goals. Then, we introduce the auxiliary variables d_r, d_v, d_k to minimize the deviations from the aspired levels, respectively. The aspired levels can be obtained by solving the following sub-problems SP(1) $\sim$ SP(3) independently:

$$\text{SP}(1)\begin{cases}
\min \quad R^* = -\hat{r}'x + \left(\max_{k\in[0,1]} g_1(k)\right)\|\hat{\Sigma}^{1/2}x\|\\
s.t. \quad x\in\mathbb{X}
\end{cases} \tag{10}$$

$$\text{SP}(2)\begin{cases}
\max \quad V^* = x'M_3(x\otimes x)\\
s.t. \quad x\in\mathbb{X}
\end{cases} \tag{11}$$

$$\text{SP}(3)\begin{cases}
\min \quad K^* = x'M_4(x\otimes x\otimes x)\\
s.t. \quad x\in\mathbb{X}
\end{cases} \tag{12}$$

These sub-problems can be solved by some common linear and non-linear programming methods, and the objective values can be integrated into the PGP framework via Minkowski distance [27] defined as follows:

$$Z = \left(\sum_{i=1}^{m}\left|\frac{d_i}{Z_i}\right|^p\right)^{1/p} \tag{13}$$

where Z_i is used to normalized the ith goal. Likewise, the goal variables are normalized by the corresponding aspired levels obtained from SP(1) $\sim$SP(3). Also, investors' preferences are prioritized by hyper-parameter λ_i. Following the procedure of [30], [13], we add 1 to all normalized goals to ensure that they are larger than 1, and therefore the normalized deviation from the desired levels would increase strictly with the exponent. As a result, all objectives are considered simultaneously by solving the following PGP problem:

$$\begin{aligned}
&\min Z = \left(1+\left|\frac{d_r}{R^*}\right|\right)^{\lambda_1} + \left(1+\left|\frac{d_v}{V^*}\right|\right)^{\lambda_2} + \left(1+\left|\frac{d_k}{K^*}\right|\right)^{\lambda_3}\\
&s.t.\begin{cases}
-\hat{r}'x + \left(\max_{k\in[0,1]} g_1(k)\right)\|\hat{\Sigma}^{1/2}x\| - d_r = R^*\\
x'M_3(x\otimes x) + d_v = V^*\\
x'M_4(x\otimes x\otimes x) - d_k = K^*\\
x\in\mathbb{X}\\
d_r, d_v, d_k \geq 0
\end{cases}
\end{aligned} \tag{14}$$

where the value of λ_i is dependent on the preference of the investor. Accordingly, we can obtain the solution of Eq. (14) for the different scenarios. Algorithm 1 illustrates the flowchart of the proposed PGP approach.

V. Preselection using machine learning-based algorithms

The effectiveness and efficiency of machine learning algorithms in financial modeling has been demonstrated by lots of scholars [15], [16], [31], [32]. Also, the performance of portfolio models could be significantly improved by the rational preselection based on machine learning algorithms. To pick out high-quality risky assets from the

Algorithm 1 PGP approach.

Input: Sample data D.

Output: Optimized portfolio weight x_{opt}.

1: Estimate the sample mean vector $\hat{r}$, co-variance matrix $\hat{\Sigma}$, co-skewness matrix M_3, and co-kurtosis matrix M_4 from D.

2: Solve the sub optimization problems SP(1) $\sim$SP(3) independently, and obtain the respective desired objective value, R^*, V^*, K^*.

3: Introduce the auxiliary variables d_r, d_v, d_k, and construct the PGP problem (14) based on the Minkowski distance (13).

4: Solve the PGP problem (14), and obtain the optimized portfolio weight x_{opt}.

5: **return** x_{opt}.

asset pool, we implement the preselection synthesizing the forecasting results retrieved from machine learning algorithms. In this section, two ensemble learning models within the scope of this work are introduced. Then, the preselection based on the chosen machine learning algorithms is proposed and explained.

A. Random Forest

The main idea of ensemble learning models is combining several base estimators into a strong one to improve forecasting accuracy [33], [34], [35]. Bagging and boosting are two widely-used strategies in modeling. Random Forest (RF) is one of the ensemble learning models taking the bagging strategy, where the predictions gained from the individual base estimators independently are averaged in RF. Suppose that there are M features in the data set, and at most $k \approx \sqrt{M}$ features would be randomly selected by the base estimator for making decision. Generally, information gain [36], [37] is the criterion when splitting a node. The final result $F(x)$ of RF would be achieved according to the voting equation as follows:

$$F(x) = \arg\max \sum_{i=1}^{n} \mathbf{I}(r_i(x) = R(x)) \tag{15}$$

where $r_i(x)$ is the forecasting result of the ith base estimator, $\mathbf{I}(\cdot)$ is the indicator function. More details about RF can be referred to [38].

B. LightGBM

LightGBM is a boosting ensemble learning model proposed by [39]. Essentially, LightGBM is a kind of gradient boosting decision tree (GBDT) algorithm [40], [41], but two novel techniques, gradient-based one-side sampling (GOSS) and exclusive feature bundling (EFB), are developed for this improved GBDT algorithm. Additionally, unlike the traditional GBDT-based algorithms such as XGBoost [42], [43], the trees in LightGBM would grow vertically, whereas horizontally in other algorithms [44], making LightGBM an effective and efficient approach in processing large-scale data sets.

Given the supervised training sample set $X = \{(x_i, y_i)\}_{i=1}^{n}$, the goal of LightGBM is to find an approximation $\hat{f}(x)$ to a certain optimal function that minimizes the expected value of the loss function $L(y, f(x))$ as follows:

$$\hat{f} = \arg\min_{f} \mathbb{E}[L(y, f(x))] \tag{16}$$

where the boosting strategy is employed in LightGBM, that is, a number of T regression trees $f_i(X)$ are built to approximate the final model in an additive form as follows:

$$f_T(X) = \sum_{i=1}^{T} f_i(X) \tag{17}$$

Accordingly, at step t, the objective function of LightGBM can be expressed as follows:

$$\Gamma_t = \sum_{i=1}^{n} L(y_i, f_{t-1}(x_i) + f_i(x_i)) \tag{18}$$

and the loss function can be rapidly approximated by Taylor expansions, hence $Gamma_t$ can be reformulated as follows:

$$\Gamma_t \approx \sum_{i=1}^{n} \left(g_i f_t(x_i) + \frac{1}{2} h_i f_t^2(x_i) \right) \tag{19}$$

where g_i and h_i represent the first-order and the second-order gradient statistics of loss function $L(\cdot)$, respectively.

Defining I_j as the sample set of leaf node j, Γ_t can further be reformulated as follows:

$$\Gamma_t = \sum_{j=1}^{J} \left((\sum_{i \in I_j} g_j) w_j + \frac{1}{2} (\sum_{i \in I_j} (h_i + \lambda) w_j^2) \right) \tag{20}$$

where w_j represents the sample weight vector of leaves. Actually, The hyper-parameters in LightGBM have significant effect on forecasting accuracy, Table I summarizes the main hyper-parameters of LightGBM used in this work.

C. Proposed preselection algorithm

To pick out high-quality risky assets as well as utilize the forecasting results of RF and LightGBM comprehensively, we develop the preselection algorithm based on Fama-French fundamental factors. Some related empirical researches have verified the effectiveness of Fama-French fundamental factors in machine learning algorithms [45], [46]. Hence, we select Fama-French five factors as the basic features fed into machine learning algorithms. Sharpe ratio (SR) is the risk-adjusted criterion for selecting risky assets in the proposed preselection algorithm, which can be calculated as follows:

$$SR = \frac{r_p - r_f}{\sigma_p} \tag{21}$$

where r_p is the return of a portfolio, r_f is the risk-free rate, σ_p is the standard deviation of a portfolio. [47] further considered the circumstance of the negative value of r_p, and provided the modified version of SR (MSR) as follows:

$$MSR = \frac{r_p - r_f}{(\sigma_p)^{(r_p - r_f)/|r_p - r_f|}} \tag{22}$$

In both in-sample and out-of-sample analyses, SR provides an intuitive comparison of portfolio performance [48], which constitutes the reason for using this measure in our preselection algorithm. The details of the proposed preselection algorithm are revealed in Algorithm 2.

TABLE I
MAIN HYPER-PARAMETERS IN LIGHTGBM.

Hyper-parameter	Interpretation
n_estimators	The maximal number of base estimators
num_leaves	The number of leaves per tree.
learning_rate	The speed of iteration.
max_depth	Describing the maximum depth of the tree, and is capable of handling overfitting.
min_data	The minimum number of the records a leaf may have. It is also used to deal with overfitting.
n_jobs	The number of threads used in LightGBM.
feature_fraction	The fraction of features selected randomly in each iteration for building trees.
bagging_fraction	Specifying the fraction of data to be used for each iteration, also used to speed up the training and avoid overfitting.
reg_alpha	Coefficient of L_1 penalty.
reg_lambda	Coefficient of L_2 penalty.

Algorithm 2 Preselection.

Input: Data set containing Fama-French five factors and returns D, Risky assets set A, Risk-free rate r_f, Random Forest, LightGBM.

Output: Selected risky assets A_{sel}.

1: Extract D_t for training and D_p for prediction from D.
2: Train RF and LightGBM based on D_t, then use the two ensemble learning models to obtain the predictive returns series Y_r and Y_l based on D_p.
3: Calculate the standard deviation of the obtained returns series, σ_r and σ_l, respectively.
4: According to Eqs. (21) and (22), calculate SR or MSR of the forecasting returns series, S_r and S_l, respectively.
5: Sort S_r and P_r with a heap Q of size M. The sorted results are Q_s and Q_r, respectively. Record the indexes of the selected risky assets in a set I.
6: Calculate the intersection of Q_s and Q_r, Q_{sel} of size m.
7: Obtain the risky assets A_{sel} from Q_{sel} based on the indexes set I.
8: **return** A_{sel};

VI. NUMERICAL EXPERIMENTS

In this section, we implement the following numerical experiments to demonstrate the effectiveness of the proposed worst-case mean-VaR-Skewness-Kurtosis (wVSK) portfolio model with preselection (wVSK-p). As a comparison, out-of-sample performances of some baseline models such as mean-variance (MV) and mean-VaR (MVaR) are also provided. Fig. 1 shows the flowchart of the designed numerical experiments. The optimization tools used in empirical researches are YALMIP [49] and CVX [50] in MATLAB R2019b, and the machine learning algorithms within the scope of this work are programmed in Python platform.

A. Data set

Considering the efficiency of market [51], we select relatively stable and efficient financial data set for the numerical experiments. To this end, the US 12 industry portfolio daily data set from Kenneth R. French is used in the numerical experiments, where the data from Jan. 2, 2019, to Mar. 11, 2020, constitutes the training set (total 300 observations), while the data from Oct. 1, 2020, to May 28, 2021, constitutes the testing set (total 166 observations). Additionally, the data from Mar. 12, 2020, to Sep. 30, 2020, is used in the proposed preselection phase. Table II presents the descriptive statistics of the training samples. From the statistical results of the Jarque-Bera test (JB) [52], we can observe that there is no normality in the samples. Therefore, the rationality of considering higher-order moments in portfolio formation holds.

B. Results of preselection

Accuracy is a crucial indicator for evaluating machine learning algorithms. Hence, based on the existing literature, the criteria of mean square error (MSE) and mean absolute error (MAE) are defined as follows:

$$\begin{aligned} MSE &= \frac{1}{n}\sum_{i=1}^{n}(y_i-\hat{y}_i)^2 \\ MAE &= \frac{1}{n}\sum_{i=1}^{n}|y_i-\hat{y}_i| \end{aligned} \tag{23}$$

where y_i indicates the true value while $\hat{y}_i$ represents the estimated value.

Fig. 2 reveals the MSE and MAE of RF and LightGBM, respectively. The X-axis indicates the industries, the left Y-axis represents the value of MSE while the right Y-axis represents the value of MAE. In the legend, LGB is short for LightGBM and is marked with triangle symbols, whereas RF is marked with star symbols. It can be observed that RF has lower MSE than LightGBM, while LightGBM shows slightly better performance of MAE than RF.

Applying the developed Algorithm 2 to the US 12 industry data set, seven industries (Durbl, Manuf, BusEq, Telcm, Shops, Hlth, Other) are selected to construct portfolio model wVSK-p. According to the conclusion of [53], holding about seven risky assets at the same time is appropriate for individual investors, which is consistent with the result of preselection.

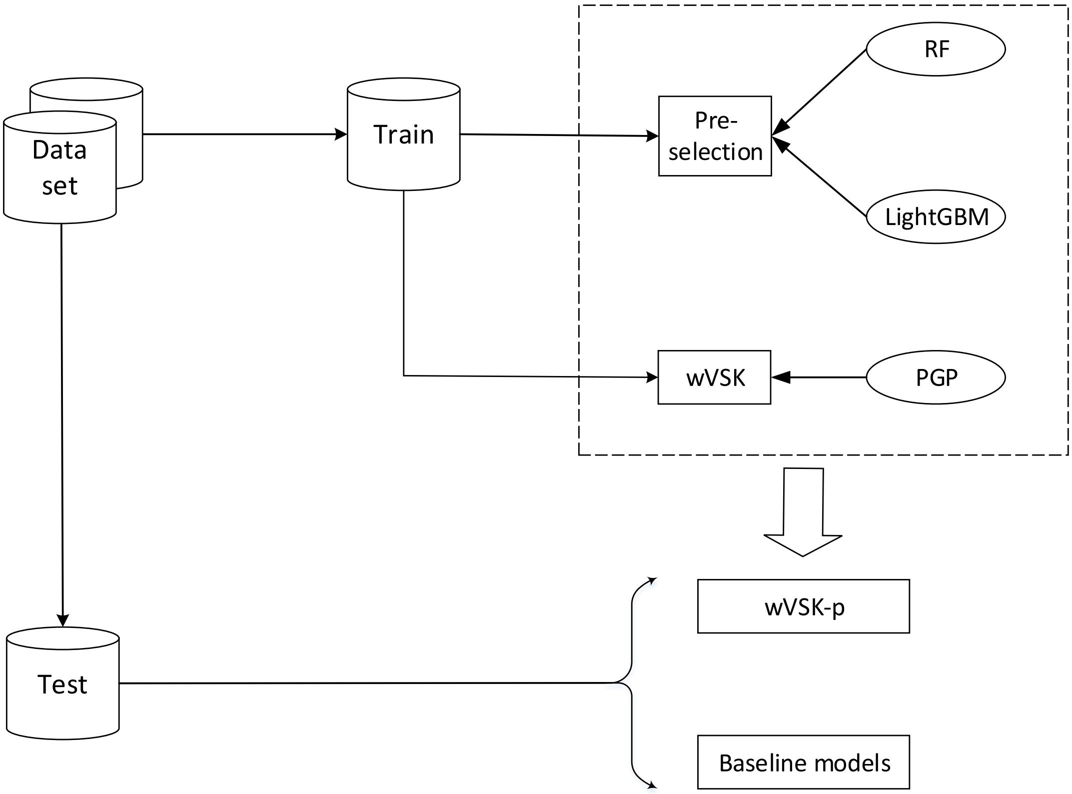

Fig. 1. Flowchart of the numerical experiments.

TABLE II
DESCRIPTIVE STATISTICS OF THE TRAINING SAMPLES.

Industry	Min($\times 10^{-2}$)	25% Quantile($\times 10^{-2}$)	Median($\times 10^{-2}$)	Mean($\times 10^{-2}$)	75% Quantile($\times 10^{-2}$)	Max($\times 10^{-2}$)	JB	p
NoDur	-5.9600	-0.2900	0.0800	0.0316	0.5150	4.5000	1054.20	0.00
Durbl	-10.2400	-0.6850	0.2300	0.0601	0.8300	6.7100	614.39	0.00
Manuf	-9.1600	-0.4900	0.0600	0.0101	0.7150	4.5200	1240.80	0.00
Enrgy	-19.7300	-0.8650	-0.0200	-0.1587	0.8250	5.4200	17074.14	0.00
Chems	-7.0300	-0.4450	0.1300	0.0117	0.5850	4.8700	1169.87	0.00
BusEq	-7.3300	-0.4250	0.2200	0.0944	0.8700	5.7900	349.55	0.00
Telcm	-6.5700	-0.4300	0.0900	0.0429	0.5800	3.9300	800.81	0.00
Utils	-7.1600	-0.3050	0.1500	0.0414	0.5250	5.4600	3013.40	0.00
Shops	-5.0300	-0.3550	0.1350	0.0503	0.5900	4.4500	399.28	0.00
Hlth	-5.6700	-0.4200	0.1000	0.0307	0.6150	4.8300	346.95	0.00
Money	-10.3500	-0.3900	0.1100	0.0185	0.6700	5.8800	2541.09	0.00
Other	-8.3300	-0.4500	0.1250	0.0159	0.6850	4.9400	1479.79	0.00

* Acronyms of the industries can be referred to the homepage of Kenneth R. French.

C. Out-of-sample performance

For the purpose of evaluating the proposed portfolio models comprehensively, some performance measures are defined as follows. ROI represents the return of investment, which measures the cumulative return of a certain portfolio. APY is short for the annualized percentage yield, which reflects the profitability of an investment intuitively. STD represents the standard deviation of a portfolio, which is a widely used indicator measuring the symmetric risk of an investment. Maximum drawdown (MDD) is the observed peak-to-trough decline during a specific period for a portfolio, which evaluating the robustness of an investment.

In the numerical experiments, we set $\lambda_1 = 1, \lambda_2 = 2, \lambda_3 = 1$ for Eq. (14) and $r_f = 3\%$ per year, the out-of-performances of different portfolio models are shown in Table III. With regard to return, wVSK-p outperforms other models with ROI 0.2978 and APY 0.4512. wVSK ranks second with ROI 0.2696 and APY 0.4063. MV obtains the best STD of 0.1261, as compared 0.1274 for MVaR, 0.1433 for wVSK, 0.1520 for wVSK-p. Obviously, the conservatism of the conventional worst-case mean-VaR model can be overcome by considering higher-order moments in our numerical experiments. In terms of SR, wVSK-p has the highest level of 2.7711, wVSK follows with 2.6261. However, wVSK reaches the best MDD of 0.0504, followed by MVaR, with 0.0548. For the tail risk, MVaR has the lowest VaR($\%5$) of 0.0119, then MV,

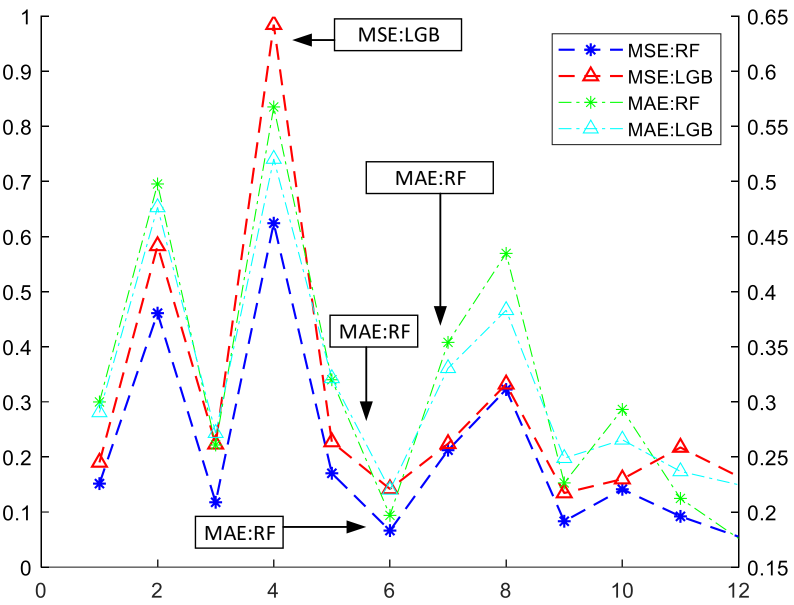

Fig. 2. Flowchart of the numerical experiments.

TABLE III
OUT-OF-SAMPLE PERFORMANCE OF PORTFOLIO MODELS.

	MV	MVaR	wVSK	wVSK-p
ROI	0.1947	0.1897	0.2696	**0.2978**
APY	0.2893	0.2817	0.4063	**0.4512**
STD	**0.1261**	0.1274	0.1433	0.1520
SR	2.0560	1.9752	2.6261	**2.7711**
MDD	0.0557	0.0548	**0.0504**	0.0720
VaR(5%)	0.0124	**0.0119**	0.0152	0.0154
Skewness	-0.4572	-0.4718	**-0.1418**	-0.4483
Kurtosis	4.0010	3.9669	4.1068	**3.7874**

0.0124. The values of skewness and kurtosis verify the actual effectiveness of higher-order moments in portfolio formation, where wVSK obtains the highest skewness of -0.1418 and wVSK-p has the lowest kurtosis of 3.7874.

Fig. 3 is visualizes the result of Table III and further demonstrate the superiority of the proposed portfolio models. During the whole test period, wVSK shows superior performance than the baseline portfolio models, but wVSK-p performs unsatisfactorily at the beginning of the test period. Considering less risky assets could be selected after preselection, wVSK-p has more difficult in diversifying risk. Thus, wVSK-p shows higher STD, MDD, VaR($\%5$) than wVSK model, which means assuming more risk to pursue higher return. Although more robust out-of-sample performance wVSK has, preselection is still a process that deserves serious consideration in the circumstance of a large number of risky assets to be managed. In a nutshell, higher-order moments play a key role in overcoming the conservatism of the worst-case mean-VaR portfolio optimization, and more analysis is to be implemented in the next section.

D. Sensitivity analysis

In this section, more combinations of the preference tuple $(\lambda_1, \lambda_2, \lambda_3)$ are carried out to further investigate the effectiveness of skewness and kurtosis of the proposed portfolio models. Fig. 4 presents the VaR($\%5$) and SR of wVSK and wVSK-p portfolio models, where $\lambda_1 = 1, \lambda_3 = 1$ are fixed while λ_2 varies within a range of $[1, 3]$. With the increase of λ_2, the VaR($\%5$) of wVSK shows a gradual downward trend, whereas the SR of wVSK inclines to rise. However, similar trends do not obviously show in the wVSK-p model. Both VaR($\%5$) and SR presents a state of oscillation in wVSK-p because of the existence of preselection. Fig. 5 shows the VaR($\%5$) and SR of wVSK and wVSK-p, where $\lambda_1 = 1, \lambda_2 = 2$ are fixed while λ_3 varies within a range of $[0.5, 2.5]$. It can be observed that the values of VaR($\%5$) for wVSK and wVSK-p decrease with the increasing value of λ_3. The financial models are paying more attention to curbing kurtosis incline to obtain portfolios with thinner tails, which is beneficial to the performance of VaR. Additionally, all combinations of the preference parameters result in wVSK and wVSK-p portfolio models with higher SR

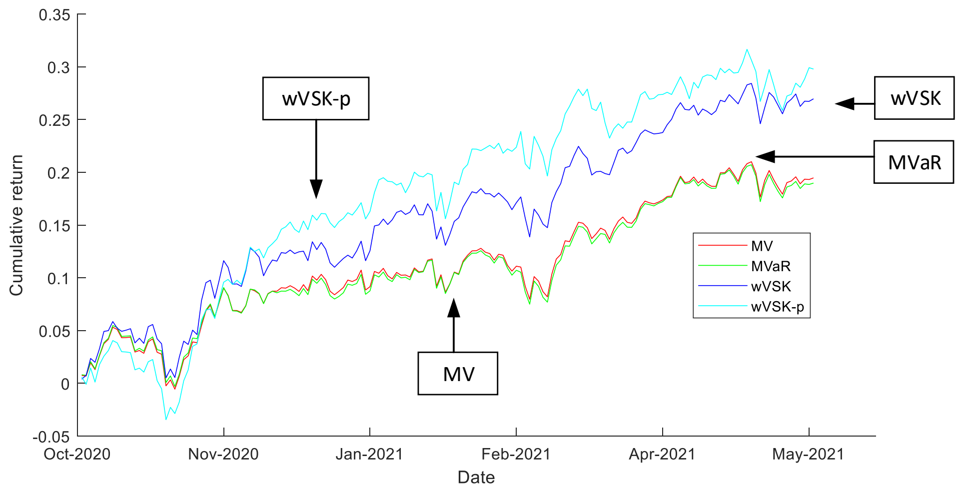

Fig. 3. Cumulative returns of different portfolio models.

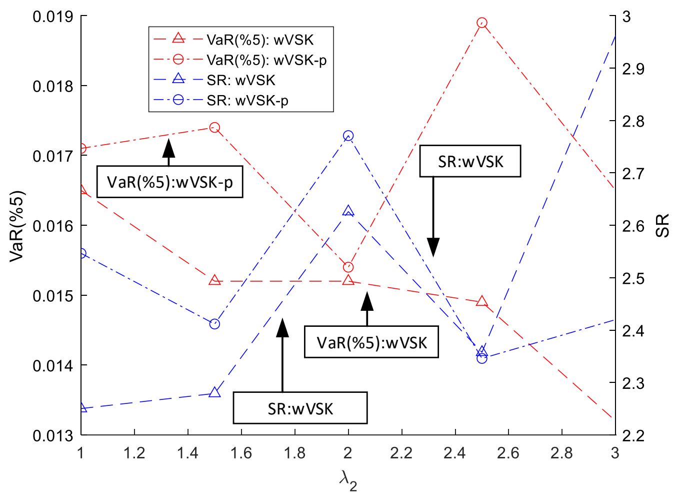

Fig. 4. VaR$(\%5)$ and SR of portfolios with different λ_2.

than MV and MVaR in the numerical experiments, which verifies that the proposed portfolio models can overcome the conservatism of the conventional worst-case robust portfolio models while maintaining the robustness of the primal optimization models to some extent.

VII. Conclusions & Discussions

Conservatism is one of the main issues in robust portfolio optimization, limiting the large-scale use of this type of model by financial practitioners. Without the assumption of normality, higher-order moments encompass intuitive approaches to obtain less conservative portfolios. Forecasting information also provides feasible methods to overcome the conservatism, which could be achieved and utilized by modern artificial intelligence techniques. This paper aims to develop the worst-case mean-VaR portfolio models considering higher-order moments, where the preselection involving Random Forest and LightGBM algorithms is designed and implemented. Some attributes concerning mean, VaR, variance, skewness, and kurtosis are integrated into the proposed portfolio models in the form of PGP.

The comparative result of the numerical experiments demonstrates the effectiveness of the proposed models. It can be seen from the out-of-sample performance of wVSK that higher ROI and APY does wVSK has, while not increasing the level of risk significantly compared to MV and MVaR models. Due to the criterion for selecting risky assets, preselection further improves the model performance in terms of return while reducing the robustness of the primal proposed model. Constrained with the number of risky assets that could be selected, some risk indicators such as STD, MDD, and VaR$(\%5)$ are exacerbated on adaptation with preselection. However, it is still a necessary process when individual investors face lots of risky assets to manage. In future research, we

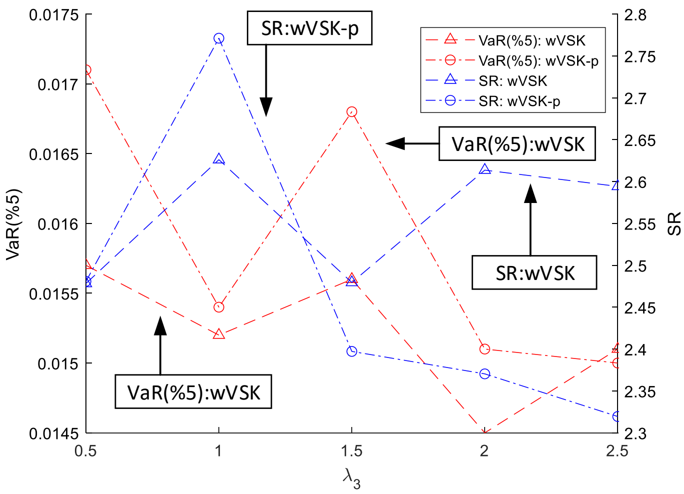

Fig. 5. VaR$(\%5)$ and SR of portfolios with different λ_3.

would develop more multi-stage portfolio models in the proposed framework and risk diversification would also be considered to enrich the existing framework.

Acknowledgments

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 thank Xiangxi Kong, a Ph.D. candidate from the College of Business, Shanghai University of Finance & Economics.

References

[1] H. M. Markowitz, "Portfolio selection," *The Journal of Finance*, vol. 7, no. 1, p. 77, 1952.
[2] E. F. Fama and K. R. French, "Size, value, and momentum in international stock returns," *Journal of financial economics*, vol. 105, no. 3, pp. 457–472, 2012.
[3] ——, *Common risk factors in the returns on stocks and bonds*. University of Chicago Press, 2021.
[4] D. Goldfarb and G. Iyengar, "Robust portfolio selection problems," *Mathematics of operations research*, vol. 28, no. 1, pp. 1–38, 2003.
[5] A. Thiele, "A note on issues of over-conservatism in robust optimization with cost uncertainty," *Optimization*, vol. 59, no. 7, pp. 1033–1040, 2010.
[6] E. Roos and D. den Hertog, "Reducing conservatism in robust optimization," *INFORMS Journal on Computing*, vol. 32, no. 4, pp. 1109–1127, 2020.
[7] S. Lotfi, M. Salahi, and F. Mehrdoust, "Adjusted robust mean-value-at-risk model: less conservative robust portfolios," *Optimization and Engineering*, vol. 18, no. 2, pp. 467–497, 2017.
[8] S. Lotfi and S. A. Zenios, "Robust var and cvar optimization under joint ambiguity in distributions, means, and covariances," *European Journal of Operational Research*, vol. 269, no. 2, pp. 556–576, 2018.
[9] L. Min, J. Dong, J. Liu, and X. Gong, "Robust mean-risk portfolio optimization using machine learning-based trade-off parameter," *Applied Soft Computing*, vol. 113, p. 107948, 2021.
[10] C. R. Harvey, J. C. Liechty, M. W. Liechty, and P. Müller, "Portfolio selection with higher moments," *Quantitative Finance*, vol. 10, no. 5, pp. 469–485, 2010.
[11] T. Joro and P. Na, "Portfolio performance evaluation in a mean–variance–skewness framework," *European Journal of Operational Research*, vol. 175, no. 1, pp. 446–461, 2006.
[12] B. Chen, J. Zhong, and Y. Chen, "A hybrid approach for portfolio selection with higher-order moments: Empirical evidence from shanghai stock exchange," *Expert Systems with Applications*, vol. 145, p. 113104, 2020.
[13] M. Aksaraylı and O. Pala, "A polynomial goal programming model for portfolio optimization based on entropy and higher moments," *Expert Systems with Applications*, vol. 94, pp. 185–192, 2018.
[14] L. Yu, S. Wang, and K. K. Lai, "Neural network-based mean–variance–skewness model for portfolio selection," *Computers & Operations Research*, vol. 35, no. 1, pp. 34–46, 2008.
[15] W. Wang, W. Li, N. Zhang, and K. Liu, "Portfolio formation with preselection using deep learning from long-term financial data," *Expert Systems with Applications*, vol. 143, p. 113042, 2020.
[16] W. Chen, H. Zhang, M. K. Mehlawat, and L. Jia, "Mean–variance portfolio optimization using machine learning-based stock price prediction," *Applied Soft Computing*, vol. 100, p. 106943, 2021.
[17] T. Hai and L. Min, "Hybrid robust portfolio selection model using machine learning-based preselection," *Engineering Letters*, vol. 29, no. 4, pp. 1626–1635, 2021.
[18] L. Haff, "An identity for the wishart distribution with applications," *Journal of Multivariate Analysis*, vol. 9, no. 4, pp. 531–544, 1979.
[19] K. Schöttle and R. Werner, "Robustness properties of mean-variance portfolios," *Optimization*, vol. 58, no. 6, pp. 641–663, 2009.
[20] F. D. Arditti and H. Levy, "Portfolio efficiency analysis in three moments: the multiperiod case," in *Financial Dec Making under Uncertainty*. Elsevier, 1977, pp. 137–150.
[21] H. Levy, "The rationale of the mean-standard deviation analysis: Comment," *The American Economic Review*, vol. 64, no. 3, pp. 434–441, 1974.
[22] P. A. Samuelson, "The fundamental approximation theorem of portfolio analysis in terms of means, variances and higher moments," in *Stochastic optimization models in finance*. Elsevier, 1975, pp. 215–220.
[23] W. Chen, Y. Wang, P. Gupta, and M. K. Mehlawat, "A novel hybrid heuristic algorithm for a new uncertain mean-variance-skewness portfolio selection model with real constraints," *Applied Intelligence*, vol. 48, no. 9, pp. 2996–3018, 2018.
[24] J. Zhai, M. Bai, and J. Hao, "Uncertain random mean–variance–skewness models for the portfolio optimization problem," *Optimization*, pp. 1–24, 2021.
[25] S. Ashfaq, U. Ayub, G. Mujtaba, N. Raza, and S. Gulzar, "Gainers and losers with higher order portfolio risk optimization," *Physica A: Statistical Mechanics and its Applications*, vol. 563, p. 125416, 2021.
[26] B. AracioĞlu, F. Demircan, and H. Soyuer, "Mean–variance–skewness–kurtosis approach to portfolio optimization: An application in istanbul stock exchange," *Ege Academic Review*, vol. 11, no. 5, pp. 9–17, 2011.
[27] K. K. Lai, L. Yu, and S. Wang, "Mean-variance-skewness-kurtosis-based portfolio optimization," in *First International Multi-*

Symposiums on Computer and Computational Sciences (IMSCCS'06), vol. 2. IEEE, 2006, pp. 292–297.

[28] D. Maringer and P. Parpas, "Global optimization of higher order moments in portfolio selection," *Journal of Global optimization*, vol. 43, no. 2, pp. 219–230, 2009.

[29] P. Chunhachinda, K. Dandapani, S. Hamid, and A. J. Prakash, "Portfolio selection and skewness: Evidence from international stock markets," *Journal of Banking & Finance*, vol. 21, no. 2, pp. 143–167, 1997.

[30] T.-Y. Lai, "Portfolio selection with skewness: a multiple-objective approach," *Review of Quantitative Finance and Accounting*, vol. 1, no. 3, pp. 293–305, 1991.

[31] R. Chowdhury, M. Mahdy, T. N. Alam, G. D. Al Quaderi, and M. A. Rahman, "Predicting the stock price of frontier markets using machine learning and modified black–scholes option pricing model," *Physica A: Statistical Mechanics and its Applications*, vol. 555, p. 124444, 2020.

[32] R. Chowdhury, M. A. Rahman, M. S. Rahman, and M. Mahdy, "An approach to predict and forecast the price of constituents and index of cryptocurrency using machine learning," *Physica A: Statistical Mechanics and its Applications*, vol. 551, p. 124569, 2020.

[33] Y. Freund and R. E. Schapire, "A decision-theoretic generalization of on-line learning and an application to boosting," *Journal of computer and system sciences*, vol. 55, no. 1, pp. 119–139, 1997.

[34] Z.-H. Zhou, *Ensemble methods: foundations and algorithms*. Chapman and Hall/CRC, 2019.

[35] T. A. Borges and R. F. Neves, "Ensemble of machine learning algorithms for cryptocurrency investment with different data resampling methods," *Applied Soft Computing*, vol. 90, p. 106187, 2020.

[36] J. T. Kent, "Information gain and a general measure of correlation," *Biometrika*, vol. 70, no. 1, pp. 163–173, 1983.

[37] B. Azhagusundari, A. S. Thanamani *et al.*, "Feature selection based on information gain," *International Journal of Innovative Technology and Exploring Engineering (IJITEE)*, vol. 2, no. 2, pp. 18–21, 2013.

[38] L. Breiman, "Random forests," *Machine learning*, vol. 45, no. 1, pp. 5–32, 2001.

[39] G. Ke, Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma, Q. Ye, and T.-Y. Liu, "Lightgbm: A highly efficient gradient boosting decision tree," *Advances in Neural Information Processing Systems*, vol. 30, pp. 3146–3154, 2017.

[40] J. H. Friedman, "Greedy function approximation: a gradient boosting machine," *Annals of Statistics*, pp. 1189–1232, 2001.

[41] ——, "Stochastic gradient boosting," *Computational statistics & data analysis*, vol. 38, no. 4, pp. 367–378, 2002.

[42] T. Chen, T. He, M. Benesty, V. Khotilovich, Y. Tang, H. Cho *et al.*, "Xgboost: extreme gradient boosting," *R Package Version 0.4-2*, vol. 1, no. 4, pp. 1–4, 2015.

[43] T. Chen and 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.

[44] X. Sun, M. Liu, and Z. Sima, "A novel cryptocurrency price trend forecasting model based on lightgbm," *Finance Research Letters*, vol. 32, p. 101084, 2020.

[45] L. Min, J. Dong, D. Liu, and X. Kong, "A black-litterman portfolio selection model with investor opinions generating from machine learning algorithms," *Engineering Letters*, vol. 29, no. 2, pp. 710–721, 2021.

[46] S. Gu, B. Kelly, and D. Xiu, "Empirical asset pricing via machine learning," *The Review of Financial Studies*, vol. 33, no. 5, pp. 2223–2273, 2020.

[47] C. L. Israelsen *et al.*, "A refinement to the sharpe ratio and information ratio," *Journal of Asset Management*, vol. 5, no. 6, pp. 423–427, 2005.

[48] M. Caporin, G. M. Jannin, F. Lisi, and B. B. Maillet, "A survey on the four families of performance measures," *Journal of Economic Surveys*, vol. 28, no. 5, pp. 917–942, 2014.

[49] J. Lofberg, "Yalmip: A toolbox for modeling and optimization in matlab," in *2004 IEEE international conference on robotics and automation (IEEE Cat. No. 04CH37508)*. IEEE, 2004, pp. 284–289.

[50] M. Grant, S. Boyd, and Y. Ye, "Cvx: Matlab software for disciplined convex programming," 2009.

[51] B. G. Malkiel, "The efficient market hypothesis and its critics," *Journal of economic perspectives*, vol. 17, no. 1, pp. 59–82, 2003.

[52] T. Thadewald and H. Büning, "Jarque–bera test and its competitors for testing normality–a power comparison," *Journal of Applied Statistics*, vol. 34, no. 1, pp. 87–105, 2007.

[53] F. D. Paiva, R. T. N. Cardoso, G. P. Hanaoka, and W. M. Duarte, "Decision-making for financial trading: A fusion approach of machine learning and portfolio selection," *Expert Systems with Applications*, vol. 115, pp. 635–655, 2019.