A Multiperiod Multiobjective Portfolio Selection Model With Fuzzy Random Returns for Large Scale Securities Data

Chen Li, Yulei Wu, *Senior Member, IEEE,* Zhonghua Lu, Jue Wang, and Yonghong Hu

***Abstract*—It is agreed that portfolio selection models are of great importance for the financial market. In this article, a constrained multiperiod multiobjective portfolio model is established. This model introduces several constraints to reflect the trading restrictions and quantifies future security returns by fuzzy random variables to capture fuzzy and random uncertainties in the financial market. Meanwhile, it considers terminal wealth, conditional value at risk (CVaR), and skewness as tricriteria for decision making. Obviously, the proposed model is computationally challenging. This situation gets worse when investors are interested in a larger financial market since the data they need to analyze may constitute typical big data. Whereafter, a novel intelligent hybrid algorithm is devised to solve the presented model. In this algorithm, the uncertain objectives of the model are approximated by a simulated annealing resilient back propagation (SARPROP) neural network which is trained on the data provided by fuzzy random simulation. An improved imperialist competitive algorithm, named IFMOICA, is designed to search the solution space. The intelligent hybrid algorithm is compared with the one obtained by combining NSGA-II, SARPROP neural network, and fuzzy random simulation. The results demonstrate that the proposed algorithm significantly outperforms the compared one not only in the running time but also in the quality of obtained Pareto frontier. To improve the computational efficiency and handle the large scale securities data, the algorithm is parallelized using MPI. The conducted experiments illustrate that the parallel algorithm is scalable and can solve the model with the size of securities more than 400 in an acceptable time.**

***Index Terms*—Fuzzy random simulation, imperialist competitive algorithm (ICA), multiperiod multiobjective portfolio selection, parallel computing.**

Manuscript received December 19, 2019; revised February 28, 2020, April 9, 2020, and April 28, 2020; accepted May 1, 2020. This work was supported in part by the National Natural Science Foundation of China under Grant 61873254 and in part by the Strategic Priority Research Program of Chinese Academy of Sciences under Grant XDC01040100. *(Corresponding authors: Yulei Wu; Zhonghua Lu.)*

Chen Li is with the Computer Network Information Center, Chinese Academy of Sciences, Beijing 100190, China, and also with the University of Chinese Academy of Sciences, Beijing 100049, China (e-mail: lichen@sccas.cn).

Yulei Wu is with the College of Engineering, Mathematics and Physical Sciences, University of Exeter, EX4 4QF Exeter, U.K. (e-mail: y.l.wu@exeter.ac.uk).

Zhonghua Lu and Jue Wang are with the Computer Network Information Center, Chinese Academy of Sciences, Beijing 100190, China (e-mail: zhlu@cnic.cn; wangjue@sccas.cn).

Yonghong Hu is with the School of Statistics and Mathematics, Central University of Finance and Economics, Beijing 100081, China (e-mail: huyonghong@cufe.edu.cn).

Color versions of one or more of the figures in this article are available online at https://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TFUZZ.2020.2992866

I. Introduction

PORTFOLIO selection has become an increasingly popular topic during the past several decades. It aims at achieving an optimal allocation of wealth among available securities. Modern portfolio selection theory originates from Markowitz's pioneering work [1] in 1952, in which Markowitz proposed his most famous portfolio model, the so-called mean-variance (MV) model. However, the assumption that investment returns are jointly elliptically (or spherically) distributed has made the MV model widely criticized by practitioners since the asymmetric investment return is more ordinary in real-world scenarios [2]. Many scholars [3], [4] including Markowitz, thus have dedicated themselves to the creative task of improving the MV model.

It is worth mentioning that most of the existing portfolio selection models are on the basis of the cognition that securities are offering random returns which can be extracted from historical data. However, even if the difficulty in accessing accurate historical data is ignored, there still exist some nonrandom factors in the financial market. For example, expert opinions and company performance are widely considered to be highly relevant to future security returns. Inspired by the concept of fuzzy sets introduced by Zadeh [5] in 1965, fuzzy portfolio selection models, in which future security returns are treated as fuzzy variables, boomed out and started to replace stochastic models in portfolio selection problems.

In fact, for most instances in the financial market, randomness and fuzziness exist simultaneously. In such situations, the portfolio models that mirror these two fundamental uncertainties seem to be more competitive. Gradually recognizing the fact mentioned above, feature extraction and representation of future security returns have received tremendous attention in both academia and industry. Some of them argued that it would be more reasonable to treat future security returns as random variables whose parameters are considered as fuzzy numbers, namely the so-called random fuzzy variables. The random fuzzy variable was first defined by Liu [6]. This concept was then developed by Hasuike *et al.* [7] and Katagiri *et al.* [8] to construct random fuzzy portfolio selection models. However, future security returns may be fuzzy variables with the key value determined by random variables, fuzzy random variables [9]. To the best of authors' knowledge, little attention has been placed in this area.

From a practical perspective, there exist many trading restrictions in the real-world financial market. As a result, some researchers [10], [11] began to take realistic constraints into consideration when building their portfolio models. However, it is noticed that most of the above studies are limited to single-period single-objective portfolio selection problem. In fact, investors are driven to a long-term investment and premeditate more criteria, to make their decisions better adapt the complex and volatile market environment. Any decision ignoring this fact may be unrealistic. To avoid falling into the above discussed traps, researchers have made several attempts. However, this puzzle has not been addressed, since most of these attempts are one-sided or rely on the model conversion technology.

There is almost no study on the long-term investment problems with realistic constraints and multiple decision criteria under uncertain environment in which randomness and fuzziness appear simultaneously. This article proposes a constrained multiperiod multiobjective portfolio model with fuzzy random returns. In this model, terminal wealth, conditional value at risk (CVaR) and skewness are served as decision criteria. Realistic constraints such as the cardinality constraint, bounding constraint and round-lot constraint are supplemented. Meanwhile, to eliminate potential possibility of turning the model into a simple accumulation of multiple single period models, an assumption associated with the investment horizons is also added. These additional extensions can make the proposed model more comprehensive and suitable for various investment scenarios. It is obvious that the model of this type can be computationally challenging and cannot be solved by traditional methods. This motivates the prosperity of studying alternative algorithms. Evolutionary algorithm (EA) has been found to be an appropriate choice. The studies of [12]–[14] are some overtures that adopt this alternative tool to settle portfolio selection problems suffering a lot for calculation difficulties. Although EAs cannot guarantee global optimal solutions, they maintain a superiority since they can successfully find high-quality near optimal solutions using less computational time. In this article, a novel intelligent hybrid algorithm is put forward for the sake of effectively resolving the proposed model.

Over the past 20 years, data has increased to a large scale in various fields [15]. The term of big data is then proposed. For the investors who are interested in a larger financial market, the related data that they must analyze constitute typical big data. In this situation, the models consisting of only a few hundred securities may take several months to solve, which cannot meet the transaction demands. Thus, it is urgent to handle the large scale securities data in portfolio selection problems. It is agreed that high-performance computing (HPC) is playing a significant role in big data processing. Inspired by this, message passing interface (MPI) technology is adopted to parallelize the newly developed intelligent hybrid algorithm so as to improve the computational efficiency and solve the proposed model subject to large scale data in an acceptable time.

The rest of this article is organized as follows. In Section II, related works are reviewed. Section III presents a brief introduction to the notions and properties of fuzzy variables and fuzzy random variables. In Section IV, the expression for available wealth at each period is described first, followed by a constrained multiperiod multiobjective portfolio model with fuzzy random returns. Section V is dedicated to presenting the intelligent hybrid algorithm which is used to solve the aforementioned model. In Section VI, the intelligent parallel hybrid algorithm adopting MPI technology is introduced to improve computational efficiency and cope with large scale securities data. Section VII demonstrates the superiority of intelligent hybrid algorithm and the scalability of its parallelization. Finally, Section VIII concludes this article.

II. Related Work

In 1952, Markowitz [1] first introduced the MV portfolio model. In this model, Markowitz described the return of portfolio by the expectation of returns and the risk by the variance. However, using variance as a risk means that the model treats desirable high investment returns and investment returns below the expected value equally, which is not realistic. Targeting at this challenge, variants of the MV model boomed. For example, Markowitz [16] used semivariance as another risk measure and introduced the mean-semivariance model. Konno and Yamazaki [17] proposed the mean-absolute-deviation model to get rid of most of the shortcomings in the MV model. Artzner *et al.* [3] creatively qualified the risk of portfolio as value at risk (VaR) and proposed the mean-VaR portfolio model. However, this model was then proven to suffer from multiple extremum problem when security returns are asymmetrically distributed. CVaR presented by Rockafellar and Uryasev [4] in 2000 addressed this problem to some extent. Based on CVaR, a new risk measure named conditional drawdown-at-risk [18] was defined and used to settle the real life portfolio selection problems. However, in the models mentioned previously, only randomness in the financial market is reflected. In these models, security returns are treated as random variables and can be derived from historical data.

In fact, the uncertainty in the financial market originates from the interaction of investors, in which fuzziness is an important characterization. In 1965, Zadeh [5] proposed the concept of fuzzy sets, which then became the basis of fuzzy techniques. With the development of fuzzy techniques, researchers began to describe the uncertainty of the financial market from a fuzzy point of view. Tanaka and Guo [19] reflected the uncertainty in the financial market by an interval given by the spreads of the portfolio returns and proposed a possibility portfolio model. Then, Huang [20] proposed two types of fuzzy chance-constrained model, and then solved them by a newly introduced fuzzy simulation-based algorithm. Gupta *et al.* [21] established comprehensive portfolio models by fuzzy mathematical programming. Hajnoori *et al.* [22] employed grey fuzzy technique to forecast stock prices and use them to form a novel constrained portfolio model. Recently, Li *et al.* [23] built a fuzzy mean-CVaR model by considering the future security returns as triangle fuzzy variables and proposed a parallel algorithm based on the fuzzy simulation for model solving. Pai [24] constructed a constrained multiobjective fractional programing model, and dealt with the uncertainty in the financial market by a strategically refined Monte Carlo simulation. Then, fuzzy decision theory-based metaheuristics are applied to obtain final optimal portfolio.

Obviously, these models all treated randomness and fuzziness separately. It makes more sense to consider twofold uncertainties when constructing portfolio models. Katagiri and Ishii [25] are the first ones who took future security returns as fuzzy random variables. Following this idea, Liu [12] constructed a general framework of fuzzy random chance-constrained programming and developed a fuzzy random simulation for uncertain objectives approximation. Li and Xu [26] proposed a novel fuzzy random λ-mean variance model in a hybrid uncertain environment by adding expert knowledge. Yoshida [27] discussed a VaR portfolio model under fuzzy random uncertainties. Recently, Sun *et al.* [28] proposed a fuzzy random portfolio selection model by taking investors' sentiments into account and discussed its efficient frontiers.

Apart from the uncertainties in the environment, real-word constraints and decision criteria are also the factors that should be paid attention to. The studies that explore on this area have been very active. Lwin *et al.* [29] extended the MV model by including return and risk as bicriteria and adding the cardinality, quantity, preassignment and round-lot constraints. Babaei *et al.* [30], Pouya *et al.* [31], Khanjarpanah and Pishvaee [32], and Kaucic [33] are recent studies that researched on constrained multiobjective portfolio models.

However, adding realistic constraints makes solving portfolio models an NP-hard problem, which cannot be addressed by exact algorithms in polynomial time. EA is a good candidate for solving the problem. Numerous studies have been carried out with the aim of applying EAs to solve constrained multi-objective portfolio optimization problem. Loraschi *et al.* [34] tested their coarse-grained distributed genetic algorithm (GA) on standard test functions and then applied it to a two-objective real-world portfolio model. In [35], an improved EA was used to solve the multiobjective model with additional real-world constraints such as cardinality and buy-in thresholds constraints. In [13], a constrained triobjective model was proposed by adopting three criteria, namely return, risk and liquidity, and cardinality, round-lots constraints. A compromise approach-based GA was then developed for model solving. Recently, Chen *et al.* [36] used a novel hybrid EA to find Pareto-optimal solutions for their multiperiod mean-variance-skewness model. Wang *et al.* [37] attempted to use a fuzzy simulation-based multiobjective particle swarm optimization (PSO) algorithm to solve their newly proposed fuzzy multiobjective model.

Solving large scale portfolio selection models that are composed of more than a few hundred securities is still a challenging issue, since their related data constitute typical big data. Targeting at this conundrum, HPC which is a widely popular method to handle data processing in various applications has been applied by some researchers. For example, Zou and Zhang [38] adopted diverse HPC techniques to handle portfolio allocation problem based on intraday high-frequency data. Li *et al.* [23] proposed a parallel hybrid intelligent algorithm by using MPI technology to solve fuzzy mean-CVaR model with 180 securities. The result demonstrated that the running time of solving the aforementioned model is within a reasonable level. Sun and Lu [39] designed a big data platform using HPC technology and used it to obtain investment strategy based on large scale financial data.

To sum up, portfolio selection has become one of the active areas in finance. Among the research effort, there are three main challenges.

1) How to reflect the twofold uncertainties in the financial market when building a portfolio model?
2) How to make the portfolios derived from models closer to the real investment scenarios?
3) How to solve the proposed model in an acceptable time even facing large scale securities data?

To cope with the above three challenges, a constrained multiperiod multiobjective portfolio model with fuzzy random returns is proposed first by considering investors' attitudes and real-world constraints. In this model, future security returns are treated as fuzzy random variables. Meanwhile, terminal wealth, CVaR, and skewness are used as tricriteria. Moreover, the cardinality, bounding, round-lot, and additional constraints are taken into consideration. Then, a novel intelligent hybrid algorithm is developed for the sake of effectively resolving the proposed model. To improve computational efficiency and cope with large scale securities data, an intelligent parallel hybrid algorithm adopting MPI technology is presented.

This article is an extended version of the previous conference paper [40]. The main new contributions are summarized as follows.

- Literatures that technically relate to the proposed work are systematically reviewed, including fuzzy random portfolio selection models, multiobjective portfolio optimization problems and relevant algorithms and HPC technology for big data analysis.
- To facilitate the understanding of the proposed scheme, the background knowledge about fuzzy variables is added, and the properties of fuzzy random variables are complemented as well.
- The introduction of the proposed model is enriched. Instead of giving the expression of available wealth in each period directly, the process of its calculation is described. Moreover, realistic constraints are explained in detail in the proposed model.
- Fuzzy random simulation and simulated annealing (SA) resilient back propagation (SARPROP) neural network are supplemented to illustrate the procedure of approximating uncertainty functions in the proposed model. The original imperialist competitive algorithm (ICA) is also presented for better comprehension of the improvement in IFMOICA. Furthermore, flowcharts of the IFMOICA and the novel intelligent hybrid algorithm are provided.
- The proposed algorithm is implemented and compared on two more datasets. For better comparison, spacing metric is added as another performance metric. In addition, the obtained Pareto frontiers of the two algorithms on the three datasets are given. The running time, speedup and parallel efficiency on the three datasets are illustrated to verify the scalability of the parallelization of the intelligent hybrid algorithm and its superiority when facing the large scale securities data.

III. Preliminaries

In this section, the relevant properties of fuzzy variables are reviewed, followed by the definition, expected value, and skewness of fuzzy random variables.

Definition 1 (Credibility space [41]): Assume that Θ is a nonempty set. For each $A \in 2^{\Theta}$, there is a nonnegative number $\mathrm{Cr}\{A\}$ that satisfies four axioms, namely normality, monotonicity, self-duality, and maximality. Then, the credibility space can be denoted by a triplet $(\Theta, 2^{\Theta}, \mathrm{Cr})$, where the function Cr is called credibility measure.

Definition 2 (Fuzzy variable [41]): Let $\mathcal{R}$ be the set of real numbers. A fuzzy variable defined on credibility space $(\Theta, 2^{\Theta}, \mathrm{Cr})$ is a mapping $\xi : \Theta \to \mathcal{R}$.

Then, its membership function can be defined by

$$\mu(x) = (2\mathrm{Cr}\{\xi = x\}) \wedge 1,\ x \in \mathcal{R}. \tag{1}$$

Theorem 1 [41]: Suppose that μ is the membership function of fuzzy variable ξ. Then, for any Borel subset A of $\mathcal{R}$

$$\mathrm{Cr}\{\xi \in A\} = \frac{1}{2}(\sup_{x\in A}\mu(x) + 1 - \sup_{x\in A^c}\mu(x)). \tag{2}$$

The independence of fuzzy variables has been discussed by many scholars from different perspectives. In this article, the definition given and proved by [42] is adopted.

Definition 3 (The independence of fuzzy variables [42]): $\xi_1, \xi_2, ..., \xi_n$ are considered to be independent, if for any sets $A_1, A_2, ..., A_n \in \mathcal{R}$

$$\mathrm{Cr}\{\cap_{i=1}^{n}\{\xi_i \in A_i\}\} = \min_{0\le i\le n}\mathrm{Cr}\{\xi_i \in A_i\}. \tag{3}$$

Theorem 2 [41]: Presume that fuzzy variables $\xi_1, \xi_2, \xi_3, \ldots, \xi_n$ are independent and $\mu_1, \mu_2, \mu_3, \ldots, \mu_n$ constitute their membership functions, respectively, such that $\mathrm{Cr}\{f(\mu_1, \mu_2, \mu_3, \ldots, \mu_n) \in A\}$ can be described as

$$\begin{aligned}\frac{1}{2}(&\sup_{f(x_1,x_2,\ldots,x_n)\in A}\ \min_{0\le i\le n}\mu_i(x_i) + 1\\ &- \sup_{f(x_1,x_2,\ldots,x_n)\in A^c}\ \min_{0\le i\le n}\mu_i(x_i)).\end{aligned} \tag{4}$$

for any set A of $\mathcal{R}^m$.

First introduced by Kwakernaak [9], fuzzy random variables can be used to describe an uncertain environment that fuzziness and randomness exist at the same time. Since then, fuzzy random variables have demonstrated their magic charm in many fields, such as optimization, model scheduling, engineering, etc., [13], [43].

Definition 4 (Fuzzy random variable [12]): Suppose that $\mathcal{F}$ defines a set of fuzzy variables. A fuzzy random variable defined on probability space $(\Omega, \Sigma, \mathrm{Pr})$ is a measurable function $\zeta : \Omega \to \mathcal{F}$. Then, for any Borel subset B of the real line $\mathcal{R}$

$$\zeta^*(B)(\omega) = \sup_{x\in B}\mu_{\zeta(\omega)}(x) \tag{5}$$

where $\zeta(\omega)$ refers to a fuzzy variable whose membership function is described as $\mu_{\zeta(\omega)}$, $\omega \in \Omega$.

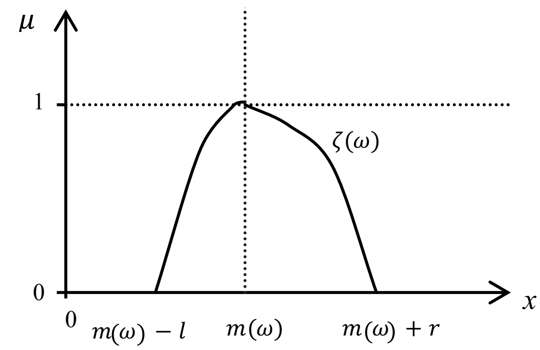


Fig. 1. *L-R* fuzzy random variable $\zeta(\omega)$.

Example 1: A fuzzy random variable ζ defined on probability space $(\Omega, \Sigma, \mathrm{Pr})$ is considered as a *L-R* fuzzy random variable if, for $\forall\omega \in \Omega$

$$\mu_{\zeta(\omega)}(x) = \begin{cases} L(\frac{m(\omega)-x}{l}), & m(\omega) - l \le x \le m(\omega)\\ R(\frac{x-m(\omega)}{r}), & m(\omega) < x \le m(\omega) + r\\ 0, & \text{otherwise}\end{cases} \tag{6}$$

where random variable $m(\omega)$ is called the peak or mean value of ζ, and positive real numbers l and r describe the left and right spreads, respectively. $L, R : [0, +\infty] \to [0, 1]$, $L(0) = R(0) = 1$ and $L(1) = R(1) = 0$. L and R are strictly monotonic decreasing functions. Symbolically, ζ is denoted with $(m(\omega), l, r)_{LR}, \forall\omega \in \Omega$ (see Fig. 1). Noted that triangular fuzzy random variables are a special case of *L-R* fuzzy random variables.

Definition 5 (Expected value of fuzzy random variable [44]): Suppose that $(\Omega, \Sigma, \mathrm{Pr})$ is a probability space. Then, the expected value of fuzzy random variable ζ can be given by

$$\begin{aligned}\mathrm{E}(\zeta) = \int_{\Omega}[&\int_0^{\infty}\mathrm{Cr}\{\zeta(\omega) \ge r\}\mathrm{d}r -\\ &\int_{-\infty}^{0}\mathrm{Cr}\{\zeta(\omega) \le r\}\mathrm{d}r]\mathrm{Pr}(d\omega).\end{aligned} \tag{7}$$

Theorem 3 [44]: Presume that ζ_1 and ζ_2 are fuzzy random variables with finite expected values, such that for any $a, b \in \mathcal{R}$

$$\mathrm{E}[a\zeta_1 + b\zeta_2] = a\mathrm{E}[\zeta_1] + b\mathrm{E}[\zeta_2]. \tag{8}$$

Definition 6 (Skewness of fuzzy random variable [36]): Let e be the finite expected value of fuzzy random variable ζ, such that its skewness, $S[\zeta]$ can be expressed as

$$\mathrm{S}[\zeta] = \mathrm{E}[(\zeta - e)^3]. \tag{9}$$

IV. Constrained Multiperiod Multiobjective Portfolio Model With Fuzzy Random Returns

As it is known that investment horizons do affect the composition of portfolio returns, in this section, the issue that how the composition of portfolio returns changes with the investment period is discussed first, followed by deriving an expression of the available wealth at the beginning of each period. Then, a constrained multiperiod multiobjective portfolio model with fuzzy random returns is established. In

TABLE I
KEY NOTATIONS USED IN THE MODEL DERIVATION

Notations	Descriptions
S_j	The collection of securities with investment horizon j, $j = 1, 2, \dots, \mathcal{H}$
W_t	The available wealth at the beginning of period t, $t = 1, 2, \dots, T + 1$
$\zeta_{t,i}$	The fuzzy random return of security i at period t which contains the invested capital apart from net profit, $t = 1, 2, \dots, T$, $i = 1, 2, \dots, n$
$x_{t,i}$	The proportion of wealth invested in security i at period t, $t = 1, 2, \dots, T$, $i = 1, 2, \dots, n$
$\boldsymbol{x_t}$	The portfolio at period t, $t = 1, 2, \dots, T$
$y_{t,i}$	A binary variable, $y_{t,i} = 1$, if wealth is assigned to security i at period t, and otherwise $y_{t,i} = 0$, $t = 1, 2, \dots, T$, $i = 1, 2, \dots, n$
$d_{t,i}$	The unit transaction cost on security i at period t, $t = 1, 2, \dots, T$, $i = 1, 2, \dots, n$
D_t	The transaction cost aroused by constructing portfolio at period t, $t = 1, 2, \dots, T$
Max_t	The maximum number of securities that constitute a portfolio at period t, $1 \le Max_t \le n$, $t = 1, 2, \dots, T$
$l_{t,i}$	The lower bound of investment proportion on security i at period t, $t = 1, 2, \dots, T$, $i = 1, 2, \dots, n$
$u_{t,i}$	The upper bound of investment proportion on security i at period t, $t = 1, 2, \dots, T$, $i = 1, 2, \dots, n$
$pc_{t,i}$	The price of security i in the secondary market at period t, $t = 1, 2, \dots, T$, $i = 1, 2, \dots, n$
$TN_{t,i}$	The total number of shares that shareholder holds on security i at period t, $t = 1, 2, \dots, T$, $i = 1, 2, \dots, n$
$N^0_{t,i}$	The current position ratio that the investment institution possesses on security i at period t, $t = 1, 2, \dots, T$, $i = 1, 2, \dots, n$
$e_{t,i}$	The minimum transaction volume for security i at period t, $t = 1, 2, \dots, T$, $i = 1, 2, \dots, n$

this model, three decision criteria, namely the terminal wealth, CVaR, and skewness, are considered. In the meantime, real-world constraints are introduced.

It is assumed that investors with initial wealth W_1 will assign their wealth to n securities in the following T periods and acquire their terminal wealth at the end of the period T. During the whole investment, they can readjust their portfolio at the beginning of each period. For better understanding, all notations used subsequently are introduced in Table I.

A. Available Wealth in Each Period

On the basis of descriptions mentioned above, future security returns are characterized by triangular fuzzy random variables, denoted by $(m_{t,i}(\omega) - l_{t,i}, m_{t,i}(\omega), m_{t,i}(\omega) + r_{t,i})$. For each ω, $m_{t,i}(\omega)$ refers to a normal distributed random variable. $l_{t,i}$ and $r_{t,i}$ are all positive constants, $t \in \{1, 2, ..., T\}$, $i \in \{1, 2, ..., n\}$.

The methods used in current research place severe limitations on security return forecasting. For example, experts' estimations have been criticized for being too subjective. The time series analysis overly relies upon historical data. Benefiting from the property that the future is independent of the past given the present, Markov process can be more flexible to the unstable financial market, especially when there exist multiple uncertainties. Therefore, the Markov process is used to predict triangular fuzzy random returns to make the model more robust. Moreover, how to determine the left and right spreads for triangular fuzzy random returns is the key in this process. First, the transition matrix for each security is constructed based on the historical price data, followed by forecasting the future return of the highest price a^h, the future return of the lowest price a^l, and the future return of the closing price a^c. The left and right spreads can be obtained by $l = a^c - a^l$ and $r = a^h - a^c$, respectively.

It is noted that transaction cost is of great importance for the final investment decision. To express terminal wealth more precisely, transaction cost is taken into consideration. Assume that there are no additional capital invested in all investment period

$$\mathrm{D}_t = \sum_{i=1}^{n} d_{t,i} |W_t x_{t,i} - W_{t-1} x_{t-1,i}| \tag{10}$$

satisfying $W_0 = 0$ and $x_{0,i} = 0$.

Next, an investment process with $\mathcal{H}$ different investment horizons is considered. In this process, the issue that how the composition of portfolio returns changes with the investment period is discussed. When $t - \mathcal{H} + 1 < 1$ (i.e., $t < \mathcal{H}$), t parts constitute the return at the beginning of period $t + 1$, namely the returns obtained from securities belonging to S_j and invested at the beginning of period $t-j+1$ ($j = 1, 2, ..., t$). The returns from securities with longer investment horizons fail to make themselves included, since these securities do not achieve expiration. When $t - \mathcal{H} + 1 \ge 1$ (i.e., $t \ge \mathcal{H}$), it is easier to analyze the composition for the reason that all securities invested in previous periods become mature. Then, W_{t+1} can be mathematically defined by

$$W_{t+1} = \begin{cases} \sum_{j=1}^{t} [W_{t-j+1} \sum_{i \in S_j} \zeta_{t-j+1,i} x_{t-j+1,i}] - \mathrm{D}_t, \\ \qquad t < \mathcal{H} \\ \sum_{j=1}^{\mathcal{H}} [W_{t-j+1} \sum_{i \in S_j} \zeta_{t-j+1,i} x_{t-j+1,i}] - \mathrm{D}_t, \\ \qquad t \ge \mathcal{H} \end{cases} \tag{11}$$

B. Three Investment Objectives

1) Expectation of Terminal Wealth: The investors who prefer long-term investments always lay strong emphasis on terminal wealth, aiming at obtaining considerable returns at a prespecified final time. In this case, the expectation of terminal wealth denoted by $\mathrm{E}[W_{T+1}]$ is taken as an objective to reflect investors' attitudes.

2) Conditional Value at Risk of Terminal Wealth: Risk measurement is another factor that draws investors' attention. As previously mentioned, many scholars devoted themselves to capturing the losses incurred in the investment in a more practical way and proposed diverse so-called risk measures.

Among these risk measures, CVaR is one of the most commonly adopted measures [4], [45], [46]. In this article, CVaR of terminal wealth is used to measure the risk that the whole investment may be exposed to.

For better explanation, the notion of VaR which is closely related to CVaR is introduced first. Presented by Morgan in 1996, VaR estimates the maximum loss that an investment may suffer during a given period under a certain confidence level. In this work, the VaR of W_{T+1}, $W_{(T+1)_{VaR}}(\beta)$ can be defined as follows:

$$W_{(T+1)_{\text{VaR}}}(\beta) = \inf\{x|\text{Cr}\{W_{T+1} \leq x\} \geq \beta\} \quad (12)$$

where β denotes the confidence level and $\beta \in (0, 1]$, Cr refers to the credibility measure.

As an improvement of VaR, CVaR is a risk measure that is used to calculate the mean loss exceeding VaR that an investment may experience during a given period under a certain confidence level. The tail of loss distribution that is ignored by VaR is considered in CVaR. Based on (12), it can be derived by

$$W_{(T+1)_{\text{CVaR}}}(\alpha) = \frac{\int_{\alpha}^{1} W_{(T+1)_{\text{VaR}}}(\beta)d\beta}{1-\alpha} \quad (13)$$

where α describes the prespecified confidence level satisfying $\alpha \in (0, 1]$.

3) Skewness of Terminal Wealth: Conventional portfolio models impose a strong assumption that the distributions of investment returns are spherical or elliptical. However, it turns out that securities with asymmetrically distributed returns are more widespread. Omitting this fact to construct a model will result in unrealistic investment decisions. The third moment of return, so-called skewness, has been demonstrated by numerous researchers [47], [48] to play a vital role in the case of asymmetric return distributions, especially in the multiperiod model. Considering its high relevance to decision making, the skewness of terminal wealth is taken as the third objective and defined as follows:

$$\text{S}[W_{T+1}] = \text{E}[(W_{T+1} - \text{E}[W_{T+1}])^3]. \quad (14)$$

C. Proposed Model

According to real-world constraints and objectives proposed above, a constrained multiperiod multiobjective portfolio model is formulated as follows:

$$\min \ W_{(T+1)_{\text{CVaR}}}(\alpha) \quad (15)$$

$$\max \ \text{E}[W_{T+1}] \quad (16)$$

$$\max \ \text{S}[W_{T+1}] \quad (17)$$

$$\text{subject to} \ \sum_{i=1}^{n} x_{t,i} = 1, \ t = 1, 2, ...T \quad (18)$$

$$x_{t,i} = 0, \ \text{for any } i \in S_j \text{ and } t + j > T + 1 \quad (19)$$

$$\sum_{i=1}^{n} y_{t,i} \leq Max_t, \ t = 1, 2, ...T \quad (20)$$

$$l_{t,i}y_{t,i} \leq x_{t,i} \leq u_{t,i}y_{t,i}, \ t = 1, 2, ...T, i = 1, 2, ...n \quad (21)$$

$$\frac{x_{t,i}W_t}{pc_{t,i}} \leq \lfloor (0.1 - N_{t,i}^0)TN_{t,i} \rfloor, \ t = 1, 2, ...T, i = 1, 2, ...n \quad (22)$$

$$x_{t,i} = z_{t,i}e_{t,i}, \ z_{t,i} \in \mathcal{N}, t = 1, 2, ...T, i = 1, 2, ...n \quad (23)$$

$$y_{t,i} \in \{0, 1\}, \ t = 1, 2, ...T, i = 1, 2, ...n \quad (24)$$

The budget constraint defined in (18) guarantees that available wealth in each investment period is invested completely. Equations (20) and (24) define the cardinality constraint that restricts the maximum number of securities that constitute a portfolio at period t to lessen transaction costs and maintain diversification. Bounding constraint is described through (21). This constraint points out that the proportion on security i included in the portfolio at period t is confined within a given interval. It embodies that forbidding administrative costs for tiny holdings is more consistent with investors' preference. Equation (22) defines institutional policies constraint that prevents the proportion of shares of a security held by an investment institution from exceeding 10%. The round-lot constraint described in (23) limits that the proportion on each security in the portfolio at period t is a multiple of its minimum transaction volume, where $\mathcal{N}$ refers to the collection of positive integers. Equation (19) describes an additional constraint to ensure that the available wealth will not be invested in the securities whose returns cannot be mature at the end of period T.

V. Proposed Intelligent Hybrid Algorithm

Considering triangular fuzzy random returns and real-world constraints, it will be an arduous task to solve the proposed model by using conventional optimization approaches. Thus, a novel intelligent hybrid algorithm (see Fig. 2) is introduced to resolve the proposed model.

A. Fuzzy Random Simulation

Liu [12] first proposed a fuzzy random simulation by incorporating random simulation and fuzzy simulation. Since then, its importance in the complicated calculations among fuzzy random variables has been proven by numerous scholars (e.g., [49]). In this work, fuzzy random simulation is utilized to derive an effective approximation of uncertain functions in the model.

For portfolio vector $\boldsymbol{x} = (\boldsymbol{x_1}, \boldsymbol{x_2}, \ldots, \boldsymbol{x_T})$, the procedure of obtaining expected value of the terminal wealth $\text{E}[W_{T+1}]$ is described in Algorithm 1.

Note that the minimum value of r obtained from (25) can be treated as an approximation to fuzzy VaR under condition $\text{Cr}\{W_{T+1}(\boldsymbol{\zeta^0}(\omega)) \leq r\} \geq \alpha$. Then, the procedure of obtaining $W_{(T+1)_{\text{CVaR}}}(\alpha)$ is summarized in Algorithm 2

$$\begin{aligned} L(r) = 0.5(&\max\{\mu_{t,i}|W_{T+1}(\boldsymbol{\zeta^0}(\omega)) \geq r\} \\ &+ 1 - \max\{\mu_{t,i}|W_{T+1}(\boldsymbol{\zeta^0}(\omega)) < r\}) \end{aligned} \quad (25)$$

where $t = 1, 2, ...T$, $i = 1, 2, ...n$. $\boldsymbol{\zeta^0}$ defines a fuzzy random vector whose randomness is determined by a random vector $\boldsymbol{X^0}$, and $\mu_{t,i}$ is the membership function of the fuzzy variable $\zeta_{t,i}^0(\omega)$.

Since $\text{S}[W_{T+1}]$ can be approximated by a process similar to $\text{E}[W_{T+1}]$, the procedure of acquiring $\text{S}[W_{T+1}]$ will not be presented here.

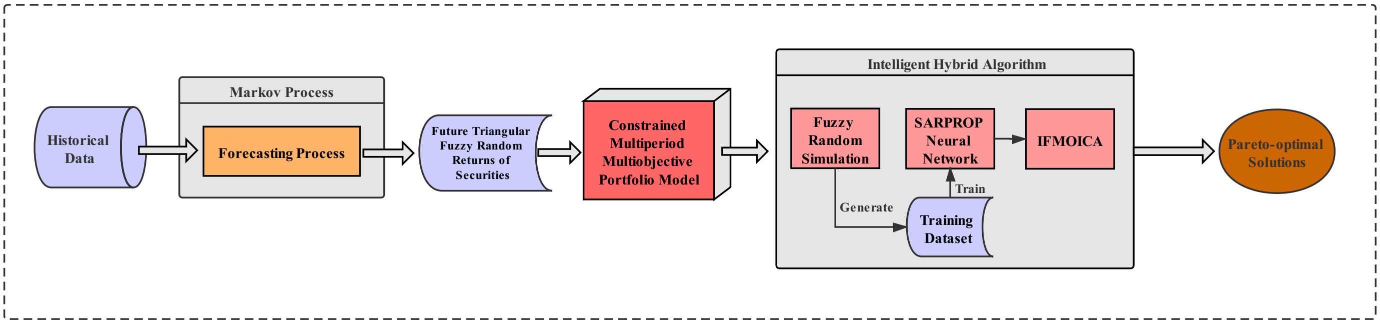


Fig. 2. Overall framework of the intelligent hybrid algorithm.

Algorithm 1: Procedure to Calculate $\mathrm{E}[W_{T+1}]$.

Input: Decision vector $\boldsymbol{x} = (\boldsymbol{x_1}, \boldsymbol{x_2}, \ldots, \boldsymbol{x_T})$, future return vector $\boldsymbol{\zeta}$, random vector $\boldsymbol{X}$ that determines the randomness of $\boldsymbol{\zeta}$, non-negative real number that is large enough M_1, positive integer that is large enough M_2
Output: $\mathcal{EW}$

set $\mathcal{EW} = 0$;
for *each* $i \in [0, M_2]$ **do**
  sample $\boldsymbol{X^0}$ according to the distribution of random vector $\boldsymbol{X}$;
  set $e_1 = 0$;
  set $e_2 = 0$;
  set $h = \frac{M_1}{M_2}$;
  for *each* $j \in [0, M_2]$ **do**
    $r_j = jh$;
    calculate $\mathrm{Cr}[j] = \mathrm{Cr}\{W_{T+1}(\boldsymbol{\zeta^0}(\omega)) \geq r_j\}$;
    $r'_j = -M_1 + jh$;
    calculate $\mathrm{Cr}'[j] = \mathrm{Cr}\{W_{T+1}(\boldsymbol{\zeta^0}(\omega)) \geq r'_j\}$;
  for *each* $j \in [0, M_2 - 1]$ **do**
    $e_1 = e_1 + \frac{h(\mathrm{Cr}[j]+\mathrm{Cr}[j+1])}{2}$;
    $e_2 = e_2 + \frac{h(\mathrm{Cr}'[j]+\mathrm{Cr}'[j+1])}{2}$;
  set $\eta^0 = e_1 - e_2$;
  $\mathcal{EW} = \mathcal{EW} + \eta^0$;
$\mathcal{EW} = \frac{\mathcal{EW}}{M_2}$;
return $\mathcal{EW}$;

Algorithm 2: Procedure to Obtain $W_{(T+1)_{\mathrm{CVaR}}}(\alpha)$.

Input: Decision vector $\boldsymbol{x} = (\boldsymbol{x_1}, \boldsymbol{x_2}, \ldots, \boldsymbol{x_T})$, future return vector $\boldsymbol{\zeta}$, random vector $\boldsymbol{X}$ that determines the randomness of $\boldsymbol{\zeta}$, confidence level α, positive integer that is large enough N
Output: $\mathcal{W}$

set $\mathcal{W} = 0$;
for *each* $i \in [0, N]$ **do**
  sample $\boldsymbol{X^0}$ according to the distribution of random vector $\boldsymbol{X}$;
  set $e = 0$;
  for *each* $j \in [0, N]$ **do**
    $\alpha_j = \alpha + j(1-\alpha)/N$;
    find the minimum value of r_j satisfying $L(r_j) \leq \alpha_j$; *// see (25)*
    $e = e + r_j(1-\alpha)/N$;
  $\mathcal{W} = \mathcal{W} + e/(1-\alpha)$;
$\mathcal{W} = \mathcal{W}/N$;
return $\mathcal{W}$;

B. Simulated Annealing Resilient Back Propagation Neural Network

Time consuming has become one of the major challenges faced by the fuzzy random simulation. Targeting at this demerit, a neural network is trained for objectives approximation while the fuzzy random simulation is used to generate the training data.

It is agreed that neural network offers an alternative way in diverse fields [50], [51]. For function approximating, it can learn from fundamental information in examples and then uniformly approximate the given functions which fully meets authors' requirements. First introduced by Treadgold and Gedeon in 1997, SARPROP [52] has been proven to be able to improve the convergence rate for neural networks training. By using SA, this method can also avoid converging to local minima. To benefit from the advantages mentioned above, the neural network trained by SARPROP is adopted for objectives approximating in the proposed algorithm. This enhancement will significantly improve computational efficiency of intelligent hybrid algorithm.

C. Improved Imperialist Competitive Algorithm

Inspired by the behavior of imperialist competition, the ICA begins with some countries. Some of them (more powerful ones) serve as imperialists and the rest are considered as colonies. In fact, the number of imperialists is essential to the performance of the ICA. Relevant literature emphasizes that 10–13% of the number of countries seems to be a good choice. In this article, the proportion is set as 10%. After the country classification, one imperialist and some colonies then constitute an empire. The number of colonies in the empire is determined by the imperialist on the basis of its power. In the subsequent assimilation operation, all colonies in each empire move toward their related imperialist. Competition is then happened among empires. Empires try to supplant each other and the weakest one loses the colonies it possesses and collapses finally. For details, please refer to [53].

As a promising choice to handle complicated portfolio models, ICA possesses some remarkable properties which are suitable for solving the proposed model. For example, it is powerful in the sense of dealing with large number of decision variables, and is less dependent on initial solutions. Nevertheless, it also has some drawbacks, e.g., premature convergence. To overcome those shortcomings, a novel improved multiobjective ICA named IFMOICA is used to conduct the search for Pareto-optimal solutions.

IFMOICA adopts a new search mechanism to promote the capability of exploration and develops a novel two-step method to select solutions.

1) Novel Two-Step Method: A reasonable and effective benchmark for solution ranking is of great importance in multiobjective portfolio problems. In IFMOICA, a newly proposed two-step method is employed to accomplish this job.

The method starts with ranking countries in the current generation by fast non-dominated sorting. Then, non-dominated countries are assorted into rank 1 and are added to a external archive A. The imperialists are also chosen from the countries in rank 1.

A metric in accordance with the objectives and the ranks given in step 1 is then proposed in the next step of method. Archive A will finally reach its maximum capacity A_{max} as the search progresses. In this case, the proposed metric serves as a quantitative measure to calculate the power of countries which help Archive A maintain high quality solutions. Additionally, it also plays an important role in creating and comparing the empires. In this article, the power of country c belonging to Rank R is illustrated as follows:

$$F_{i,c} = \frac{|f_{i,c} - f_i^{\text{worst}}|}{f_i^{\max} - f_i^{\min}} \tag{26}$$

$$\text{Cost}_c = \sum_{i=1}^{N_o} [\frac{F_{i,c}}{\sum_{j=1}^{N_R} F_{i,j}}](R-1)N_o \tag{27}$$

$$P_c = \frac{1}{\text{Cost}_c} \tag{28}$$

where $f_{i,c}$ represents the ith objective value of country c. The maximum value, minimum value, and worst value of the ith objective in the current generation are denoted by $f_i^{\max}$, $f_i^{\min}$, and f_i^{worst}, respectively. N_o refers to the number of objectives in the model. N_R describes the number of countries belonging to Rank R.

Each objective value is normalized twice. The first time [see (26)] implemented within the whole current generation would handle the issue that objective values are too concentrated, and the second [the first part of (27)] occurs within the rank to guarantee the dominance of R in the cost value. All normalized objectives values are summed to generate the cost value, and the power of the country can be denoted with the reciprocal of the cost value.

2) New Search Mechanism: A notable feature of original ICA is that all colonies in empires perform movements toward their relevant imperialist in light of the attraction mechanism. Limitations of this operation may result in the so-called prematurity or poor convergence speed. It is clear that maintaining diversity in countries has become an urgent issue.

Developed by Yang [54], the firefly algorithm (FA) has been proven to have the potential to outperform other EAs, e.g., GA and PSO, in some fields [55]. FA captures the flash pattern of fireflies and holds many highlights. Automatic subdivision is one of the most outstanding highlights. Benefiting from the behavior of fireflies, the whole population in FA can be divided into subgroups automatically, and each subgroup is allowed to swarm separately. This feature helps FA preserve diversity of the whole population, which is exactly what the ICA needs. Based on the fact mentioned above, a new search mechanism is introduced by adopting the flashing characteristics of fireflies. Two actions compose the new search mechanism:

- Simultaneous movement of colony toward its imperialist and the most powerful imperialist: As noted, the movement to the most powerful imperialist does not exist in the original ICA. However, in fact, it seems a fair inference that a colony would desire to move toward the most powerful imperialist since each colony wants to reach a better socio-political position. Aroused by this fact, a new move operation is added. Then, the first action which combines the moving mechanism in FA with this new move operation can be defined as follows:

$$\begin{aligned} \text{col}_i^{\text{new}} &= \text{col}_i^{\text{old}} + M\beta_0 e^{-\gamma r_{i,\text{imp}}^2}(\text{imp}_{\text{col}_i} - \text{col}_i^{\text{old}}) \\ &+ (1-M)\beta_0 e^{-\gamma r_{i,\text{stro}}^2}(\text{imp}_{\text{stro}} - \text{col}_i^{\text{old}}) \\ &+ \alpha(\epsilon - \frac{1}{2}) \end{aligned} \tag{29}$$

where the new position and old position of colony i are denoted by $\text{col}_i^{\text{new}}$ and $\text{col}_i^{\text{old}}$, respectively. $\text{imp}_{\text{col}_i}$ and imp_{stro} refer to the positions of the related imperialist of colony i and the most powerful imperialist, separately. M is the move rate satisfying $M \in [0.1, 0.9]$. The same as in FA, γ and β_0 define the light absorption coefficient and the maximum attractiveness value, respectively. r is the Cartesian distance from the current colony to the chosen one. The last term is randomization with a parameter α and a random vector ϵ drawn uniformly from $[0, 1]$.
- Movement caused by the interaction of colonies: This action takes its inspiration from the interaction between the colonies, and assumes that all colonies are obliged to spread their effect in order to compete with each other and improve socio-political position. Interaction of colonies is reflected in (30) and (31),

$$\begin{aligned} \text{col}_i^{\text{new}} &= \text{col}_i^{\text{old}} + \beta_0 e^{-\gamma r_{j,jj}^2}(\text{col}_j - \text{col}_{jj}) \\ &+ \alpha(\epsilon - \frac{1}{2}), \quad \text{if } P_j > P_{jj} \end{aligned} \tag{30}$$

$$\begin{aligned} \text{col}_i^{\text{new}} &= \text{col}_i^{\text{old}} + \beta_0 e^{-\gamma r_{jj,j}^2}(\text{col}_{jj} - \text{col}_j) \\ &+ \alpha(\epsilon - \frac{1}{2}), \quad \text{if } P_j \le P_{jj} \end{aligned} \tag{31}$$

where col_j and col_{jj} refer to the positions of colony j and colony jj, respectively. P_j and P_{jj} denote their power, respectively. Colonies j and jj are randomly filtered from the colonies within the empire that colony i belongs to.

Colony i would tend to move toward colony j if colony j is more powerful than colony jj and toward colony jj, otherwise.

Note that, to guarantee that colonies move to the prospective direction,

$$\mathrm{col}_i = \begin{cases} \mathrm{col}_i^{\mathrm{new}} & \mathrm{col}_i^{\mathrm{new}} \text{ is better than } \mathrm{col}_i^{\mathrm{old}} \\ \mathrm{col}_i^{\mathrm{old}} & \text{otherwise} \end{cases} \tag{32}$$

where col_i refers to the position of colony i in the current generation. col_i would maintain if colony i does not reach a better position after moving.

The flow chart of IFMOICA is illustrated in Fig. 3. Apart from the abovementioned innovation, the repair method is adopted to cope with constraints. Each country that does not satisfy all the constraints considered needs to be repaired before being accepted as a member. The algorithm will continue until stopping criterion is met, i.e., there only exists one empire that controls all countries or the maximum number of iterations is reached.

D. The Intelligent Hybrid Algorithm

On the basis of all of the above knowledge, a more effective and powerful intelligent hybrid algorithm is proposed. In this algorithm, a trained SARPROP neural network provides approximate objectives with the help of fuzzy random simulation and the newly developed IFMOICA searches solution space to form the Pareto-optimal solutions. The pseudocode of the intelligent hybrid algorithm is described in Algorithm 3.

VI. Intelligent Parallel Hybrid Algorithm

In fact, multiple objective evaluations (especially the training dataset generation part) still make computational time a significant challenge for the proposed algorithm, which means that the Pareto frontier cannot be obtained in a satisfying time especially when facing the large scale securities data. However, it is obvious that the proposed algorithm holds an inherent parallel structure, implying that algorithmic parallelization can be a good choice to improve computational efficiency. In this article, the MPI technology is applied for the purpose of improving the efficiency of training data generation and exploring a larger search space.

It is well known that the master-slave paradigm is very easy to implement and manage. Consequently, the intelligent hybrid algorithm is parallelized based on master-slave paradigm, and its flowchart is illustrated in Fig. 4. Like in all of master-slave parallel algorithms, apart from handling its own computational task, the master processor in the parallelization takes the responsibility of coordinating all involved operations, while slave processors only perform the tasks given to them. In detail, each processor implements Markov process first to predict future fuzzy random security returns which would avoid unnecessary data exchange. After that, the master processor informs slave processors the size of data that they should generate. For load balancing, the sizes are generally equal. Then, each processor preforms fuzzy random simulation to complete this task. Master processor begins to gather the obtained data and spell them into one dataset when all the simulations are finished. Afterward, each processor trains a SARPROP neural network on the basis of the previous dataset followed by sending their weight matrixes to the master processor. A weights-choosing procedure is then carried out by master processor for weight matrixes selection. In this procedure, the weight matrix with highest accuracy on the validation set is selected and sent to all slave processors. Next, each processor executes IFMOICA and implements migration after some decades (which is different for different runs) by sending their Pareto-optimal solutions to master processor. Moreover, the master processor computes the current Pareto-optimal front and sends it back. It should be noted that all processors maintain the same number of countries and implement migrations synchronously. The master processor outputs the Pareto-optimal solutions when the set termination

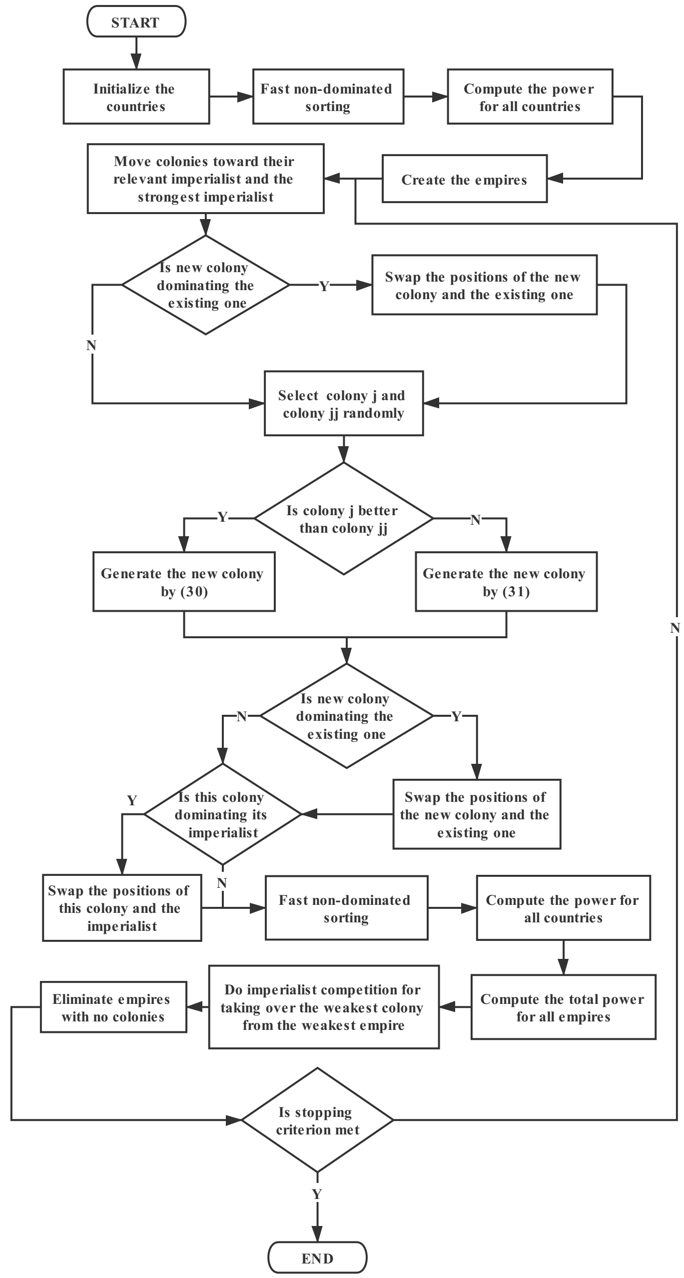


Fig. 3. Flowchart of IFMOICA.

Algorithm 3: The Intelligent Hybrid Algorithm

Input: Historical data on the security returns, the size of the archive A $A_{\max}$, population size $PopSize$

Output: The archive A

for $i \leq$ *the given size of training dataset* **do**
  generate decision vector $\boldsymbol{x} = (\boldsymbol{x_1}, \boldsymbol{x_2}, \ldots, \boldsymbol{x_T})$ randomly;
  data$[i][0 : T * n - 1] \leftarrow (\boldsymbol{x_1}, \boldsymbol{x_2}, \ldots, \boldsymbol{x_T})$;
  data$[i][T * n]$ = E$[W_{T+1}]$; *// see Algorithm 1*
  data$[i][T * n + 1]$ = $W_{(T+1)_{\text{CVaR}}}(\alpha)$; *// see Algorithm 2*
  data$[i][T * n + 2]$ = S$[W_{T+1}]$;
nn $\leftarrow$ initialize a SARPROP neural network;
while $Iters \leq MaxIter$ *and* $mse > DesiredErr$ **do**
  train_on_data (nn, data); *// train the neural network*
saveWeights (weightsfile);
for $i \in [1, PopSize]$ **do**
  Countries$[i]$.portfolio $\leftarrow$ initialize a country randomly;
  Countries$[i]$.objectives $\leftarrow$ nn.outPut (Countries$[i]$.portfolio);
Fast_Non_Dominated_Sort (Countries, $PopSize$);
Power (Countries); *// calculate power for all countries*
CreateEmpire (); *// create empires*
if *the number of countries with rank* $1 < A_{\max}$ **then**
  maintain the archive A with countries with rank 1;
else
  maintain the archive A with the best $A_{\max}$ countries with rank 1, ranked by power;
while *stopping criterion not met* **do**
  Assimilation (); *// assimilation operator and update archive A*
  Competition (); *// imperialist Competition*
  EliminateEmpire (); *// eliminate the empire with no colonies*
return archive A;

condition is met.

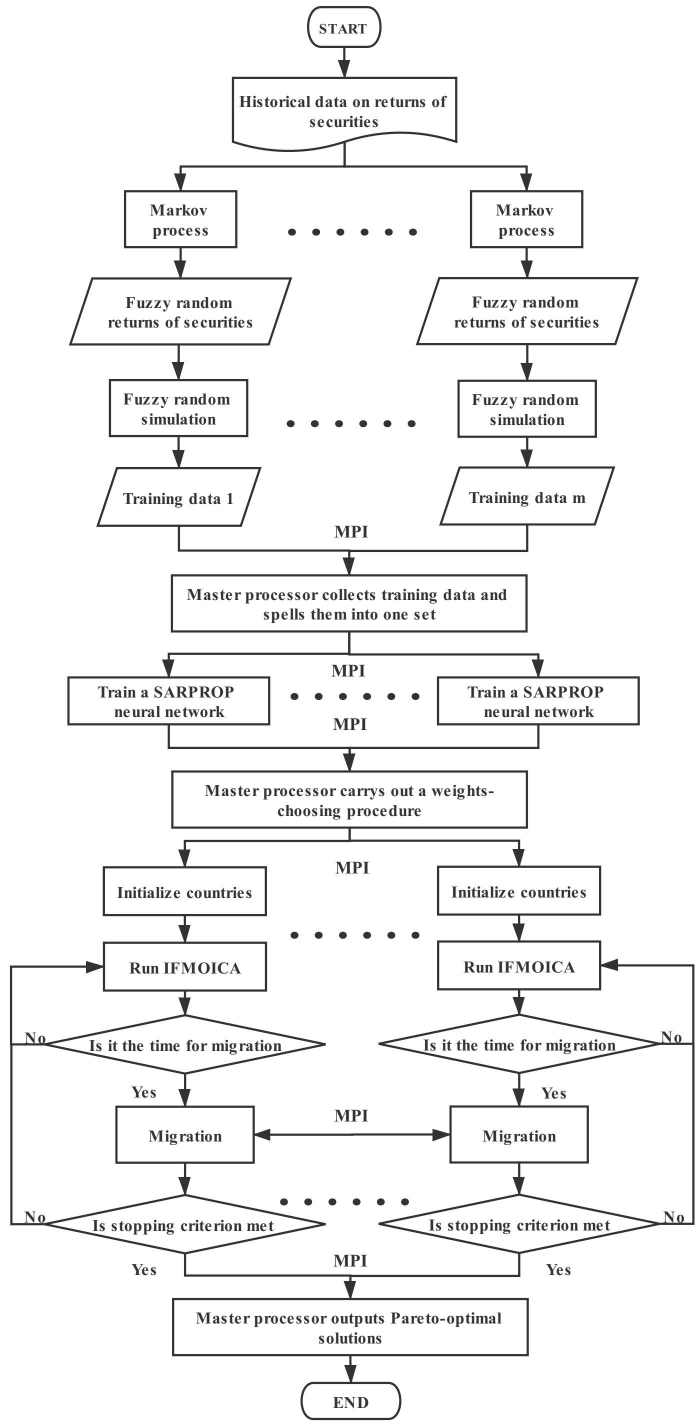


Fig. 4. Flowchart of the intelligent parallel hybrid algorithm.

VII. Performance Evaluation

In this section, parameter settings for the proposed model and algorithms are introduced first. After hybridizing NSGA-II with fuzzy random simulation and SARPROP neural network to form an algorithm named FRNSGA-II, a comparison between the intelligent hybrid algorithm and FRNSGA-II is given. In this comparison, the running time and two widely adopted performance metrics are used to evaluate the performance of algorithms on three different real historical financial datasets. Whereafter, the intelligent parallel hybrid algorithm is tested on the same three data sets to verify its scalability.

A. Parameters Setup

For all datasets used, presume that the securities that investors may be interested in are with three investment horizons. Securities with higher volatility are considered to be more suitable for shorter investment horizon. Meanwhile, as mentioned before, the triangular fuzzy random returns are predicted by Markov process based on historical data from the Shanghai Stock Exchange. Additionally, the initial wealth is set as $W_1 = 1$. Table II is devoted to summarizing the parameters in the model. It can be noticed that different securities in different investment periods take the identical unit transaction cost. In fact, the unit transaction cost in the financial market is composed of transfer fee, stamp duty, and commission. The first one is often ignored in academic

TABLE II
PARAMETERS FOR THE PROPOSED MODEL

α	Max_t	$d_{t,i}$	$l_{t,i}$	$u_{t,i}$	$TN_{t,i}$	$N^0_{t,i}$
0.9	10	0.003	0.05	1	$[8e^6, 2e^{10}]$	$[0, 0.085]$

TABLE III
GENERAL PARAMETERS FOR IFMOICA AND NSGA-II

Parameters	IFMOICA	NSGA-II
population size	600	600
number of generation	2000	2000
archive size	100	–
number of imperialists	60	–
moving rate	0.6	–
maximum attractiveness value	3	–
absorption coefficient	0.00001	–
mutation probability	–	0.1
crossover probability	–	0.5

research due to its small proportion. In this study, it is assumed that the last two parts of each security are the same in the overall investment period, and the unit transaction cost is thus set to 0.003 which is an appropriate choice for most of the case studies. Some general parameters for IFMOICA and NSGA-II are specified in Table III.

B. Experiment Results and Analysis

1) Quality Indicator: Several performance metrics have been developed to measure qualitative characteristics of Pareto-optimal solutions obtained from the multiobjective EAs. In this article, spacing (S) and hypervolume (HV) metric are used.

Spacing (S): S metric measures the variance of nearest distance between elements in non-dominated frontier found. It numerically describes the spread of elements in the obtained non-dominated frontier and is defined as follows:

$$S = \sqrt{\frac{1}{|Q|-1}\sum_{i=1}^{|Q|}(\overline{d}-d_i)^2} \tag{33}$$

$$d_i = \min_j(\sum_{m=1}^{M} |f_m^i(x) - f_m^j(x)|),\ j = 1, 2, ..., |Q| \tag{34}$$

where Q refers to non-dominated frontier generated by an algorithm, and $|Q|$ defines the number of elements in Q. $\overline{d}$ is the mean of all d_i. M denotes the dimension of the element in Q. A smaller value of S indicates that the elements in the found non-dominated frontier are more evenly spaced.

Hypervolume (HV): The HV metric is defined to obtain the volume enclosed by elements in the found non-dominated frontier with respect to the objective space, namely a summation of hypercubes bounded by a reference point. This metric captures both the proximity of the found non-dominated frontier to the true Pareto frontier and the diversity of the former. Mathematically,

$$\text{HV} = \text{volume}(\cup_{s=1}^{|Q|}\mathcal{V}_s) \tag{35}$$

where Q denotes the non-dominated frontier generated by an algorithm. $\mathcal{V}_s$ refers to the hypercube from element s to the reference point. $|Q|$ describes the number of members in Q. A larger value of this metric is more preferred when comparing the results of the two algorithms.

For a more accurate computation, the objective space is normalized by (36).

$$f_i = \frac{f_i - f_i^{\min}}{f_i^{\max} - f_i^{\min}} \tag{36}$$

where $f_i^{\max}$ and $f_i^{\min}$ refer to the maximum and minimum value of the ith objective produced by all runs of the two algorithms. Generally, the reference point is formed by worst values of the objectives. In this work, it is set as $(1, 0, 0)$.

2) Comparison of the Algorithms: A set of experiments are carried out to provide the comparison between the intelligent hybrid algorithm and FRNSGA-II in terms of running time, HV, S metrics and Pareto frontier. In this comparison, historical data from the Shanghai Stock Exchange (i.e., Datasets 1-3) is used, which covers the price data over a period of 23 years from April 1996 to April 2019. On Dataset 1, a six-month portfolio selection problem with 10 securities is considered, and a dataset of 6144 data point generated from fuzzy random simulation is used as the training dataset for SARPROP neural network. On Dataset 2, the size of problem increases to 30 securities. Accordingly, the size of training set is raised to 12288. A three-month large scale portfolio selection problem with 402 securities is based on Dataset 3, in which the number of training set data points is 36864.

As the problems of this type are computationally difficult, the true Pareto frontiers are not known. Thus, all non-dominated solutions obtained from the two algorithms are integrated to form the best known Pareto frontier which is a reference set (BKPF).

The running time, HV, S metrics and Pareto frontier of the two algorithms on the three datasets are summarized in Fig. 5-7. For simplicity, only the running time of the last part in the two algorithms is compared. The results show that the intelligent hybrid algorithm is significantly faster than FRNSGA-II for all the problem instances. For example, compared to FRNSGA-II, a dramatic decrease (nearly 96.4%) is seen in the mean value for running time of the intelligent hybrid algorithm on Dataset 3. Moreover, the intelligent hybrid algorithm holds smaller mean values for S metric and larger mean values for HV on Dataset 2 and Dataset 3. On Dataset 1, the performance of the two algorithms is almost comparable since the intelligent hybrid algorithm obtains a better mean value for HV and a slightly underperformed mean value for S metric.

From the perspective of illustration, the comparison between the obtained Pareto frontiers of the algorithms and BKPF on the three datasets is also presented. On Dataset 1, the Pareto frontiers generated by the two algorithms are all difficult to differentiate visually from BKPF, demonstrating that they will be competitive to each other. As the size of securities increases, the intelligent hybrid algorithm begins to show its superiority and maintains a better Pareto frontier even when the number of securities exceeds 400. It can be inferred from

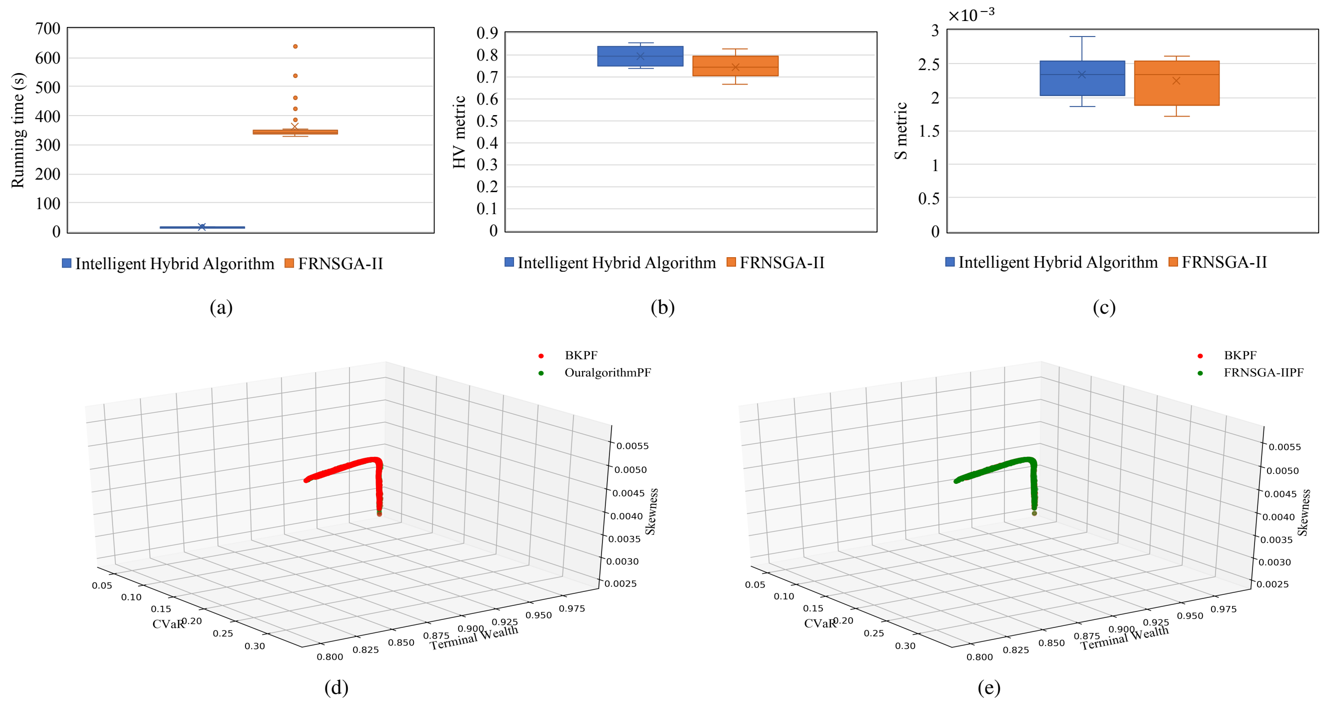


Fig. 5. Comparison between the intelligent hybrid algorithm and FRNSGA-II on Dataset 1. (a) Running time. (b) HV metric. (c) S metric. (d) Pareto frontiers of the intelligent hybrid algorithm and BKPF. (e) Pareto frontiers of FRNSGA-II and BKPF.

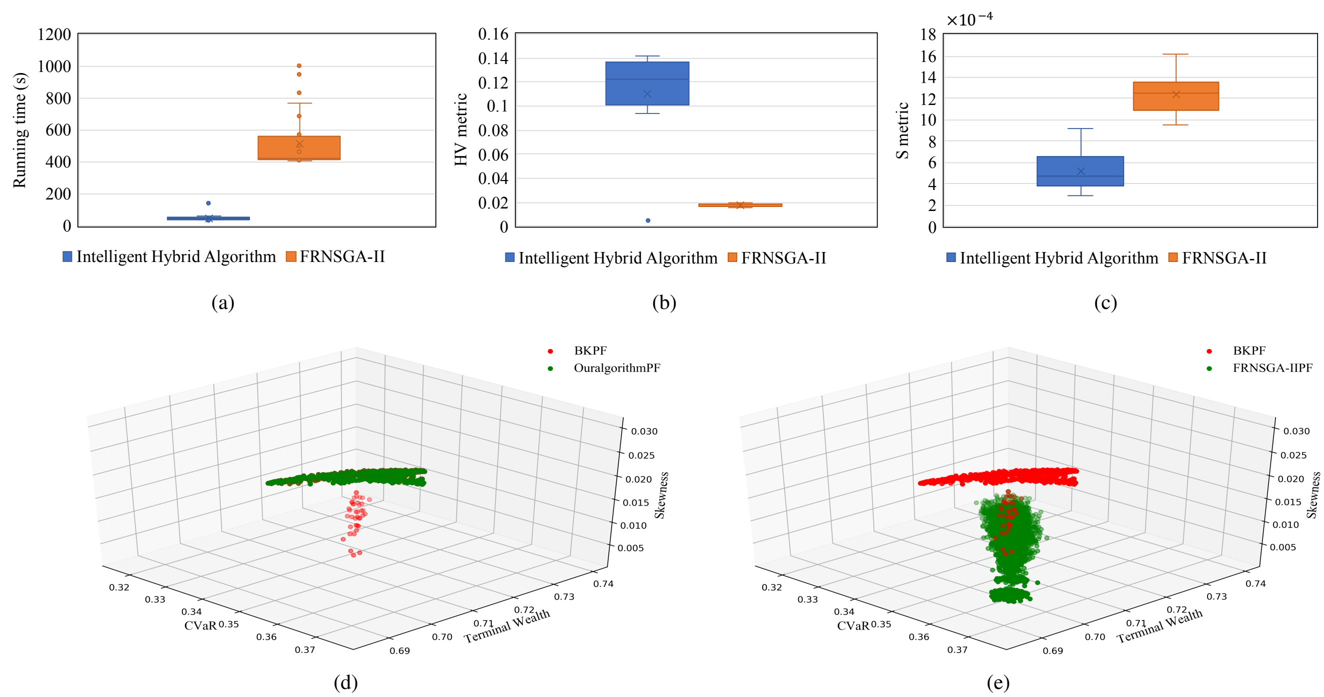


Fig. 6. Comparison between the intelligent hybrid algorithm and FRNSGA-II on Dataset 2. (a) Running time. (b) HV metric. (c) S metric. (d) Pareto frontiers of the intelligent hybrid algorithm and BKPF. (e) Pareto frontiers of FRNSGA-II and BKPF.

the abovementioned facts that the intelligent hybrid algorithm can solve the constrained multiperiod multiobjective portfolio model efficiently even facing large scale securities data and will achieve a better performance over FRNSGA-II.

3) Scalability of the Intelligent Parallel Hybrid Algorithm: Speedup and parallel efficiency are significantly vital to proving the scalability of a parallel algorithm. In this work, the running time, speedup, and parallel efficiency on the three datasets are illustrated in Figs. 8-10. The results show that running time can be reduced dramatically for all the problem instances by conducting parallelization. For example, with 3072 processors, it only takes nearly 5 h to solve the model with the size of securities up to 402 which will be an impossible task for a single processor. Furthermore, the results also confirm that the parallel algorithm is scalable to some extent. Take speedup and parallel efficiency on Dataset 1 as an example, the parallel

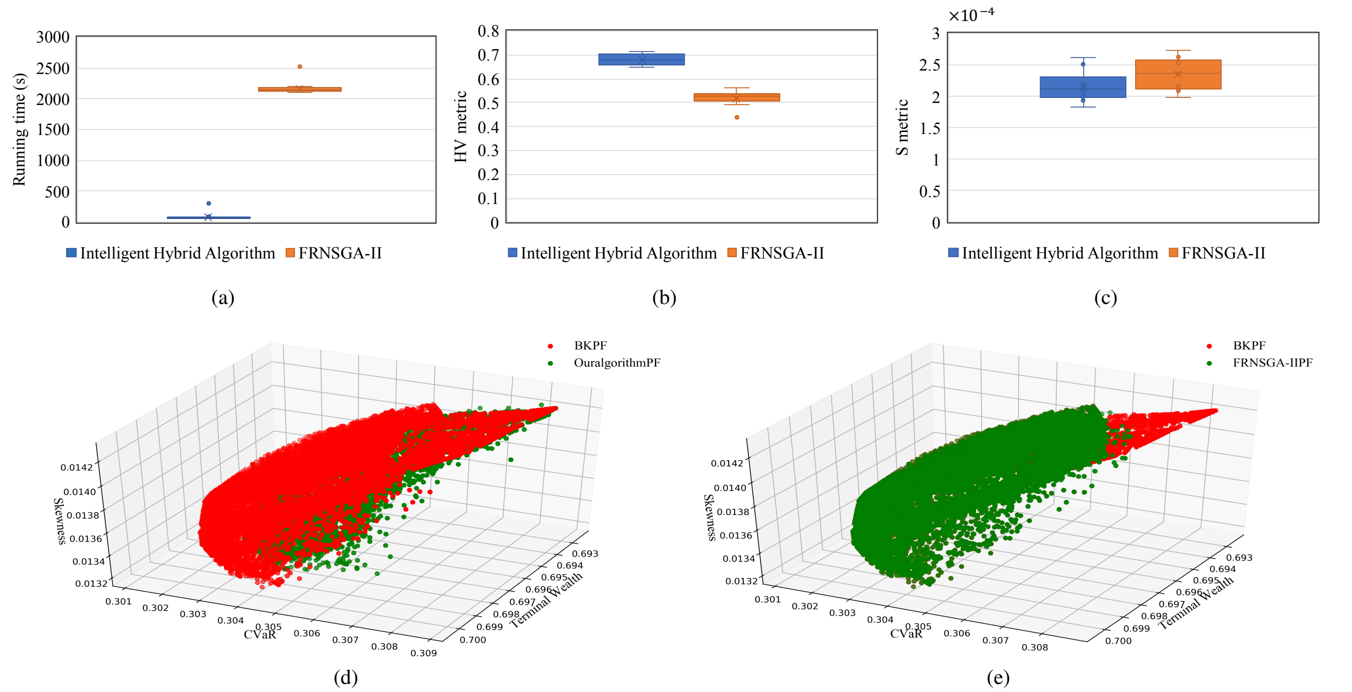


Fig. 7. Comparison between the intelligent hybrid algorithm and FRNSGA-II on Dataset 3. (a) Running time. (b) HV metric. (c) S metric. (d) Pareto frontiers of the intelligent hybrid algorithm and BKPF. (e) Pareto frontiers of FRNSGA-II and BKPF.

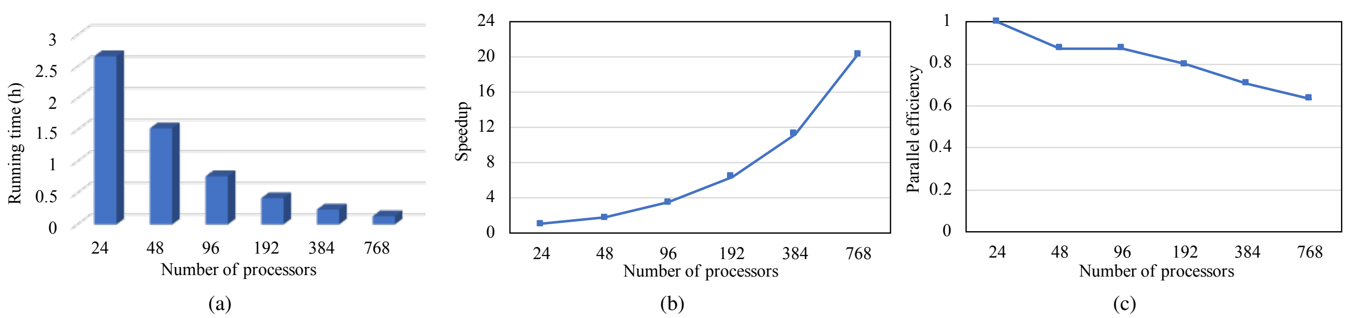


Fig. 8. Running time, speedup and parallel efficiency on Dataset 1. (a) Running time. (b) Speedup. (c) Parallel efficiency.

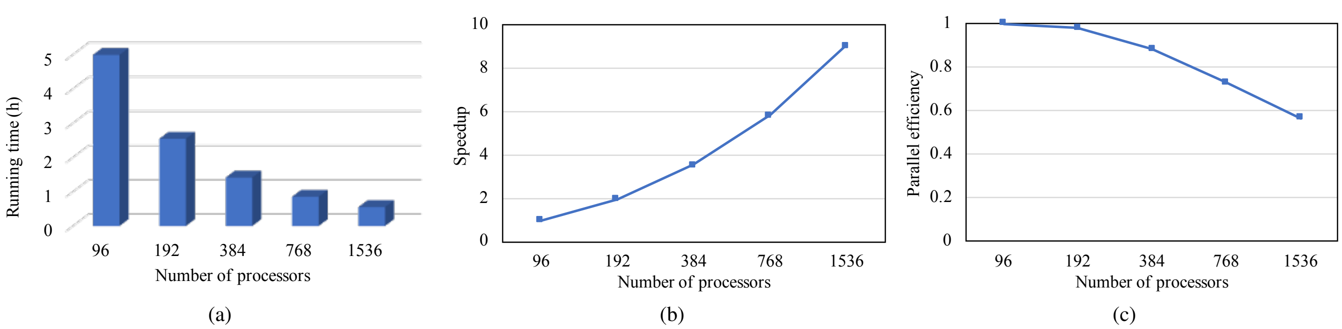


Fig. 9. Running time, speedup and parallel efficiency on Dataset 2. (a) Running time. (b) Speedup. (c) Parallel efficiency.

algorithm maintains good speedup and parallel efficiency when the number of processors is relatively small. As the number of processors increases, a significantly drop is seen in the performance of speedup and parallel efficiency as a result of frequent communications and massive data exchange among processors. Despite the decline, the parallel efficiency remains above 60%. The speedup and parallel efficiency on Dataset 2 and Dataset 3 have a similar performance with Dataset 1.

As an overall result, the parallelization of the intelligent hybrid algorithm not only improves the computational efficiency but also ensures the superiority when dealing with large scale securities data.

VIII. Conclusion

In this article, a multiperiod multiobjective portfolio model with real-world constraints was developed to handle the portfolio selection problem based on large scale securities data in

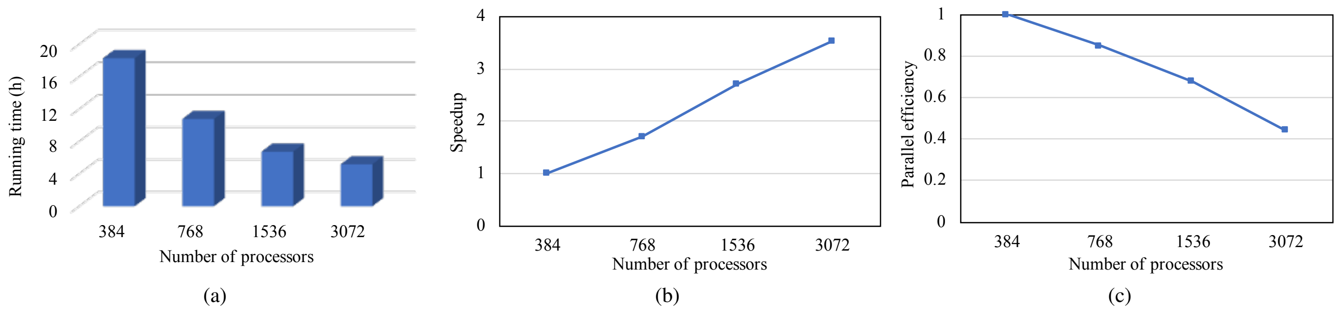


Fig. 10. Running time, speedup and parallel efficiency on Dataset 3. (a) Running time. (b) Speedup. (c) Parallel efficiency.

the uncertain market. In this model, the randomness and fuzziness in the financial market were captured by treating future security returns as fuzzy random variables. Moreover, terminal wealth, CVaR, and skewness were considered as tricriteria for decision making. Aiming to effectively resolve the proposed model, a novel intelligent hybrid algorithm was introduced. In this algorithm, a trained SARPROP neural network provided approximate objectives with the help of fuzzy random simulation, and the newly proposed IFMOICA was used to form the Pareto-optimal solutions. It was demonstrated that the intelligent hybrid algorithm achieves a better performance over FRNSGA-II. The improvement is not only in the running time but also in the quality of the Pareto frontier. To improve computational efficiency and cope with large scale securities data, an intelligent parallel hybrid algorithm adopting MPI technology was presented. The experiments certified that the parallel algorithm has a good scalability and could solve the large scale portfolio selection problem consisting of securities up to 402 in a reasonable time.

For future work, the proposed model can be extended with more trading constraints. Furthermore, the proposed algorithm can be tested on larger historical financial datasets and generalized to more fuzzy random portfolio selection problems.

References

[1] H. Markowitz, "Portfolio selection," *The journal of finance*, vol. 7, no. 1, pp. 77–91, 1952.

[2] E. F. Fama and R. Roll, "Parameter estimates for symmetric stable distributions," *Journal of the American Statistical Association*, vol. 66, no. 334, pp. 331–338, 1971.

[3] P. Artzner, F. Delbaen, J. M. Eber, and D. Heath, "Coherent measures of risk," *Mathematical finance*, vol. 9, no. 3, pp. 203–228, 1999.

[4] R. T. Rockafellar and S. Uryasev, "Optimization of conditional value-at-risk," *Journal of risk*, vol. 2, no. 3, pp. 21–41, 2000.

[5] L. A. Zadeh, "Fuzzy sets," *Information and control*, vol. 8, no. 3, pp. 338–353, 1965.

[6] B. Liu, "Random fuzzy dependent-chance programming and its hybrid intelligent algorithm," *Information sciences*, vol. 141, no. 3-4, pp. 259–271, 2002.

[7] T. Hasuike, H. Katagiri, and H. Ishii, "Portfolio selection problems with random fuzzy variable returns," *Fuzzy Sets and systems*, vol. 160, no. 18, pp. 2579–2596, 2009.

[8] H. Katagiri, T. Uno, K. Kato, H. Tsuda, and H. Tsubaki, "Random fuzzy multi-objective linear programming: Optimization of possibilistic value at risk (pVaR)," *Expert Systems with Applications*, vol. 40, no. 2, pp. 563–574, 2013.

[9] H. Kwakernaak, "Fuzzy random variables—I. Definitions and theorems," *Information sciences*, vol. 15, no. 1, pp. 1–29, 1978.

[10] D. X. Shaw, S. Liu, and L. Kopman, "Lagrangian relaxation procedure for cardinality-constrained portfolio optimization," *Optimisation Methods & Software*, vol. 23, no. 3, pp. 411–420, 2008.

[11] D. Bertsimas and R. Shioda, "Algorithm for cardinality-constrained quadratic optimization," *Computational Optimization and Applications*, vol. 43, no. 1, pp. 1–22, 2009.

[12] B. Liu, "Fuzzy random chance-constrained programming," *IEEE Transactions on Fuzzy systems*, vol. 9, no. 5, pp. 713–720, 2001.

[13] J. Li and J. Xu, "Multi-objective portfolio selection model with fuzzy random returns and a compromise approach-based genetic algorithm," *Information Sciences*, vol. 220, pp. 507–521, 2013.

[14] Y. J. Liu, W. G. Zhang, and W. J. Xu, "Fuzzy multi-period portfolio selection optimization models using multiple criteria," *Automatica*, vol. 48, no. 12, pp. 3042–3053, 2012.

[15] M. Chen, S. Mao, and Y. Liu, "Big data: A survey," *Mobile networks and applications*, vol. 19, no. 2, pp. 171–209, 2014.

[16] H. Markowitz, *Portfolio selection: Efficient diversification of investments*. New York, NY, USA: Wiley, 1959.

[17] H. Konno and H. Yamazaki, "Mean-absolute deviation portfolio optimization model and its applications to tokyo stock market," *Management science*, vol. 37, no. 5, pp. 519–531, 1991.

[18] A. Chekhlov, S. Uryasev, and M. Zabarankin, "Portfolio optimization with drawdown constraints," in *Supply chain and finance*. World Scientific, 2004, pp. 209–228.

[19] H. Tanaka and P. Guo, "Portfolio selection based on upper and lower exponential possibility distributions," *European Journal of operational research*, vol. 114, no. 1, pp. 115–126, 1999.

[20] X. Huang, "Fuzzy chance-constrained portfolio selection," *Applied mathematics and computation*, vol. 177, no. 2, pp. 500–507, 2006.

[21] P. Gupta, M. K. Mehlawat, and A. Saxena, "Asset portfolio optimization using fuzzy mathematical programming," *Information Sciences*, vol. 178, no. 6, pp. 1734–1755, 2008.

[22] A. Hajnoori, M. Amiri, and A. Alimi, "Forecasting stock price using grey-fuzzy technique and portfolio optimization by invasive weed optimization algorithm," *Decision Science Letters*, vol. 2, no. 3, pp. 175–184, 2013.

[23] C. Li, Z. Lu, Y. Hu, F. Liu, and J. Wang, "A parallel hybrid intelligent algorithm for fuzzy mean-cvar portfolio model," in *2017 IEEE 19th International Conference on High Performance Computing and Communications; IEEE 15th International Conference on Smart City; IEEE 3rd International Conference on Data Science and Systems (HPCC/SmartCity/DSS)*. IEEE, 2017, pp. 348–355.

[24] G. A. V. Pai, "Fuzzy decision theory based metaheuristic portfolio optimization and active rebalancing using interval type-2 fuzzy sets," *IEEE Transactions on Fuzzy Systems*, vol. 25, no. 2, pp. 377–391, 2017.

[25] H. Katagiri and H. Ishii, "Fuzzy portfolio selection problem," in *IEEE SMC'99 Conference Proceedings. 1999 IEEE International Conference on Systems, Man, and Cybernetics (Cat. No. 99CH37028)*, vol. 3. IEEE, 1999, pp. 973–978.

[26] J. Li and J. Xu, "A novel portfolio selection model in a hybrid uncertain environment," *Omega*, vol. 37, no. 2, pp. 439–449, 2009.

[27] Y. Yoshida, "An estimation model of value-at-risk portfolio under uncertainty," *Fuzzy Sets and Systems*, vol. 160, no. 22, pp. 3250–3262, 2009.

[28] W. Sun, W. Zhang, and W. Xu, "Portfolio optimization strategy under fuzzy random environment with investor sentiment," in *2015 11th International Conference on Natural Computation (ICNC)*. IEEE, 2015, pp. 1240–1245.

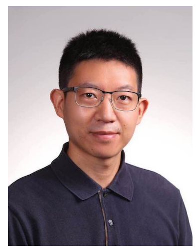
[29] K. Lwin, R. Qu, and G. Kendall, "A learning-guided multi-objective evolutionary algorithm for constrained portfolio optimization," *Applied Soft Computing*, vol. 24, pp. 757–772, 2014.

[30] S. Babaei, M. M. Sepehri, and E. Babaei, "Multi-objective portfolio optimization considering the dependence structure of asset returns,"

European Journal of Operational Research, vol. 244, no. 2, pp. 525–539, 2015.
[31] A. R. Pouya, M. Solimanpur, and M. J. Rezaee, "Solving multi-objective portfolio optimization problem using invasive weed optimization," *Swarm and Evolutionary Computation*, vol. 28, pp. 42–57, 2016.
[32] H. Khanjarpanah and M. S. Pishvaee, "A fuzzy robust programming approach to multi-objective portfolio optimisation problem under uncertainty," *International Journal of Mathematics in Operational Research*, vol. 12, no. 1, pp. 45–65, 2018.
[33] M. Kaucic, "Equity portfolio management with cardinality constraints and risk parity control using multi-objective particle swarm optimization," *Computers & Operations Research*, vol. 109, pp. 300–316, 2019.
[34] A. Loraschi, M. Tomassini, A. Tettamanzi, and P. Verda, "Distributed genetic algorithms with an application to portfolio selection problems," in *Artificial neural nets and genetic algorithms*. Berlin, Germany: Springer, 1995, pp. 384–387.
[35] F. Streichert, H. Ulmer, and A. Zell, "Evolutionary algorithms and the cardinality constrained portfolio optimization problem," in *Operations Research Proceedings 2003*. Springer, 2004, pp. 253–260.
[36] 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.
[37] B. Wang, Y. Li, S. Wang, and J. Watada, "A multi-objective portfolio selection model with fuzzy value-at-risk ratio," *IEEE Transactions on Fuzzy Systems*, vol. 26, no. 6, pp. 3673–3687, 2018.
[38] J. Zou and H. Zhang, "High-frequency financial statistics through high-performance computing," in *Conquering Big Data with High Performance Computing*. Berlin, Germany: Springer, 2016, pp. 233–252.
[39] Y. Sun and Z. Lu, "Financial quantitative big data platform based on high performance computing," in *2019 IEEE International Conference on Computational Science and Engineering (CSE) and IEEE International Conference on Embedded and Ubiquitous Computing (EUC)*. IEEE, 2019, pp. 303–307.
[40] C. Li, Y. Wu, Z. Lu, and Y. Hu, "An intelligent parallel hybrid algorithm for multi-objective multi-period portfolio selection models with fuzzy random returns," in *2019 IEEE Intl Conf on Parallel and Distributed Processing with Applications, Big Data and Cloud Computing, Sustainable Computing and Communications, Social Computing and Networking (ISPA/BDCloud/SocialCom/SustainCom)*. IEEE, 2019, pp. 459–465.
[41] B. Liu, "A survey of credibility theory," *Fuzzy optimization and decision making*, vol. 5, no. 4, pp. 387–408, 2006.
[42] Y. K. Liu and J. Gao, "The independence of fuzzy variables with applications to fuzzy random optimization," *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, vol. 15, no. supp02, pp. 1–20, 2007.
[43] Z. Wang, L. Yang, L. Zhao, N. Cao, and Y. Lu, "A dual-objective vendor-managed inventory model for a single-vendor multi-retailer supply chain with fuzzy random demand," *Journal of Intelligent & Fuzzy Systems*, vol. 35, no. 1, pp. 211–222, 2018.
[44] Y. K. Liu and B. Liu, "A class of fuzzy random optimization: expected value models," *Information Sciences*, vol. 155, no. 1-2, pp. 89–102, 2003.
[45] S. Alexander, T. F. Coleman, and Y. Li, "Minimizing CVaR and VaR for a portfolio of derivatives," *Journal of Banking & Finance*, vol. 30, no. 2, pp. 583–605, 2006.
[46] H. Yao, Z. Li, and Y. Lai, "Mean–CVaR portfolio selection: A non-parametric estimation framework," *Computers & Operations Research*, vol. 40, no. 4, pp. 1014–1022, 2013.
[47] H. Konno and K. I. Suzuki, "A mean-variance-skewness portfolio optimization model," *Journal of the Operations Research Society of Japan*, vol. 38, no. 2, pp. 173–187, 1995.
[48] X. Li, Z. Qin, and S. Kar, "Mean-variance-skewness model for portfolio selection with fuzzy returns," *European Journal of Operational Research*, vol. 202, no. 1, pp. 239–247, 2010.
[49] H. Ke and J. Ma, "Modeling project time–cost trade-off in fuzzy random environment," *Applied Soft Computing*, vol. 19, pp. 80–85, 2014.
[50] G. Srivastava, C. N. S. V. Kumar, V. Kavitha, N. Parthiban, and R. Venkataraman, "Two-stage data encryption using chaotic neural networks," *Journal of Intelligent and Fuzzy Systems*, vol. 38, no. 3, pp. 2561–2568, 2020.
[51] H. Xu and G. Srivastava, "Automatic recognition algorithm of traffic signs based on convolution neural network," *Multimedia Tools and Applications*, vol. 79, pp. 11 551–11 565, 2020.
[52] N. Treadgold and T. Gedeon, "The sarprop algorithm, a simulated annealing enhancement to resilient back propagation," in *Proceedings International Panel Conference on Soft and Intelligent Computing*, 1996, pp. 293–298.
[53] E. Atashpaz-Gargari and C. Lucas, "Imperialist competitive algorithm: An algorithm for optimization inspired by imperialistic competition," in *2007 IEEE congress on evolutionary computation*. IEEE, 2007, pp. 4661–4667.
[54] X. S. Yang, *Nature-inspired metaheuristic algorithms*. Frome, U.K., Luniver press, 2008.
[55] N. Ghatasheh, H. Faris, I. Aljarah, and R. M. Al-Sayyed, "Optimizing software effort estimation models using firefly algorithm," *Journal of Software Engineering and Applications*, vol. 8, pp. 133–142, 2019.

Chen Li received the B.S. degree in network engineering from the North China Electric Power University, Baoding, China, in 2010. She is currently working toward the Ph.D. degree in computer software and theory with the Computer Network Information Center, Chinese Academy of Sciences, Beijing, China, and also with the University of Chinese Academy of Sciences, Beijing, China.

She is a Visiting Ph.D. Student with the College of Engineering, Mathematics and Physical Sciences, University of Exeter, Exeter, U.K. Her current research interests include high performance computing technology in the computational finance.



Yulei Wu (Senior Member, IEEE) received the B.Sc. degree (1st Class Hons.) in computer science and the Ph.D. degree in computing and mathematics from the University of Bradford, Bradford, U.K., in 2006 and 2010, respectively.

He is currently a Senior Lecturer with the Department of Computer Science, College of Engineering, Mathematics and Physical Sciences, University of Exeter, Exeter, U.K. His main research interests include computer networks, networked systems, software defined networks and systems, network management, and network security and privacy.

Zhonghua Lu received the B.S. and M.S. degrees from the Shaanxi Normal University, Xi'an, China, in 1986 and 1991, respectively, and the Ph.D. degree from the Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing, China, in 2002, all in mathematics.

She is a Professor with the Computer Network Information Center, Chinese Academy of Sciences. Her current research interests include high performance computing technology and its applications in computational finance.

Jue Wang received the B.S. degree in information and computation science from the North China Electric Power University, Baoding, China, in 2003, and the Ph.D. degree in computer science from the University of Science and Technology Beijing, Beijing, China, in 2009.

He is an Associate Professor with the Department of Artificial Intelligence Technology and Application Development, Computer Network Information Center (CNIC), Chinese Academy of Sciences, Beijing, China. He is Group Head of Computational Material and Associate Head of Youth Innovation Promotion Association, CNIC. His current research interests include parallel computing and artificial intelligence platform.

Yonghong Hu received the B.S. degree in mathematics and the M.S. degree in applied mathematics from the Shaanxi Normal University, Xi'an, China, in 1988 and 2002, respectively, and the Ph.D. degree in forestry economics and management from the Beijing Forestry University, Beijing, China, in 2008.

He is a Professor with the School of Statistics and Mathematics, Central University of Finance and Economics, Beijing. His current research interests include financial econometrics, synthetically evaluation method.