

Research Article

Research on Risk Identification System Based on Random Forest Algorithm-High-Order Moment Model

Li-Jun Liu ¹, Wei-Kang Shen ¹ and Jia-Ming Zhu ²

¹School of Economics, Hebei GEO University, Shijiazhuang 050031, China

²Institute of Quantitative Economics, Anhui University of Finance and Economics, Bengbu 233030, China

Correspondence should be addressed to Jia-Ming Zhu; zhujm1973@163.com

Received 2 March 2021; Revised 16 March 2021; Accepted 24 March 2021; Published 8 April 2021

Academic Editor: Huihua Chen

Copyright © 2021 Li-Jun Liu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

With the continuous development of the stock market, designing a reasonable risk identification tool will help to solve the irrational problem of investors. This paper first selects the stocks with the most valuable investment value in the future through the random forest algorithm in the nine-factor model and then analyzes them by using the higher-order moment model to find that different investors' preferences will make the weight of the portfolio change accordingly, which will eventually make the optimal return and risk set of the composition of the portfolio change. The risk identification system designed in this paper can provide an effective risk identification tool for investors and help them make rational judgments.

1. Introduction

As China's stock market continues to develop, how to identify the risk of many financial assets and investor preferences and reduce the bias of investors' investment behavior due to speculative and subjective nature has become a growing concern for investors. At present, there are three main ways for investors to evaluate the stock market: first, technical analysis, which is based on the study of stock price fluctuation patterns; second, fundamental analysis, which is based on the study of the intrinsic value of stocks; and third, quantitative analysis, which is based on the modeling of historical information of stocks. With the development of data analysis technology, quantitative analysis has become the mainstream research approach; and from the perspective of strategic stock selection, the value-growth investment strategy is a perfect blend of traditional value-based and active growth investment strategies, which considers assets that are somehow undervalued at the current stage but at the same time have a better potential for sustained growth, to have room for investment. Therefore, stock selection using the value-growth investment strategy can yield more stable investment returns; in terms of indicator stock selection, stock selection is actually the study of stock

impact factors, and stock prices are determined by a multidimensional spatial system of factor composition. Therefore, stock selection actually explores the problem of optimal classification in a multidimensional space.

With the continuous development in recent years, many scholars have applied machine learning methods to stock selection strategies; Fan and Palaniswami were the first to apply support vector machine algorithms to stock selection [1]; Kim et al. also used vector machines for stock price prediction with good results [2]; Yu et al. used support vector machines for stock market prediction by combining the selection properties of genetic algorithms to improve the efficiency of the model [3]; Genuer et al., Bin Li et al., and Ladyzynski et al. applied the random forest algorithm to the classification prediction problem with relatively good experimental results [4–6]; and Heaton et al. studied the application on reasset pricing and financial risk control based on deep learning [7].

In terms of investor risk identification, Sharpe proposed the CAPM model reflecting the relationship between mean return and variance risk, which laid the foundation for risk identification [8]; Chen et al. proved the existence problem of the most portfolio with kurtosis containing skewness in 2006 and used the method of segment-by-segment linear

approximation and nonlinear transformation in 2007 by converting it into a solvable linear programming problem for solution [9–11]; Tang and Tang established a fluctuation model of higher-order moments and a theoretical framework of higher-order moments on CAPM based on the definition of higher-order moments [12]; Peng studied the distribution of risky asset returns from the perspective of having time variability and used the portfolio complete distribution information to study the optimal optimization problem [13]; Wang et al. proposed a multifactor structure state static MV model by proving the optimal solution value under the condition that short selling is not allowed [14]; Níguez et al. studied the effect of higher-order risk attitudes and statistical moments on the optimal allocation of risky assets in a standard portfolio selection model, using data from the United States to illustrate the introduction of these higher-order effects on the decision to maximize expected utility relative importance [15]; and Yue et al. provided a corresponding optimization treatment of the higher-order moments model [16–18].

The random forest algorithm and multifactor model are better used in stock selection, while the higher-order moment model is also more comprehensive in risk classification compared to the second-order moment. This paper is a risk identification system based on data mining methods and using a mature factor stock selection system, combined with investor preferences, designed to provide investors with effective stock risk identification tools, establish a more rational investment philosophy, to better promote the sound development of the overall stock market.

2. Data Sources, Data Processing, and Assumptions

2.1. Source of Data. The data in this paper are obtained from the RESET database. The data selected in the stock selection phase are all listed company stock data, with the time date of January 1, 2016, to December 31, 2019, and the data type is quarterly; the data selected in the risk identification phase are daily stock returns after stock selection, with the time date of January 1, 2018, to December 31, 2019, and the data type is daily.

2.2. Processing of Data. Due to the large amount of data in the raw data and the presence of missing values, this paper preprocesses the data according to the completeness as well as the usefulness of the information. The main processing is divided into the following processes: data sieving. The original data has more redundancy and missing values. Therefore, firstly, null records are removed for screening. Secondly, this paper considers stocks with normal business status, so the ST stock category is removed: data integration. The data are classified in the stock selection stage, where two classifications are chosen in this paper: dichotomous classification and quintuple classification. In the second classification, those with quarterly returns greater than 0 are classified as 1 class and those with quarterly returns less than 0 are classified as 0 class; in the fifth classification, those with

quarterly returns greater than 0.05 are classified as 2 classes, those with quarterly returns less than 0.05 but greater than 0.01 are classified as 1 class, those with quarterly returns greater than less than 0.01 but greater than -0.01 are classified as 0 class the quarterly return less than -0.01 but greater than -0.05 is classified as -1 , and the quarterly return less than -0.05 is classified as -2 .

2.3. Research Hypothesis

2.3.1. Hypothesis 1: The Multifactor Stock Selection Model Based on the Random Forest Algorithm Is Effective in Stock Selection Strategy. The random forest algorithm is widely used in stock selection strategies, and in combination with a well-established multifactor model, it makes the results of stock selection reliable. This paper also uses the random forest algorithm and the nine-factor model for stock selection and refers to a large amount of literature to show that the method has high accuracy and stability. Therefore, the first hypothesis of this paper is that the multifactor model based on the random forest algorithm is also effective in this paper.

2.3.2. Hypothesis 2: A Risk Set Consisting of Variance, Skewness, and Kurtosis Is a More Effective Measure of Risk. Traditionally, the volatility of risk can be measured by variance, but the mean and antivarience measures alone are not sufficient for investment strategies. When financial time series meet normal distribution, mean and variance measures are valid; however, in reality, most financial time series do not meet normal distribution, but more skewed distribution, so skewness and kurtosis coefficients need to be added to measure together. This paper proposes the hypothesis that the risk set consisting of variance, skewness, and kurtosis is a more effective measure of risk.

2.3.3. Hypothesis 3: Investors with Different Risk Preferences Differ in Their Asset Allocation Weights, but Portfolio Assets Are Consistent with the General Laws of Finance. Different investors differ in their risk preferences due to differences in sentiment, information gathering, and so forth, which ultimately leads to differences in the weighting of resources in the asset pool. Under the assumption that the asset pool is the same, the optimal return of the portfolio will be different due to the difference in the weighting of different investment preferences. However, the portfolio assets still satisfy the category of financial assets and conform to the basic laws of finance. Therefore, this paper puts forward the hypothesis that there are differences in the weighting of the assets allocated by investors with different risk preferences, but the portfolio assets conform to the general laws of finance.

3. Stock Selection Strategy and Test Based on Random Forest Algorithm

3.1. Random Forest Stock Selection Results. The random forest algorithm is essentially a collection of tree classifiers, the base classifier constructed by a decision tree algorithm

without pruning, and the final one based on majority voting to determine the classification result [19–22]. The random forest algorithm is a decision tree model formed based on the bagging framework, which contains the set of all trees [23], and the process of forming each tree is shown in Figure 1.

The factor selection of this model mainly refers to the research results of Cao et al.'s nine-factor stock selection model [21] and Wang et al.'s eight-factor stock selection model [20] and adds appropriate factors on the basis of the value-growth strategy [24–26], so that the whole factor system contains both value and growth factors. Therefore, the factors selected in this paper are net sales margin reflecting the profitability index of the company's sales revenue; current ratio reflecting the company's short-term solvency; long-term debt ratio reflecting the company's long-term solvency; earnings per share growth rate reflecting the growth level of the company's stock earnings; total assets growth rate reflecting the growth level of the company's total assets; and return on assets reflecting the level of profitability of the company. These indicators are divided into value and growth factors according to the factor dimension, while the response factors are the corresponding quarterly returns. Table 1 shows the information about each impact factor.

Table 2 shows the information on stock selection by the random forest algorithm. Since the random forest algorithm has a better model prediction in stock selection [21], the top six stocks in terms of the voting score are selected to construct the portfolio under the condition that they satisfy the prediction categorized as 2 under five classifications and as 1 under two classifications.

3.2. Test of Stock Selection Results. Table 3 shows the tests of the above stock selection results. The tests are selected for the monthly return, monthly risk-free return, and an annual return of the corresponding stocks in 2020. Comparing the quarterly returns of the corresponding stocks with the risk-free rate, it is found that 3/4 of the stocks with code 000768 have a greater share than the risk-free rate; 7/12 of the stocks with codes 000897, 600313, and 600416 have a greater share than the risk-free rate; and almost half of the stocks with codes 600728 and 600737 have a greater share than the risk-free rate, which indicates that the stocks selected by the random forest have a higher

probability of having investment value; from the annual return of the corresponding stocks, almost all of them are greater than the annual risk-free rate, and the annual return of codes 000768, 600313, and 6000416 is more than 50 times the annual risk-free rate, which reflects that the random forest algorithm and the nine-factor model can better select stocks with investment value for investors. This reflects that the random forest algorithm and the nine-factor model can better select stocks with investment value for investors.

4. Risk Classification Based on Higher-Order Moment Model

4.1. Preparation of the Model. Based on Markowitz's mean-variance model, third-order moments (skewness) and fourth-order moments (kurtosis) are added to measure asymmetric risk and kurtosis risk of financial assets, forming a portfolio model with higher-order moment risk including skewness and kurtosis. It is assumed that the market has no transaction costs, no taxes, and no short selling and that the assets in the market can be split indefinitely. At the beginning of the period, investors allocate their wealth across risky assets in N risky assets in the proportion $X = (X_1, X_2, \dots, X_N)'$ and calculate the investment return on the N assets at the end of the period by setting its return vector as $R = (R_1, R_2, \dots, R_N)'$ and the expected return vector is $\mu = E(R) = (E(R_1), \dots, E(R_N))'$, so that a portfolio consisting of N risky assets is formed, and the investment return of the portfolio is $R_p = X'R$.

H is the $N \times N$ dimensional variance-covariance matrix of the asset portfolio (shorthand covariance array), S is the $N \times N^2$ dimensional skewness-covariance matrix of the asset portfolio (shorthand covariance array), and K is the $N \times N^3$ dimensional kurtosis-covariance matrix of the asset portfolio (shorthand covariance array), defined as follows:

$$\begin{cases} H = E[(R - \mu)(R - \mu)'] = \{h_{ij}\}, \\ S = E[(R - \mu)(R - \mu)' \otimes (R - \mu)'] = \{s_{ijm}\}, \\ K = E[(R - \mu)(R - \mu)' \otimes (R - \mu)' \otimes (R - \mu)] = \{k_{ijmn}\}, \end{cases} \quad (1)$$

in the formula

$$\begin{cases} \{h_{ij}\} = E[(R_i - \mu_i)(R_j - \mu_j)], & i, j = 1, 2, \dots, N, \\ \{s_{ijm}\} = E[(R_i - \mu_i)(R_j - \mu_j)(R_m - \mu_m)], & i, j, m = 1, 2, \dots, N, \\ \{k_{ijmn}\} = E[(R_i - \mu_i)(R_j - \mu_j)(R_m - \mu_m)(R_n - \mu_n)], & i, j, m, n = 1, 2, \dots, N. \end{cases} \quad (2)$$

Calculate the expectation, variance, skewness, and kurtosis of the portfolio:

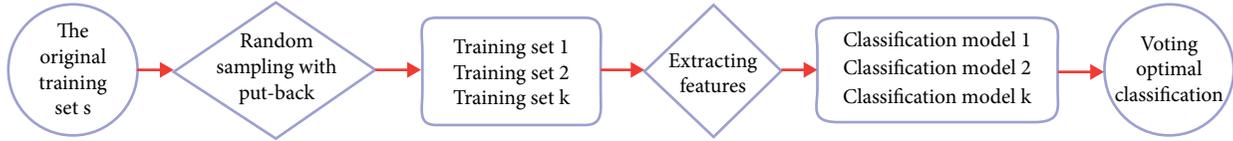


FIGURE 1: Formation rules for each tree.

TABLE 1: The information about each impact factor.

Index serial number	Index name	Index abbreviation	Factor dimension	Indicator meaning
1	Net profit rate of sales	NPR	Value factor	Profitability
2	Current ratio	CRT	Value factor	Solvency
3	Long-term debt ratio	LDR	Value factor	Solvency
4	Growth rate of earnings per share	EPS	Growth factor	Growth ability
5	Growth rate of total assets	TRR	Growth factor	Growth ability
6	Turnover of total assets	TRA	Growth factor	Operating capacity
7	Asset liability ratio	DR	Value factor	Operating capacity
8	Average return on net assets	ROE	Growth factor	Profitability
9	Return on assets	ROA	Growth factor	Profitability

TABLE 2: Random forest stock selection results.

Stock code	Preds-NQMclass5	Preds-NQMclass2	Poll count ranking
000768	2	1	1
000897	2	1	2
600313	2	1	3
600416	2	1	4
600728	2	1	5
600737	2	1	6

TABLE 3: Test of stock selection results.

Stock code	≥Proportion of RF	<Proportion of RF	Y-ret
000768	3/4	1/4	1.2466
000897	7/12	5/12	0.1857
600313	7/12	5/12	1.0922
600416	7/12	5/12	1.9384
600728	5/12	7/12	-0.1866
600737	1/2	1/2	0.159

$$\begin{cases}
 \mu_p(X) = E(R_p) = X'E(R) = X'\mu, \\
 \sigma_p^2(X) = E[(R_p - \mu_p)^2] = E[X'(R - \mu)]^2 = X'HX, \\
 s_p^3(X) = E[(R_p - \mu_p)^3] = E[X'(R - \mu)]^3 = X'S(X \otimes X), \\
 k_p^4(X) = E[(R_p - \mu_p)^4] = E[X'(R - \mu)]^4 = X'K(X \otimes X \otimes X).
 \end{cases} \quad (3)$$

The above equation is the expectation, variance, skewness, and kurtosis of the portfolio and holds under the assumption that the expectation, variance, skewness, and kurtosis of the risky asset return $R_i, i = 1, 2, \dots, N$ exist, where \otimes denotes the Kronecker product of matrices, H is the covariance array of the portfolio, S is the covariance array, and K is the covariance kurtosis array.

4.1.1. Calculating Higher-Order Moment Risk. Figures 2–7 show the volatility of the return series for each portfolio member separately. It can be found that there is a large variability in return volatility for each stock in terms of amplitude, aggregation, and the corresponding point in time. Not only is the volatility of each stock's return dramatic, but the different volatility states also reflect the variability of each stock's risk.

The distribution state of financial time series has an important influence on the regular variation of financial time series. When the distribution of financial time series satisfies the normal distribution, the regular variation of financial time series can be better analyzed by the first-order moment mean and second-order moment variance (i.e., mean-variance model). However, when the financial time series does not satisfy the normal distribution, it is far from sufficient to describe and analyze the characteristics only by means and variances of two orders of moments. In particular, in the perspective of risk measurement, nonnormal distribution implies the existence of asymmetric risk and kurtosis risk. Asymmetric risk is represented by the skewness coefficient, which has a negative value, and a significant left “thick tail”; if the skewness coefficient is positive, it has a significant right “thick tail.” The kurtosis risk is expressed by the kurtosis coefficient; if the kurtosis coefficient is positive, it means that the distribution of the series is steeper than the normal distribution; and if the kurtosis coefficient is negative, it means that the distribution of the series is more moderate than the normal distribution. According to the characteristics of financial markets, there are significant left deviation and overpeak effects in the time series of financial market returns. The left deviation means that the probability of a decline in the time series of returns is much greater than the probability of a rise in the time series of returns, and the superpeak means that the probability of the occurrence of extreme values is also greatly increased, and the phenomenon of “spikes” appears. Therefore, to deal with the actual

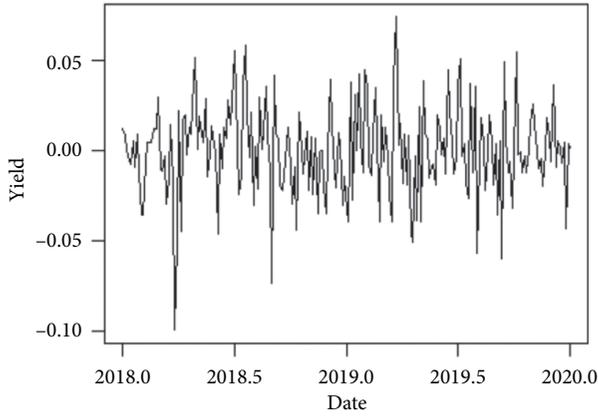


FIGURE 2: Yield of C000768.

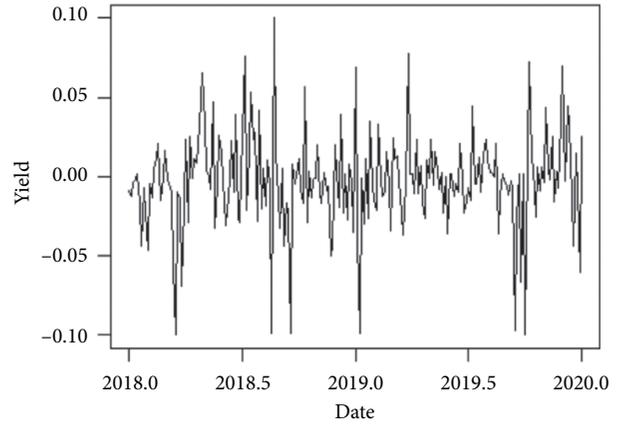


FIGURE 5: Yield of C600416.

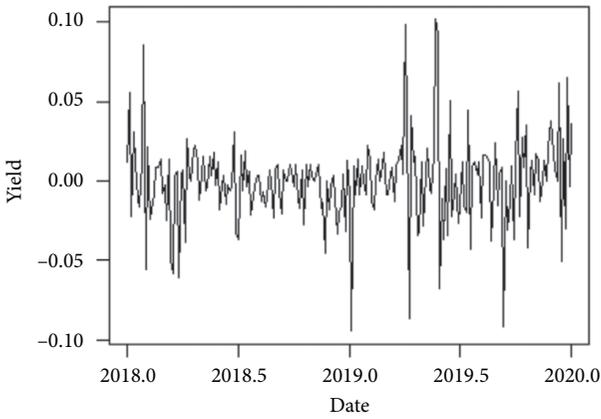


FIGURE 3: Yield of C000897.

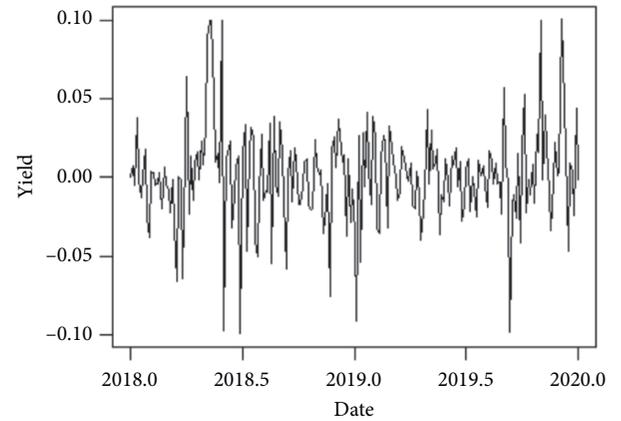


FIGURE 6: Yield of C600728.

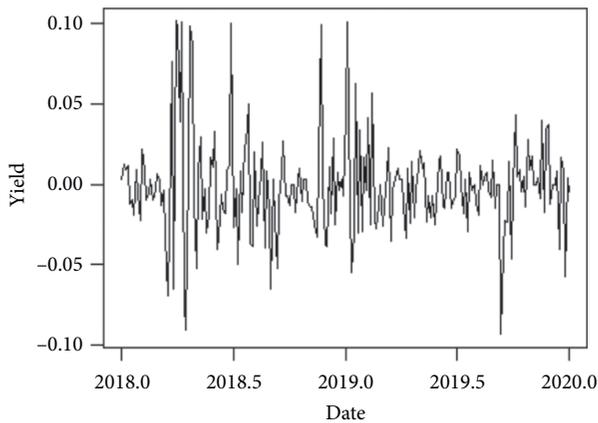


FIGURE 4: Yield of C600313.

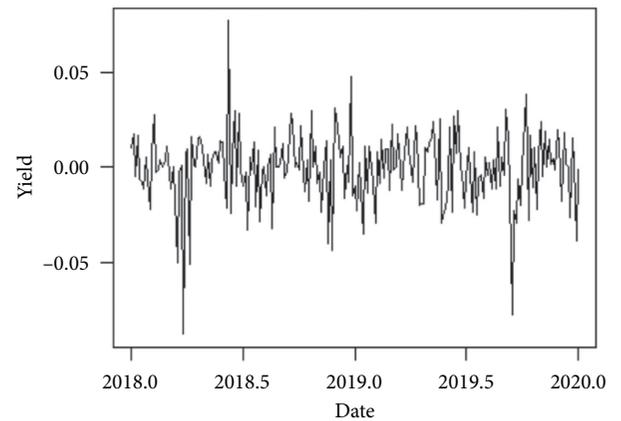


FIGURE 7: Yield of C600737.

more general financial time series, the third-order moment skewness is needed to measure asymmetric risk and the fourth-order moment kurtosis is needed to measure kurtosis risk.

The variance, skewness, and kurtosis of the returns are calculated according to equations (1)–(3), and in Table 4, descriptive statistics are given for each stock return series. During the sample observation period, only the third and

fifth columns of stocks in Table 4 have negative mean returns, while the rest are positive. And in terms of skewness, only the fifth column is negative, implying that there is a possibility of falling returns; the rest are all positive which implies that there is an upward trend in the returns of the other stocks. The kurtosis statistic indicates a thicker tail characteristic than the normal distribution. The test of JB statistic shows that all stock return series obey nonnormal

TABLE 4: Basic information of yield.

Classification	C000768	C000897	C600313	C600416	C600728	C600737
Mean	0.000	-0.001	0.000	-0.001	0.001	0.000
Median	0.000	0.000	0.000	-0.002	0.001	0.000
Maximum	0.100	0.102	0.102	0.100	0.100	0.100
Minimum	-0.100	-0.100	-0.101	-0.100	-0.100	-0.099
Sd	0.022	0.026	0.032	0.029	0.029	0.022
Skew	0.108	0.253	0.595	-0.107	0.224	0.124
Kurt	5.697	6.261	5.454	4.954	5.430	7.131
JB	146.064	217.272	148.451	77.086	121.848	341.889
P value	0.000	0.000	0.000	0.000	0.000	0.000
Normality	Non	Non	Non	Non	Non	Non

distribution. From the overall results, it is clear that the portfolio analysis through the second-order moments model alone is too limited and does not meet the actual distribution needs. Therefore, a more precise and quantitative analysis of financial risk can be conducted through the overall research and analysis from the perspective of higher-order moments.

5. Risk Identification Based on PGP Technique: High-Order Moment Model

5.1. *Constructing the M-V-S-K Model.* A multiobjective optimization technique is used to combine the four conflicting individual objectives in equation (3). The portfolio model with higher-order moment risk is constructed by maximizing the first-order moments (expectation) and third-order moments (skewness) while minimizing the second-order moments (variance) and fourth-order moments (kurtosis) [27].

$$\begin{cases} \max \mu_p(X) = X' \mu, \\ \min \sigma_p^2(X) = X' H X, \\ \max s_p^3(X) = X' S(X \otimes X), \\ \min k_p^4(X) = X' K(X \otimes X \otimes X), \\ X' I = 1, \\ X_i \geq 0 (i = 1, 2, \dots, N), \end{cases} \quad (4)$$

where $I = (1, 1, \dots, 1)'$ is the $N \times 1$ vector whose elements are all 1.

5.2. *Techniques for Solving the Mean-High-Order Moment Model.* Here, for the mean-high-order moment model for the corresponding solution, the PGP technique proposed by Lai et al. is used, where first multiple-objective problems are converted into single-objective problems, which are first considered separately, and then the individual objectives are combined [28, 29]. Satisfaction levels $\tilde{\mu}_p, \tilde{\sigma}_p^2, \tilde{s}_p^3, \tilde{k}_p^4$ are determined, and the satisfaction level represents the optimal level for a particular single objective without considering other objective conditions:

$$\text{s.t.} \begin{cases} \max \tilde{\mu}_p(X) = X' \mu, \\ X' I = 1, \\ X_i \geq 0, (i = 1, 2, \dots, N), \end{cases} \quad (5)$$

$$\text{s.t.} \begin{cases} \min \tilde{\sigma}_p^2(X) = X' H X, \\ X' I = 1, \\ X_i \geq 0, (i = 1, 2, \dots, N), \end{cases} \quad (6)$$

$$\text{s.t.} \begin{cases} \max \tilde{s}_p^3(X) = X' S(X \otimes X), \\ X' I = 1, \\ X_i \geq 0, (i = 1, 2, \dots, N), \end{cases} \quad (7)$$

$$\text{s.t.} \begin{cases} \min \tilde{k}_p^4(X) = X' K(X \otimes X \otimes X), \\ X' I = 1, \\ X_i \geq 0, (i = 1, 2, \dots, N). \end{cases} \quad (8)$$

Find the minimum and optimal combination weights $X = (X_1, X_2, \dots, X_N)'$:

$$\text{Min } Z = \left| \frac{d_1}{\mu_p} \right|^{\alpha_1} + \left| \frac{d_2}{\sigma_p^2} \right|^{\alpha_2} + \left| \frac{d_3}{s_p^3} \right|^{\alpha_3} + \left| \frac{d_4}{k_p^4} \right|^{\alpha_4}, \quad (9)$$

$$\text{s.t.} \begin{cases} X' \mu + d_1 = \tilde{\mu}_p(X), \\ X' H X - d_2 = \tilde{\sigma}_p^2(X), \\ X' S(X \otimes X) + d_3 = \tilde{s}_p^3(X), \\ X' K(X \otimes X \otimes X) - d_4 = \tilde{k}_p^4(X), \\ X' I = 1, \\ X_i \geq 0, (i = 1, 2, \dots, N), \\ d_j \geq 0, (j = 1, 2, 3, 4). \end{cases} \quad (10)$$

In the previous equation, d_1 represents the deviation between the optimal expectations σ_p^2 and $\tilde{\sigma}_p^2$, d_2 represents the deviation between the optimal variances s_p^3 and \tilde{s}_p^3 , d_3 represents the deviation between the optimal skewness k_p^4 and skewness \tilde{k}_p^4 , and d_4 represents the deviation between the optimal kurtosis k_p^4 and kurtosis \tilde{k}_p^4 . $\alpha_1, \alpha_2, \alpha_3$, and α_4 represent the degree of investor preference for the mean,

TABLE 5: Portfolio risk identification results.

Code	Combination type Preference	The weight of the best portfolio					
		C000768	C000897	C600313	C600416	C600728	C600737
C1	1, 1, 0, 0	0.2387	0.1094	0.1360	0.0141	0.1679	0.3329
C2	1, 1, 0, 1	0.2207	0.1359	0.1301	0.0564	0.1382	0.3178
C3	1, 1, 1, 0	0.2139	0.0940	0.1326	0.0271	0.2016	0.3298
C4	1, 1, 1, 1	0.2146	0.1286	0.1281	0.0597	0.1515	0.3164
C5	1, 1, 1, 2	0.1971	0.1129	0.1321	0.0322	0.1751	0.3496
C6	1, 1, 2, 1	0.2076	0.1078	0.1245	0.0666	0.1792	0.3133
C7	1, 1, 2, 2	0.2057	0.0625	0.1213	0.0483	0.2348	0.3264
C8	Equal weight	0.1667	0.1667	0.1667	0.1667	0.1667	0.1667

variance, skewness, and kurtosis, respectively. The optimal weights $X = (X_1, X_2, \dots, X_N)'$ derived from equation (7) then form the basis for optimal portfolio investment selection under the high-order moment risk condition.

5.3. Model Calculation Results. The two columns in Table 5 indicate that, in the mean-high-order moment risk model, the measure of preference set will consist of four indicator parameters together (A_1, A_2, A_3, A_4) representing the preference for expected return, variance risk, skewness, and kurtosis, respectively. Here, we choose the values to represent the degree of preference: 0 means no preference for some indicators, 1 means a preference for some indicators, and 2 means a special preference for some indicators. In this paper, we coded different preference combinations for each different characteristic. When (A_1, A_2, A_3, A_4) takes the value of (1,1,0,0), it means more consideration of expected return and variance risk; when (A_1, A_2, A_3, A_4) takes the value of (1,1,0,1), it means more consideration of expected return, variance, and kurtosis risk. When (A_1, A_2, A_3, A_4) is (1, 1, 1, 1), the expected return, variance, and kurtosis are considered; when (A_1, A_2, A_3, A_4) is (1,1,1,2), the expected return, variance, and kurtosis are considered and kurtosis risk is considered; when (A_1, A_2, A_3, A_4) is (1, 1, 1, 2), the expected return, variance, and kurtosis are considered; when (A_1, A_2, A_3, A_4) is (1,1,2,1), it means that the expected return, variance, and kurtosis are taken into account and skewness risk is taken into account; when (A_1, A_2, A_3, A_4) is (1,1,2,2), it means that the expected return, variance, skewness, and kurtosis risk are taken into account; finally, the equal-weight portfolio is also calculated for comparative analysis. The weight of each asset reflects the degree of contribution to the portfolio assets of which it is a part, and each asset contributes to the portfolio including risk as well as return. The asset allocations under different risk preferences are calculated according to equations (4)–(10), and the third to eighth columns in Table 5 show the optimal weight allocations for the portfolio under different investment preferences. This indicates that the asset contributes more to the portfolio return and risk.

The results of the corresponding portfolio's return and risk are also calculated according to equations (4)–(10), and in Table 6, columns three to six are the optimal set of mean returns and risks taken by different investment

TABLE 6: Optimal return of the portfolio over the risk set.

Code	Combination type Preference	Best return and risk			
		Mean	Variance	Skewness	Kurtosis
C1	1, 1, 0, 0	0.0231	2.0512	-1.6999	24.9089
C2	1, 1, 0, 1	0.0145	1.9768	-1.7154	23.5019
C3	1, 1, 1, 0	0.0252	2.0843	-1.6058	25.2483
C4	1, 1, 1, 1	0.0156	1.9870	-1.6896	23.6106
C5	1, 1, 1, 2	0.0216	2.0450	-1.6488	24.4657
C6	1, 1, 2, 1	0.0183	2.0203	-1.6266	24.0459
C7	1, 1, 2, 2	0.0275	2.1462	-1.4748	26.1060
C8	Equal weight	0.0020	2.0976	-1.6049	25.4302

preferences for the asset allocation. The results are consistent with the basic rule of financial markets: the higher the risk, the higher the return. In this paper, the risk is measured by the risk set of the combination of variance risk, skewness risk, and kurtosis risk. The results for the portfolio with preferences (C1-C7 portfolio) also show that the higher the absolute value of the risk set parameter, the higher the return. In the mean-higher-order moment risk model, the maximum expected return is 0.0231 for the C1 portfolio, which only considers variance risk and requires the distribution of wealth according to its corresponding weight ratio, and the higher-order moment risk is high despite the diversification, especially the kurtosis risk, which reaches 24.9089. The maximum expected return varies depending on the type of risk considered. The maximum expected return of 0.0145 decreases when only kurtosis risk is considered, but the maximum expected return of 0.0252 decreases when only skewness risk is considered; from the C1 and C4 portfolios, the maximum expected return of 0.0156 decreases when the investor focuses on the higher-order moment risk; from the C4, C5, and C6 portfolios, more attention to a certain risk will make the maximum return different degrees of increase, but compared to C7 portfolio for comparison, it is found that the maximum expected return of C7 portfolio has a large increase. Compared to other portfolios, this portfolio has the highest risk but lowest return. The above results found that investors must consider the risk factor when obtaining better returns; when investors switch from focusing only on returns to considering risk, the maximum expected return obtained will vary, but the maximum expected return obtained when considering risk more is constantly increasing.

6. Conclusions and Related Recommendations

6.1. Conclusion. In this paper, we first combine the random forest algorithm with the nine-factor model for stock selection based on the value-growth investment strategy, the stocks with more votes are counted as the final result of stock selection, and the results of the stock selection are tested to find that the method is equally valid for use in this paper.

In terms of risk measurement, a risk set consisting of variance, skewness, and kurtosis is chosen to measure risk. It is found that the risk set of variance, skewness, and kurtosis is a more reasonable measure of the characteristics of financial time series than the traditional variance measure, and the risk set is a more effective measure of risk characteristics.

In terms of risk identification, this paper finds that different investors' preferences will lead to corresponding changes in the portfolio weights, which will eventually lead to changes in the optimal return and risk set of the portfolio. The higher-order moment risk model has a significant effect on measuring risk and can also well diversify the higher-order moment risk. Unlike only the variance risk measure of risk, the risk set consisting of the third-order moment skewness risk and the fourth-order moment kurtosis risk is more suitable for analyzing the generalized financial time series risk. The above empirical results also show that the larger the absolute value of the coefficients of the risk set, the higher the overall risk and the greater the risk obtained; the higher-order moment model subdivides the risk, thus achieving a more accurate risk diversification.

6.2. Related Policy Recommendations

6.2.1. Improve the Market Structure and Prevent Financial Risks. China's stock market started late and developed for a short period and is still defective in many aspects [30–31]. The Science and Technology Innovation Board, New Third Board Selective Layer, GEM Registration System Reform, and SSE Index Reform launched in recent times are undoubtedly important initiatives to improve the market structure. Regulators and policymakers should continue to promote market structure reform, strive to improve trading mechanisms, encourage institutional investors to enter the market, vigorously develop financial derivatives, and regulate market trading practices on this basis. At the same time, relevant departments should also reduce administrative intervention in the market, so that the market rises and falls according to its own value development, to create a good market environment. At present, China implements the limit of up and down and T+1 trading system, but there are some problems. For example, the shorting mechanism is not perfect, there are many small- and medium-sized investors in the stock market and few institutional investors. Pension funds, enterprise annuities, public welfare funds, and other funds into the market restrictions have only been slowly relaxed in recent years. These circumstances are likely to lead to investors blindly following the wind operation, chasing up and down, the intrinsic value of the stock not able to be

reflected, and then producing herding effect, intensifying the risk of the stock market. The current financial derivatives such as stock index futures and financing and financing instruments, which are used to stabilize the stock market and prevent risks, have fewer types and higher thresholds and are rarely accessed by ordinary investors, and there is still room for further improvement of financial derivatives.

6.2.2. Improve the Information Disclosure System and Strengthen the Supervision of Listed Companies. The stock market suffers from the problem of untimely and inaccurate information disclosure, and false public opinion is easily directed to mislead investors. small- and medium-sized investors do not have access to correct information, while some institutional investors have the news in advance. In this situation of asymmetric and incomplete information, small and medium investors are prone to herd mentality and irrational behavior such as chasing up and killing down and herd effect. Therefore, the regulator should strive to improve the information disclosure mechanism, ensure the timeliness and accuracy of information disclosure, and put forward clear requirements on the time and manner of information disclosure to enhance the transparency of stock market information and improve market efficiency. At the same time, the supervision and management of the listed companies should be strengthened, penalties should be increased, and the punishment mechanism should be improved to prevent irregular disclosures, false propaganda, malicious speculation of stock prices, and insider trading, to effectively protect the interests of small- and medium-sized investors.

6.2.3. Strengthen Investor Education and Improve Investor Quality. The vast majority of participants in China's stock market are small- and medium-sized investors, and small- and medium-sized investors themselves often lack the corresponding theoretical knowledge and are prone to gambler's mentality and frequent operations when investing, which will intensify the herding effect in the market and is not conducive to the healthy and stable development of the market. Therefore, small- and medium-sized investors who have just entered the market should carry out certain investment education, so that they master the necessary basic knowledge of investment. At the same time, through television, the network, and other media tools to encourage the majority of investors to learn the relevant investment knowledge, establish the correct investment concept and guide investors to rational investment, to improve the overall quality of investors. As for investors themselves, they should also continue to learn professional knowledge, establish the concept of value investment, adhere to rational investment, keep a clear head, and not blindly follow the wind investment.

Data Availability

The data used to support the findings of this study are included within the article.

Disclosure

Li-Jun Liu and Wei-Kang Shen are co-first authors.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Authors' Contributions

Li-Jun Liu and Wei-Kang Shen contributed equally to this paper,

Acknowledgments

This study was supported by the National Social Science Foundation of China under the key project "Research on Policy Tool Selection and Methodological Innovation for Stabilizing Growth and Adjusting Structure" (15AZD006).

References

- [1] A. Fan and M. Palaniswami, "Stock selection using support vector machines," in *Proceedings of the International Joint Conference on Neural Networks (IJCNN'01) (Cat. No. 01CH37222)*, pp. 1793–19798, IEEE, Washington, DC, USA, July 2001.
- [2] K.-j. Kim, "Financial time series forecasting using support vector machines," *Neurocomputing*, vol. 55, no. 1-2, pp. 307–319, 2003.
- [3] L. Yu, S. Wang, and K. K. Lai, "Mining stock market tendency using GA-based support vector machines," vol. 3828, pp. 336–345, in *Proceedings of the First International Workshop on Internet and Network Economics (LNCS)*, vol. 3828, Springer, Hong Kong, China, December 2005.
- [4] B. Li, P. Zhao, S. C. H. Hoi, and V. Gopalkrishnan, "PAMR: passive aggressive mean reversion strategy for portfolio selection," *Machine Learning*, vol. 87, no. 2, pp. 221–258, 2012.
- [5] B. Li, S. C. H. Hoi, P. Zhao, and V. Gopalkrishnan, "Confidence weighted mean reversion strategy for online portfolio selection," *ACM Transactions on Knowledge Discovery from Data*, vol. 7, no. 1, pp. 1–38, 2013.
- [6] P. Ladyzynski, K. Zbikowski, and P. Grzegorzewski, "Stock trading with random forests, trend detection tests and force index volume indicators," *Artificial Intelligence and Soft Computing*, vol. 7895, pp. 441–452, 2013.
- [7] J. B. Heaton, N. G. Witte, and J. H. Witte, "Deep learning for finance: deep portfolios," *Applied Stochastic Models in Business and Industry*, vol. 33, no. 1, pp. 3–12, 2016.
- [8] W. F. Sharpe, "Capital asset prices: a theory of market equilibrium under conditions of risk," *The Journal of Finance*, vol. 19, no. 3, pp. 425–442, 1964.
- [9] Z. J. Chen and Z. X. Ye, "Optimal portfolio model with higher order moment constraints and approximate linear programming angle," *Journal of Engineering Mathematics*, vol. 25, no. 6, pp. 1005–1012, 2008.
- [10] Z. J. Chen, *Optimal Portfolio Model and Algorithm with Higher Order Moment Constraints*, Shanghai Jiao-tong University, Shanghai, China, 2007.
- [11] Z. H. Shuai, *Robust Portfolio Selection Based on S-type Utility Function and Higher Moment*, Xiamen University, Xiamen, China, 2009.
- [12] Z. P. Tang and Y. Tang, "Analysis of the theoretical framework of high order moment volatility modeling and its application (Group 4 Financial Bank)," 2016.
- [13] S. Z. Peng, *Research on Portfolio Optimization Based on Higher Order Moments*, Harbin Institute of Technology, Harbin, China, 2012.
- [14] Y. P. Wang, Z. P. Chen, and Y. N. Chen, "Research on multi factor portfolio selection model," *Journal of Engineering Mathematics*, vol. 29, no. 2, pp. 807–814, 2012.
- [15] T.-M. Níguez, I. Paya, and D. Peel, "Pure higher-order effects in the portfolio choice model," *Finance Research Letters*, vol. 19, pp. 255–260, 2016.
- [16] W. Yue and Y. P. Wang, "A new fuzzy multi-objective higher order moment portfolio selection model for diversified portfolios," *Physica A: Statistical Mechanics and Its Applications*, vol. 465, pp. 124–140, 2017.
- [17] M. Aksarayl 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.
- [18] B. L. Chen, J. D. Zhong, and Y. Y. Chen, "A hybrid approach for portfolio selection with higher-order moments: empirical evidence from Shanghai Stock Exchange," *Expert Systems with Applications*, vol. 145, pp. 1–11, 2020.
- [19] K. C. Lv, H. F. Yan, and C. Chen, "Research on quantitative investment strategy based on CSI 300 component stocks," *Journal of Guangxi Normal University (Natural Science Edition)*, vol. 37, no. 1, pp. 1–12, 2019.
- [20] S. Y. Wang, Z. F. Cao, and M. Z. Chen, "Application of random forest in quantitative stock selection," *Operation Research and Management*, vol. 25, no. 3, pp. 163–168, 2016.
- [21] Z. F. Cao, H. Ji, and B. C. Xie, "Using random forest algorithm to select high quality stocks," *Journal of Capital University of Economics and Trade*, vol. 16, no. 2, pp. 21–27, 2014.
- [22] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 5, pp. 5–32, 2001.
- [23] L. Breiman, "Bagging predictors," *Machine Learning*, vol. 24, no. 2, pp. 123–140, 1996.
- [24] D. G. Mcpeek, "Growth & value investment strategy," *Wall Street Transcript*, vol. 177, no. 7, p. 1, 2007.
- [25] W. Saleh, "Value-growth investment strategy: evidence based on the residual income valuation model," *International Journal of Economics and Finance*, vol. 3, no. 4, pp. 233–244, 2011.
- [26] M. A. V. Machado, "Value or growth strategy? Empirical evidence in Brazil," *Revista Brasileira De Gestao De Negocios*, vol. 15, no. 46, 2013.
- [27] C. X. Jiang and S. Y. Zhang, "Comparison between M-V-S-K analysis and utility function analysis of higher moment portfolio investment," *Statistics and Decision*, vol. 15, pp. 12–15, 2008.
- [28] K. K. Lai, L. Yu, and S. Wang, "Mean-variance-skewness-kurtosis-kurtosis-based portfolio optimization," in *Proceedings of the First International Multi-Symposiums on Computer and Computational Sciences, 2006 (IMSXXS'06)*, June 2006.
- [29] R. Jagannathan and T. Ma, "Risk reduction in large portfolios: why imposing the wrong constraints helps," *The Journal of Finance*, vol. 58, no. 4, pp. 1651–1683, 2003.
- [30] J. M. Zhu, Y. Chen, and S. Zhang, "Analysis of the impact of climate change on national vulnerability based on fuzzy comprehensive evaluation," *Discrete Dynamics in Nature and Society*, vol. 2020, Article ID 3527540, 10 pages, 2020.

- [31] J.-B. Liu, J. Zhao, and J. Min, "On the Hosoya index of graphs formed by a fractal graph," *Fractals-Complex Geometry Patterns and Scaling*, vol. 27, no. 8, 2019.
- [32] J.-M. Zhu, L. Wang, and J.-B. Liu, "Eradication of Ebola based on dynamic programming," *Computational and Mathematical Methods in Medicine*, vol. 2016, Article ID 1580917, 9 pages, 2016.