Mathematics and Systems Science 2024, 2(1), 2859.
https://doi.org/10.54517/mss.v2i1.2859

Article

A portfolio optimization model under uncertain random environment

Yanrui Su*, Yanjiao Song, Chenyi Liu

School of Mathematics and Statistics, Nanjing University of Science and Technology, Nanjing 210094, Jiangsu, China
*** Corresponding author:** Yanrui Su, yrsu@njust.edu.cn

CITATION

Su Y, Song Y, Liu C. A portfolio optimization model under uncertain random environment. Mathematics and Systems Science. 2024; 2(1): 2859.
https://doi.org/10.54517/mss.v2i1.2859

ARTICLE INFO

Received: 29 May 2024
Accepted: 22 June 2024
Available online: 30 June 2024

Abstract: Uncertain events frequently occur in today's financial markets. Consequently, the issue of portfolio selection is becoming increasingly significant. This paper thoroughly considers the complexities of stock returns in real-world scenarios and employs uncertain differential equations (UDE), uncertain time series analysis (UTSA), stochastic differential equations (SDE), and random time series analysis (RTSA) to predict stock returns, thereby enhancing the accuracy of these predictions. Furthermore, this paper addresses investors' preferences and the limitations of using variance as a measure of investment risk. It introduces a risk preference factor and proposes an uncertain random mean-lower variance model. Finally, a genetic algorithm is utilized to solve the model, and numerical simulations are conducted to demonstrate the model's practicality.

Keywords: uncertain random portfolio; uncertain differential equation; uncertain time series analysis; stochastic differential equation; random time series analysis; genetic algorithm

1. Introduction

The portfolio problem is concerned with determining the optimal decision among various stocks with a view to achieving a proper balance between two conflicting goals: return on investment and risk. In 1952, Markowitz [1] introduced the portfolio problem, presenting the classic mean-variance model, which laid the foundation for various models in modern finance. This model aims to minimize risks while ensuring certain returns or maximize returns within a specified risk tolerance. Subsequently, numerous scholars have studied portfolio optimization models. Jin and Zhang [2] extended the traditional mean variance model, considering the influence of high-order moments, and proposed a new method for robust portfolio optimization, which provides cutting-edge theoretical support for portfolio optimization. Yang and Li [3] combined stochastic dominance theory with portfolio optimization to propose new perspectives and methods, enriching the theoretical framework of portfolio optimization. Moreover, other researchers have further explored portfolio optimization, such as Konno and Suzuki [4], introducing the skewness of asset returns and emphasizing the crucial role of the third derivative of the utility function in optimal portfolio selection. Huang [5] proposed a new risk model based on the redefined concept of portfolio risk. Krejic et al. [6] incorporated fixed costs and impact costs as nonlinear functions of trading activities into the optimal portfolio model and studied the portfolio optimization problem of VaR risk measurement considering trading costs. Li and Shu [7] defined skewness and its calculation formula in uncertain stochastic environments and demonstrated the effectiveness and applicability of skewness through portfolio selection problems.

Variance has traditionally served as a fundamental tool for assessing risk in portfolio selection. However, this approach is subject to criticism for penalizing deviations from the mean, regardless of whether they are above or below the mean. Consequently, with the advancement of portfolio theory, scholars have introduced various alternative risk metrics, including Value at Risk (VaR) [6], skewness [7], and others. Despite these innovations, the distribution function often exhibits thick-tailed phenomena [8–10], where variance alone may fail to comprehensively capture the entirety of risk. In response to this challenge, numerous scholars have put forth high-order moment models [11,12] that incorporate asset returns in portfolio construction.

Traditional portfolio optimization models are primarily based on probability theory. However, the dynamics and complexities of financial markets render this approach inadequate for studying portfolio optimization by relying on probability theory alone. Consequently, many scholars have turned to fuzzy set theory [13] as a means of addressing the portfolio optimization problem. Deng and Li [14] introduced a fuzzy portfolio optimization model incorporating loan constraints, assuming asset returns as fuzzy numbers. Pahade and Jha [15] proposed a mean-variance-skewness model that utilizes trapezoid fuzzy variables to account for skewness. Kwakernaak introduced fuzzy-random variables based on fuzzy set theory to tackle mixed stock portfolio problems, regardless of historical data availability.

Subsequently, in 2007, Liu [16] introduced uncertainty theory to better manage subjective uncertainty, followed by a revision in 2010 [17]. With the advancement of uncertainty theory, research on uncertain portfolios has flourished. Yu and Wang [18] reviewed the application of uncertainty modeling in portfolio optimization and proposed various cutting-edge modeling methods and theories to help better manage and respond to uncertainty in investment decisions. Cai and Zhu [19] explored robust portfolio optimization under Conditional Value at Risk (CVaR) and proposed new methods to address the impact of uncertainty, particularly in risk management and capital allocation. Liu and Zhou [20] studied the uncertainty problem in dynamic portfolio optimization using stochastic programming methods, providing advanced dynamic optimization techniques suitable for complex market environments.

In the context of a complex and volatile financial landscape, uncertainty and randomness frequently coexist. Liu [21] integrated probability theory and uncertainty theory, introducing the concept of uncertain random variables and chance theory. Building upon the foundations of chance theory, Qin [22] proposed a mean-variance portfolio optimization model within an uncertain random environment. Additionally, risk measures for uncertain random variables were defined to gauge risk amidst uncertain random returns, leading to the establishment of portfolio optimization models tailored to uncertain random returns. Mehlawat et al. [23] conducted a detailed investigation into high-order moment portfolio optimization within an uncertain random environment. Treanja [24] introduced the optimization problem of interval-valued Kuhn-Tucker pseudoconvexity (KT pseudoinvix), introduced interval-valued curve integration and pseudoconvexity theory, and provided a new method to solve optimization problems under uncertainty and randomness. This paper aims to further explore uncertain random portfolios based on the groundwork laid by the research of aforementioned scholars.

In this study, a heuristic algorithm, specifically the genetic algorithm, is employed. The genetic algorithm (GA), initially proposed by Holland in 1975 [25], is distinguished as a randomized adaptive global search algorithm renowned for its formidable search capabilities. It addresses the issue of multiple individuals in the population simultaneously, thereby reducing the probability of becoming trapped in local optima by evaluating a multitude of solutions within the search space.

This article combines the unified objective method of GA and multi-objective optimization, which has a certain degree of innovation. Traditional multi-objective optimization problems involve multiple objective functions and usually require trade-offs and compromises between these objectives. The model in this article combines multiple objective functions into a single objective function, innovatively transforming complex multi-objective problems into single objective problems, enabling genetic algorithms to be applied to these problems. By optimizing a single objective function, the solution space can be explored more intensively, thereby improving the global optimality of the final solution. Meanwhile, in the dynamic environment of reality, the goals and constraints of the problem may constantly change. Our method enables genetic algorithms to quickly adapt to these changes and provide effective optimization solutions.

In assessing risk, the lower difference is employed in lieu of variance. In contrast to this, the traditional approach to variance considers all segments of the income distribution, encompassing both positive and negative returns. Nevertheless, in practical applications, greater concern is often directed towards the risk associated with negative returns, given their significant impact on investors. Hence, the concepts of lower difference and semi-variance are introduced.

The investment portfolio optimization model we study aims to allocate stock investments reasonably to maximize returns and minimize risks. In the future, these models may play important roles in multiple aspects. Firstly, with the widespread application of artificial intelligence and big data, models can provide personalized investment advice for intelligent investment advisory platforms, automatically adjusting investment portfolios based on users' risk preferences and market conditions [26]. Secondly, the model will be able to handle multiple asset classes (such as bonds, commodities, real estate, etc.), achieve asset allocation in global markets, and balance the risks and returns of different markets [27]. Finally, with the development of financial technology, models can combine techniques such as natural language processing and sentiment analysis to extract investment signals from unstructured data (such as news and social media sentiment) and make more comprehensive investment decisions [28].

The remainder of this paper is structured as follows: Section 2 presents a review of key definitions in uncertainty theory and random theory. Section 3 outlines four methodologies for forecasting stock return rates: uncertain differential equations (UDE), uncertain time series analysis (UTSA), stochastic differential equations (SDE), and random time series analysis (RTSA). Section 4 presents an uncertain stochastic stock model that incorporates dual objectives of mean-lower difference to evaluate the portfolio along with liquidity constraints and upper and lower limit constraints. Then the model is transformed by normalization. Section 5 outlines the development of the GA for deriving the optimal solution to the proposed model. Section 6 presents the results of numerical simulations that demonstrate the effectiveness of both the proposed model and GA. Finally, conclusions drawn from the findings are presented.

2. Preliminary

Uncertain methods are well-suited for addressing complex and highly uncertain environments. We can utilize UDE and UTSA to forecast return rates. The stochastic approach offers the advantage of capturing the dynamic behavior of the system by incorporating random variables, which can more accurately simulate market behavior. This approach can be applied across various fields, allowing us to use SDE and RTSA to predict return rates. This section introduces some fundamental concepts and formulas related to UDE, UTSA, SDE, RTSA, and chance theory.

2.1. Uncertainty theory

Uncertainty theory was established by Liu in 2007 and has since been explored by numerous researchers. Nowadays, uncertainty theory has evolved into a distinct branch of mathematics dedicated to the analysis of uncertain phenomena. This section will introduce α -path and inverse uncertainty distribution, which are integral components of the broader field of uncertainty theory.

Definition 1([29]). *An uncertain differential equation (UDE),*

$$\boldsymbol{dX_t = f(t, X_t)dt + g(t, X_t)dC_t} \tag{1}$$

which is defined to have an α -path X_t^α if it solves the corresponding ordinary differential equation,

$$dX_t^\alpha = f(t, X_t^\alpha)dt + |g(t, X_t^\alpha)|\Phi^{-1}(\alpha)dt, \tag{2}$$

where $\Phi^{-1}(\alpha)$ is the inverse standard normal uncertainty distribution, i.e.,

$$\Phi^{-1}(\alpha) = \frac{\sqrt{3}}{\pi}\ln\frac{\alpha}{1-\alpha}, \alpha \in (0,1). \tag{3}$$

Lemma 1([30]). *Let* X_t *and* X_t^α *be the solution and* α *-path of the uncertain differential equation respectively. Then,*

$$dX_t = f(t, X_t)dt + g(t, X_t)dC_t, \tag{4}$$

$$\widetilde{E}[X_t] = \int_0^1 X_t^\alpha d\alpha. \tag{5}$$

Lemma 2([30]). *It is clear that a Liu process* C_t *is a normal uncertain process with expected value 0 and variance* t^2*, i.e.,*

$$C_t \sim \mathrm{N}(0, t).$$

Furthermore, C_t *has an uncertainty distribution,*

$$\Phi_t(x) = \left(1 + exp\left(-\frac{\pi x}{\sqrt{3}t}\right)\right)^{-1}. \tag{6}$$

Lemma 3([30]). *For an uncertain variable* ξ *with regular uncertainty distribution*

$\Phi(x)$*, its expected value can be obtained by,*

$$\widetilde{\mathrm{E}}[\xi]=\int_0^1 \Phi^{-1}(\alpha)\mathrm{d}\alpha. \tag{7}$$

Lemma 4([30])**.** *For an uncertain variable* ξ *with regular uncertainty distribution* $\Phi(x)$ *and finite expected value* e*, its variance can be obtained by,*

$$V[\xi]=\int_0^1 \left(\Phi^{-1}(\alpha)-e\right)^2 \mathrm{d}\alpha. \tag{8}$$

Lemma 5([31])**.** *Let* ξ *be an uncertain variable that follows a normal uncertainty distribution* $\mathrm{N}\ (e,\sigma)$ *with unknown expected value* e *and unknown variance* σ^2*. Then the test for the hypotheses*

$H_0: e=e_0$ and $\sigma=\sigma_0$ versus $H_1: e\neq e_0$ or $\sigma\neq\sigma_0$

at significance level α *is*

$$W=\{(z_1,z_2,\cdots,z_n)\text{: there are more than } \alpha \text{ of indexes } i\text{'s with } 1\le i\le n \text{ such that } z_i<\Phi_0^{-1}\left(\frac{\alpha}{2}\right) \text{ or } z_i>\Phi_0^{-1}\left(1\text{-}\frac{\alpha}{2}\right)\Big\},$$

where Φ_0^{-1} is the inverse uncertainty distribution of $\mathrm{N}\ \left(e_0,\sigma_0\right)$, i.e.,

$$\Phi_0^{-1}(\alpha)=e_0+\frac{\sigma_0\sqrt{3}}{\pi}ln\frac{\alpha}{1-\alpha}.$$

Lemma 6([32])**.** *Phillips-Perron test (PP test) addresses the issues of heteroscedasticity and serial correlation in the interference term* ε_t *of a regression model. The three forms of the regression model are as follows:*

$$p_t=\beta p_{t-1}+\varepsilon_t. \tag{9}$$

$$p_t=\beta p_{t-1}+\mu+\varepsilon_t. \tag{10}$$

$$p_t=\beta p_{t-1}+\sum_{j=1}^{p-1}\phi_j^*\,\Delta p_{t-j}+\mu+\alpha t+\varepsilon_t. \tag{11}$$

The null hypothesis is $H_0{:}\,\beta=1$*, and the alternative hypothesis is* $H_1{:}\,\beta<1$*. When the PP test statistic exceeds the corresponding critical value, the original hypothesis that the time series is non-stationary cannot be rejected, indicating the presence of unit roots. Conversely, if the PP test statistic is less than or equal to the critical value, the time series is deemed stationary and is free of unit roots.*

2.2. Probability theory

The RTSA analyses historical data to predict future moments. SDE is often used to describe certain financial and economic phenomena, such as changes in stock prices. This section describes some of the basics of RTSA and SDE.

Definition 2([33])**.** *If we have a time series* x_t*, then for any lag* k*, the Auto-Correlation Function (ACF)* $\rho(k)$ *can be expressed as*

$$\rho(k)=\frac{Cov(x_t,x_{t+k})}{Var(x_t)}, \tag{12}$$

where $Cov(x_t,x_{t+k})$ is the covariance of the observations at time t and time $t+k$

and $Var(x_t)$ is the variance of the time series x_t.

Definition 3([33])**.** *If we have a time series* x_t*, then the Partial Auto-Correlation Function (PACF)* ϕ_{kk} *for any lag* k *can be expressed as*

$$\phi_{kk} = \frac{\hat{E}\left[\left(x_t - \hat{E}x_t\right)\left(x_{t-k} - \hat{E}x_{t-k}\right)\right]}{\hat{E}\left[\left(x_{t-k} - \hat{E}x_{t-k}\right)^2\right]}, \tag{13}$$

where $\hat{E}(x_t)$ denotes expectation, which is the mean, and the expected value of the forecast is referred to as $\hat{E}x_t$.

Definition 4([33])**.** *The Akaike information criterion (AIC) is a measure of the goodness of fit of a statistical model. It is based on the concept of entropy and can be used to assess the complexity of the estimated model and the quality of the data fitted by this model. The formula for AIC is as follows:*

$$AIC = 2k - 2\,ln(L), \tag{14}$$

where k represents the number of parameter estimates in the model and L denotes the maximum likelihood function of the model.

Definition 5 ([33])**.** *The Bayesian Information Criterion (BIC) is another model selection criterion that is very similar to the AIC, but is more stringent when dealing with model complexity. The BIC formula is as follows:*

$$BIC = k\,ln(n) - 2\,ln(L), \tag{15}$$

where n is the number of observations, k is the number of estimated parameters in the model, and L is the maximum log likelihood of the model fit.

Lemma 7([34])**.** *(*$It\hat{o}'s$ *formula) Suppose that* $X(\cdot)$ *has a stochastic differential*

$$dX = Fdt + GdW, \tag{16}$$

for$F \in \mathbb{L}^1(0,T)$, $G \in \mathbb{L}^2(0,T)$. *Assume* $u: \mathbb{R} \times [0,T] \to \mathbb{R}$ *is continuous and that* $\frac{\partial u}{\partial t}$, $\frac{\partial u}{\partial x}$, $\frac{\partial^2 u}{\partial x^2}$ *exist and are continuous. Let*

$$Y(t) := u(X(t), t). \tag{17}$$

Then Y *has the stochastic differential*

$$\begin{aligned} dY &= \frac{\partial u}{\partial t}dt + \frac{\partial u}{\partial x}dX + \frac{1}{2}\frac{\partial^2 u}{\partial x^2}G^2 dt \\ &= \left(\frac{\partial u}{\partial t} + \frac{\partial u}{\partial x}F + \frac{1}{2}\frac{\partial^2 u}{\partial x^2}G^2\right)dt + \frac{\partial u}{\partial x}GdW. \end{aligned} \tag{18}$$

The above formula is called $It\hat{o}'s$ *formula or* $It\hat{o}'s$ *chain rule.*

Lemma 8([33])**.** *The Augmented Dickey-Fuller (ADF) test is a statistical method used to determine whether a time series contains a unit root, which indicates non-stationarity, or if it is stationary. The null hypothesis posits that the series has a unit root (i.e., it is non-stationary), while the alternative hypothesis a that the series is stationary. The ADF test typically employs an autoregressive model with lagged terms to assess the stationarity of the series.*

In the case of a no-drift-no-trend term, the ADF test equation is as follows:

$$\Delta y_t = \rho y_{t-1} + \sum_{j=1}^{p-1} \phi_j^* \Delta y_{t-j} + u_t. \tag{19}$$

In the case of a with-drift-no-trend, the ADF test is as follows:

$$\Delta y_t = \rho y_{t-1} + \sum_{j=1}^{p-1} \phi_j^* \Delta y_{t-j} + \mu + u_t. \tag{20}$$

In the case of with-drift-with-trend, the ADF test is as follows:

$$\Delta y_t = \rho y_{t-1} + \sum_{j=1}^{p-1} \phi_j^* \Delta y_{t-j} + \mu + \alpha t + u_t. \tag{21}$$

In Equations (19–20), $\rho = \beta - 1 = (\sum_{j=1}^{p} \phi_i) - 1$, $\phi_j^* = -\sum_{i=j+1}^{p} \phi_i, j = 1, 2, \ldots p-1$. *The time series* y_t *is the independent variable, while* α *is a constant value. The coefficient* β *measures the trend's strength, and* ρ *represents the serial autoregressive coefficient.*

Lemma 9([33])**.** *The Ljung-Box (LB) test evaluates serial autocorrelation in time series analysis. The null hypothesis asserts that* y_t *is a white noise series, while the alternative suggests otherwise. The LB statistic is calculated as follows:*

$$\begin{aligned} \mathrm{Var}(\hat{\rho}_k) &= \frac{n-k}{n(n+2)}, \\ LB &= n(n+2) \sum_{k=1}^{m} \left(\frac{\hat{\rho}_k^2}{n-k} \right) \sim \chi^2(m), \end{aligned} \tag{22}$$

where n is the number of samples and $\hat{\rho}_k^2$ is the correlation coefficient at k th order lag; this statistic follows a chi-square distribution with m degrees of freedom. The null hypothesis is rejected at the significance level α if the p-value is less than or equal to α, indicating that the series is not a white noise. If the p-value exceeds α, the series is considered white noise.

2.3. Chance theory

In 2013, Liu introduced uncertain random variables to model complex systems characterized by uncertainty and randomness. This section will introduce chance measures, uncertain random variables, and chance distributions.

Definition 2([35])**.** *If the uncertain random variable* ξ *has a finite expected value, then its average lower deviation is*

$$\sigma_-^2 = E\left[\left(\xi - E(\xi)\right)^-\right], \tag{23}$$

among which $\left(\xi - E(\xi)\right)^- = max\{0, E(\xi) - \xi\}$.

Lemma 10([30])**.** *Assume that* $\eta_1, \eta_2, \cdots, \eta_m$ *are independent random variables with probability distributions* $\Psi_1, \Psi_2, \cdots, \Psi_m$, *and let* $\tau_1, \tau_2, \cdots, \tau_n$ *be independent uncertain variables with uncertainty distributions* $\Upsilon_1, \Upsilon_2, \cdots, \Upsilon_n$, *respectively. If* f *is* *a measurable function, then the uncertain random variable*

$$\xi = f(\eta_1, \eta_2, \cdots, \eta_m, \tau_1, \tau_2, \cdots, \tau_n) \tag{24}$$

has a chance distribution

$$\Phi(x) = \int_{\Re^m} F(x; y_1, y_2, \cdots, y_m) d\Psi_1(y_1) d\Psi_2(y_2) \cdots d\Psi_m(y_m), \tag{25}$$

where

$$F(x; y_1, y_2, \cdots, y_m) = \mathcal{M}\{f(y_1, y_2, \cdots, y_m, \tau_1, \tau_2, \cdots, \tau_n) \leq x\} \tag{26}$$

is the uncertainty distribution of $f(y_1, y_2, \cdots, y_m, \tau_1, \tau_2, \cdots, \tau_n)$ *for any real numbers* $y_1, y_2, \cdots, y_m$, *and is determined by* $\Upsilon_1, \Upsilon_2, \cdots, \Upsilon_n$.

3. Forecast stock return rates

In light of the available historical data related to stocks, this section outlines four methods for predicting stock returns: UDE, UTSA, SDE, and RTSA.

3.1. UDE predicts stock return rates

Assuming that the price of a stock at time t is X_t, the stock price is predicted by the geometric Liu process as follows:

$$dX_t = eX_t dt + \sigma X_t dC_t. \tag{27}$$

The analytic solution of Equation (27) is

$$X_t = X_0 \, exp(et + \sigma C_t). \tag{28}$$

Equation (27) has an α -path X_t^α as follows:

$$X_t^\alpha = X_0 \, exp(et + \sigma \Phi^{-1}(\alpha) t), \tag{29}$$

where $\alpha \in (0,1)$, e and σ are the log-drift and log-diffusion, respectively, $\Phi^{-1}(\alpha)$ is the inverse standard normal uncertainty distribution, and

$$\Phi^{-1}(\alpha) = \frac{\sqrt{3}}{\pi} ln\left(\frac{\alpha}{1-\alpha}\right). \tag{30}$$

According to **Lemma 1**, the expected price of stock is

$$\tilde{E}[X_t] = \int_0^1 X_t^\alpha \, d\alpha. \tag{31}$$

Let $\delta(t)$ represent the return rates at a future moment t, the simple return rates at moment t can be expressed as

$$\delta(t) = \frac{X_t - X_{t-1}}{X_{t-1}}, \tag{32}$$

The expected return rates on the stock at time t can be obtained from Equations (31) and (32) as

$$\widetilde{E}[\delta(t)] = \int_0^1 \frac{X_t^\alpha - X_{t-1}^\alpha}{X_{t-1}^\alpha} d\alpha$$
$$= \int_0^1 \left(\frac{\exp\left(e_1 t + \sigma_1 \Phi^{-1}(\alpha) t\right)}{\exp\left(e_2 (t-1) + \sigma_2 \Phi^{-1}(\alpha)(t-1)\right)} - 1 \right) d\alpha. \tag{33}$$

In Equation (33), e_1 and σ_1 are the parameters at time t, while e_2 and σ_2 are the parameters at time $t-1$.

Based on the definition of the Liu process, the term

$$\frac{C_{t_{i+1}} - C_{t_i}}{t_{i+1} - t_i} \sim \mathcal{N}(0,1). \tag{34}$$

Furthermore, X_t obeys a log-normal uncertain distribution as follows:

$$\ln X_t \sim \mathrm{N}\left(\ln X_0 + et, \sigma t\right). \tag{35}$$

Thus, the sample data constructed by variables e and σ should adhere to the standard normal distribution of uncertainty as follows:

$$\frac{X_{t_{i+1}} - X_{t_i} - eX_{t_i}\left(t_{i+1} - t_i\right)}{\sigma X_{t_i}\left(t_{i+1} - t_i\right)} \sim \mathrm{N}\left(0,1\right). \tag{36}$$

Assume that there are n stock price observations $x_{t_1}, x_{t_2}, \cdots, x_{t_n}$ at time $t_1, t_2, \cdots, t_n$ with $t_1 < t_2 < \cdots < t_n$, respectively. By substituting X_{t_i} and $X_{t_{i+1}}$ with the observations x_{t_i} and $x_{t_{i+1}}$, respectively, we can express the equation as

$$h_i(e,\sigma) = \frac{x_{t_{i+1}} - x_{t_i} - ex_{t_i}(t_{i+1} - t_i)}{\sigma x_{t_i}(t_{i+1} - t_i)}, i = 1,2,\cdots,n-1, \tag{37}$$

which are real functions of the parameters e and σ. The estimates of e and σ are denoted by e^* and σ^*, respectively. Using the method of moments [33], we can get the system of equations:

$$\begin{cases} \dfrac{1}{n-1} \displaystyle\sum_{i=1}^{n-1} h_i\left(e^*, \sigma^*\right) = 0, \\ \dfrac{1}{n-1} \displaystyle\sum_{i=1}^{n-1} \left(h_i(e^*, \sigma^*)\right)^2 = 1. \end{cases} \tag{38}$$

The solution to the system of Equation (38) is

$$\begin{cases} e^* = \dfrac{1}{n-1} \displaystyle\sum_{i=1}^{n-1} \frac{x_{t_{i+1}} - x_{t_i}}{x_{t_i}(t_{i+1} - t_i)}, \\ \sigma^* = \sqrt{\dfrac{1}{n-1} \displaystyle\sum_{i=1}^{n-1} \left(\frac{x_{t_{i+1}} - x_{t_i}}{x_{t_i}(t_{i+1} - t_i)}\right)^2 - \left(\frac{1}{n-1} \sum_{i=1}^{n-1} \frac{x_{t_{i+1}} - x_{t_i}}{x_{t_i}(t_{i+1} - t_i)}\right)^2}. \end{cases} \tag{39}$$

Using the definition of residuals, we generate $n-1$ residuals $\varepsilon_2, \varepsilon_3, \cdots, \varepsilon_n$ of the Equation (29) corresponding to the observed data. In order to ascertain whether the Equation (29) fits the observed data of stock price, it is necessary to test whether the linear uncertainty distribution $\mathcal{L}(0,1)$ fits the $n-1$ residuals $\varepsilon_2, \varepsilon_3, \cdots, \varepsilon_n$, i.e.,

$$\varepsilon_2, \varepsilon_3, \cdots, \varepsilon_n \sim \mathcal{L}(0,1).$$

Given a significance level α, it follows the test is

$$W = \left\{ (z_2, z_3, \cdots, z_n) : \text{ there are more than } \alpha \text{ of indexes } i\text{'s with } 2 \le i \le n \text{ such that } z_i < \frac{\alpha}{2} \text{ or } z_i > 1 - \frac{\alpha}{2} \right\}. \tag{40}$$

If the vector of the $n-1$ residuals $\varepsilon_2, \varepsilon_3, \cdots, \varepsilon_n$ belongs to the test W, i.e.,

$$(\varepsilon_2, \varepsilon_3, \cdots, \varepsilon_n) \in W,$$

then the Equation (29) is not a good fit to the observed data of stock price. If

$$(\varepsilon_2, \varepsilon_3, \cdots, \varepsilon_n) \notin W,$$

then the Equation (29) is a good fit to the observed data of stock price.

3.2. UTSA predicts stock return rates

An uncertain time series is a sequence of imprecisely observed values, each of which is characterized by an uncertain variable. Mathematically, an uncertain time series is represented by

$$X = \{X_1, X_2, \cdots, X_n\},$$

where X_t are imprecisely observed values (uncertain variables) at time t. A basic problem of UTSA is to predict the value of X_{n+1} based on previously observed values $X_1, X_2, \cdots, X_n$.

In order to model the time series, Yang and Liu [36] suggested an uncertain autoregressive model,

$$X_t = a_0 + \sum_{i=1}^{k} a_i X_{t-i} + \varepsilon_t, \tag{41}$$

where $a_0, a_1, \cdots, a_k$ are unknown parameters, k is called the order of the autoregressive model, and ε_t is a disturbance term.

The least squares estimate of $a_0, a_1, \cdots, a_k$ in the autoregressive model (41) is the solution of the minimization problem

$$\min_{a_0, a_1, \ldots, a_k} \sum_{t=k+1}^{n} \tilde{E}\left[\left(X_t - a_0 - \sum_{i=1}^{k} a_i X_{t-i} \right)^2 \right]. \tag{42}$$

Denoting the optimal solution by $a_0^*, a_1^*, \cdots, a_k^*$, the fitted autoregressive model is determined by

$$X_t = a_0^* + \sum_{i=1}^{k} a_i^* X_{t-i}. \tag{43}$$

Then, for each index $t\left(t = k+1, k+2, \cdots, n\right)$, the t th residual is

$$\varepsilon_t = X_t - a_0^* - \sum_{i=1}^{k} a_i^* X_{t-i} \tag{44}$$

Hence, a suitable approach to estimate the expected value of disturbance term is the average of expected values of residuals, i.e.,

$$\hat{e} := \frac{1}{n-k} \sum_{t=k+1}^{n} \tilde{E}\,[\varepsilon_t], \tag{45}$$

and the variance of the disturbance term can be estimated by

$$\hat{\sigma}^2 := \frac{1}{n-k} \sum_{t=k+1}^{n} \tilde{E}\,[(\varepsilon_t - \hat{e})^2], \tag{46}$$

where $\hat{\varepsilon}_t$ are the t th residuals, $t = k+1, k+2, \cdots, n$, respectively.

Based on the time series $X_1, X_2, \cdots, X_n$, it can infer that the uncertain autoregressive model is

$$X_t = a_0^* + \sum_{i=1}^{k} a_i^* X_{t-i} + \mathcal{N}(\hat{e}, \hat{\sigma}). \tag{47}$$

In order to test whether the UTSA model fits the observed data, we need to evaluate whether the normal uncertainty distribution N $(\hat{e}, \hat{\sigma})$ corresponds to the $n-k$ residuals $\varepsilon_{k+1}, \varepsilon_{k+2}, \cdots, \varepsilon_n$ determined by Equation (44), i.e.,

$$\varepsilon_{k+1}, \varepsilon_{k+2}, \cdots, \varepsilon_n \sim \mathcal{N}(\hat{e}, \hat{\sigma}). \tag{48}$$

Given a significance level α, it follows from **Lemma 5** that the test is

$$W = \{(z_{k+1}, z_{k+2}, \cdots, z_n)\text{: there are more than } \alpha \text{ of indexes } t\text{'s with } k+1 \le t \le n \text{ such that } z_t < \Phi^{-1}\left(\frac{\alpha}{2}\right) \text{ or } z_t > \Phi^{-1}\left(1\text{-}\frac{\alpha}{2}\right)\}$$

where

$$\Phi^{-1}(\alpha) = \hat{e} + \frac{\hat{\sigma}\sqrt{3}}{\pi} ln \frac{\alpha}{1-\alpha}. \tag{49}$$

If the vector of the $n-k$ residuals $\varepsilon_{k+1}, \varepsilon_{k+2}, \cdots, \varepsilon_n$ belongs to W, i.e.,

$$(\varepsilon_{k+1}, \varepsilon_{k+2}, \cdots, \varepsilon_n) \in W,$$

then reject the null hypothesis that means the model is not a good fit to the observed data. If

$$(\varepsilon_{k+1}, \varepsilon_{k+2}, \ldots, \varepsilon_n) \notin W,$$

then accept the null hypothesis that means the model is a good fit to the observed data.

The forecast's uncertain variable X_{n+1} follows a normal uncertainty distribution $\mathcal{N}(\hat{\mu}, \hat{\sigma})$, i.e.,

$$\Psi(z) = \left(1 + exp\left(\frac{\pi(\hat{\mu} - z)}{\sqrt{3}\hat{\sigma}}\right)\right)^{-1}. \tag{50}$$

Taking α as the confidence level, it is easy to verify that the minimum interval $[a, b]$ with $\hat{\Psi}(b) - \hat{\Psi}(a) \geq \alpha$ is

$$\left[\hat{\mu} - \frac{\hat{\sigma}\sqrt{3}}{\pi} ln\frac{1+\alpha}{1-\alpha}, \hat{\mu} + \frac{\hat{\sigma}\sqrt{3}}{\pi} ln\frac{1+\alpha}{1-\alpha}\right].$$

Since M $\left\{a \leq \hat{X}_{n+1} \leq b\right\} \geq \hat{\Psi}(b) - \hat{\Psi}(a) = \alpha$, the α confidence interval of X_{n+1} is

$$\hat{\mu} \pm \frac{\hat{\sigma}\sqrt{3}}{\pi} \ln\frac{1+\alpha}{1-\alpha}.$$

3.3. SDE predict stock return rates

Let S_t denote the price of a stock at time t, the following SDE can be obtained from the Black-Scholes model:

$$dS_t = \mu S_t dt + \sigma S_t dW_t, \tag{51}$$

where μ and σ represent the drift and volatility of a stock, respectively, and W_t represents the Standard Brownian Motion.

Given the initial price S_0 of a stock at time $t = 0$, according to $It\hat{o}'s$ formula, Equation (51) can be written in the following form, as follows:

$$\frac{dS_t}{S_t} = \mu dt + \sigma dW_t, \tag{52}$$

the solution of Equation (51) can be obtained as follows:

$$S_t = S_0\, exp\left(\left(\mu - \frac{\sigma^2}{2}\right)t + \sigma W_t\right). \tag{53}$$

Since Equation (51) implies

$$S_t = S_0 + \int_0^t \mu\, S_t ds + \int_0^t \sigma\, S_t dW_t \tag{54}$$

and $\hat{E}\left(\int_0^t \sigma\, S_t dW_t\right) = 0$, we have,

$$\hat{E}(S_t) = S_0\, exp(\mu t). \tag{55}$$

The simple return rates on the stock at time t is

$$\xi(t) = \frac{S_t - S_{t-1}}{S_{t-1}}, \tag{56}$$

and according to Equation (55), the expected return rates at time t is

$$\widehat{E}(\xi_t) = \frac{\widehat{E}(S_t) - \widehat{E}(S_{t-1})}{\widehat{E}(S_{t-1})} = \frac{\exp(\mu_1 t)}{\exp\left(\mu_2 (t-1)\right)} - 1. \tag{57}$$

In Equation (57), μ_1 is the parameter at t, while μ_2 is the parameter at $t-1$.

To predict the price of a stock at time t, it is necessary to estimate the two unknown parameters μ and σ in Equation (53). Since

$$ln\, S_t \sim \mathcal{N}\left(ln\, S_0 + \left(\mu - \frac{\sigma^2}{2}\right)t, \sigma^2 t\right), \tag{58}$$

it is easy to obtain the probability density function of S_t:

$$f(S_t; \mu, \sigma) = \frac{1}{S_t \sigma \sqrt{2\pi t}} exp\left[-\frac{\left[ln\left(\frac{S_t}{S_0}\right) - \left(\mu - \frac{\sigma^2}{2}\right)t\right]^2}{2\sigma^2 t}\right]. \tag{59}$$

The probability density function of S_t can be used to obtain the likelihood function

$$L(S_t; \mu, \sigma^2) = \prod_{t=0}^{n} \frac{1}{S_t \sigma \sqrt{2\pi t}} exp\left[-\frac{\left[ln\left(\frac{S_t}{S_0}\right) - \left(\mu - \frac{\sigma^2}{2}\right)t\right]^2}{2\sigma^2 t}\right], \tag{60}$$

and the maximum likelihood estimation method can be used to obtain the estimated values μ^* and σ^*, respectively.

To ensure the accuracy of the predicted stock return rates S_t at t, it is essential to conduct hypothesis testing on the estimated parameter values μ^* and σ^*. In this section, only the hypothesis testing steps for the obtained parameter value μ^* are provided. Similarly, the parameter value σ^* can be obtained.

The null hypothesis, denoted by H_0, and alternative hypothesis, denoted by H_1, should be presented as follows:

$$H_0: \mu = \mu^*, \quad \text{VS} \quad H_1: \mu \neq \mu^*.$$

Construct a test statistic Z that follows a standard normal distribution and assume that the original hypothesis is true, then

$$Z = \frac{ln\, S_t - \left(ln\, S_0 + \left(\mu^* - \frac{{\sigma^*}^2}{2} t\right)\right)}{\sigma^* \sqrt{t}} \sim \mathcal{N}(0,1). \tag{61}$$

Given the significance level α, the following equation can be obtained:

$$P\left(|Z| \geq Z_{\frac{\alpha}{2}}\right) = \alpha, \tag{62}$$

where $Z_{\frac{\alpha}{2}}$ represents the $\frac{\alpha}{2}$ supremum of Z, and the rejection domain can be further determined as

$$|Z| \geq Z_{\frac{\alpha}{2}}. \tag{63}$$

Bring historical stock return rate data into formula (61) to obtain the specific value of Z. In the event that the value falls within the rejection domain, the original null hypothesis must be rejected; otherwise, it must be accepted. Only when the historical data of a certain stock passes hypothesis testing can this method be employed to predict its future return rates.

3.4. RTSA predict stock return rates

The ARIMA model is a statistical model for time series analysis and forecasting. It is based on the assumption that there exists a linear relationship between the current observations and the past observations, which can be used to describe the autocorrelation of the time series data, denoted as AR (p). The MA model treats a time series as a past noise of a number of periods as a weighted average. The current observation is obtained from the past white noise through a specific linear combination, denoted as MA (q). The ARMA (p, q) model is a combination of AR (p) and MA (q) with the order denoting the order of the AR and MA parts, respectively. The ARIMA (p, d, q) model introduces a difference operation on the basis of ARMA (p, q), which is used to handle non-stationary time series (seasonality). The difference operation enables the ARIMA model to transform a non-stationary time series into a stationary time series, which can then be modeled using the ARMA model. The mathematical formulation is as follows:

$$\begin{cases} Y_t = d + \sum_{i=1}^{p} \phi_i Y_{t-i} + \varepsilon_t - \sum_{j=1}^{q} \theta_j \varepsilon_{t-j}, \\ \phi_p \neq 0, \theta_q \neq 0, \\ \hat{E}(\varepsilon_t) = 0, Var(\varepsilon_t) = \sigma_\varepsilon^2, \hat{E}(\varepsilon_t \varepsilon_s) = 0, \forall s \neq t, \\ \hat{E}(x_s \varepsilon_t) = 0, s < t, \end{cases} \tag{64}$$

where Y_t is the return rates at time t, ε_t represents white noise at time t, $\mu, \phi_1, \phi_2, \cdots, \phi_p$ are AR coefficients, and $\theta_1, \theta_2, \cdots, \theta_q$ are MA coefficients.

The following outlines the steps involved in the construction of an ARIMA model [33]:

Firstly, the ADF test is conducted in accordance with **Lemma 8**, with the objective of determining the degree of smoothness exhibited by the data. In the event that the data is found to be non-stationary, a differencing operation is then performed on it until the resulting differenced data is deemed to be smooth. In the case of economic time series, the number of differences, d, is typically limited to 0, 1, or 2.

Subsequently, the white noise test is conducted in accordance with the stipulations of **Lemma 9**. In the event that the sequence is identified as a non-white noise sequence, it is then ready for subsequent prediction.

The ACF and PACF are computed for smooth non-white noise sequences in order to identify the ARMA model, as demonstrated in **Table 1**. After the smoothing process, the partial autocorrelation function is truncated while the autocorrelation function is trailing, thereby enabling the construction of an AR model. Conversely, if the partial autocorrelation function is trailing while the autocorrelation function is truncated, an MA model can be built. Finally, if both the partial autocorrelation function and the autocorrelation function are trailing, the sequence is suitable for the ARIMA model. The BIC criterion can be employed to order the model and determine the p and q parameters, thereby enabling the identification of the optimal model.

Then, the identified model must be tested to ascertain whether its residual sequence is white noise. If it is not white noise, it indicates that there is still useful information in the residuals, which must be modified in the model or further extracted. Once the models have been identified, the parameters of each model are determined. Finally, the following formula is applied to predict the results.

$$\hat{E}[Y_t] = d + \sum_{i=1}^{p} \phi_i Y_{t-i}. \tag{65}$$

Table 1. Model ordering.

Model	**ACF**	**PACF**
AR (p)	tail off	p-order cut-off
MA (q)	q-order cut-off	tail off
ARMA (p, q)	tail off	cut-off [1]

4. Uncertain random portfolio model

In the considering of n stocks, let n_1 represent the number of stock prices that satisfy the UDE, n_2 denote the stock return rates that satisfy the UTSA, n_3 indicate the stock prices that satisfy the SDE, and n_4 signify the stock return rates that satisfy the RTSA. It follows that $n = n_1 + n_2 + n_3 + n_4$. Simulate the dynamics of stock return rates with UTSA, RTSA. The UDE and SDE simulate the stock price and, consequently, the stock return rates. For sake of argument, let's assume that the return rates on all stocks are independent. To facilitate the narration, the following symbols are introduced as shown in **Table 2**.

Table 2. Symbol specification.

Symbol	**Symbol specification**
x_i	The investment proportion of i th stock with UDE
y_i	The investment proportion of j th stock with UTSA
p_k	The investment proportion of k th stock with SDE
q_m	The investment proportion of m th stock with RTSA
δ_i	The return rates of i th stock with UDE

Table 2. (*Continued*).

Symbol	**Symbol specification**
ζ_j	The return rates of j th stock with UTSA
ξ_k	The return rates of k th stock with SDE
η_m	The return rates of m th stock with RTSA
h	The investor's target return rates
R	The total return rates on the portfolio
V^-	Lower variance of the portfolio
a	Lower turnover rate
b	An investor's maximum risk tolerance
c	Maximum investment proportion for each stock

For the sake of expression, we have

$$\mathbf{x} = \left(x_1, x_2, \cdots, x_{n_1}\right)^T, \boldsymbol{\delta} = \left(\delta_1, \delta_2, \cdots, \delta_{n_1}\right)^T,$$
$$\mathbf{y} = \left(y_1, y_2, \cdots, y_{n_2}\right)^T, \boldsymbol{\zeta} = \left(\zeta_1, \zeta_2, \cdots, \zeta_{n_2}\right)^T,$$
$$\mathbf{p} = \left(p_1, p_2, \cdots, p_{n_3}\right)^T, \boldsymbol{\xi} = \left(\xi_1, \xi_2, \cdots, \xi_{n_3}\right)^T,$$
$$\mathbf{q} = \left(q_1, q_2, \cdots, q_{n_4}\right)^T, \boldsymbol{\eta} = \left(\eta_1, \eta_2, \cdots, \eta_{n_4}\right)^T.$$

Obviously, $\mathbf{x}^{\mathrm{T}}\boldsymbol{\delta}$, $\mathbf{y}^{\mathrm{T}}\boldsymbol{\zeta}$, $\mathbf{p}^{\mathrm{T}}\boldsymbol{\xi}$ and $\mathbf{q}^{\mathrm{T}}\boldsymbol{\eta}$ are portfolio return rates satisfying UDE, UTSA, SDE, and RTSA, respectively. Then the expected total return rate is

$$\begin{aligned}
R &= E\left[\mathbf{x}^{\mathrm{T}}\boldsymbol{\delta} + \mathbf{y}^{\mathrm{T}}\boldsymbol{\zeta} + \mathbf{p}^{\mathrm{T}}\boldsymbol{\xi} + \mathbf{q}^{\mathrm{T}}\boldsymbol{\eta}\right] \\
&= E\left[x_1\delta_1 + x_2\delta_2 + \cdots + x_{n_1}\delta_{n_1}\right] + E\left[y_1\zeta_1 + y_2\zeta_2 + \cdots + y_{n_2}\zeta_{n_2}\right] \\
&\quad + E\left[p_1\xi_1 + p_2\xi_2 + \cdots + p_{n_3}\xi_{n_3}\right] + E\left[q_1\eta_1 + q_2\eta_2 + \cdots + q_{n_4}\eta_{n_4}\right] \\
&= x_1\widetilde{E}\left[\delta_1\right] + x_2\widetilde{E}\left[\delta_2\right] + \cdots + x_{n_1}\widetilde{E}\left[\delta_{n_1}\right] \\
&\quad + y_1\widetilde{E}\left[\zeta_1\right] + y_2\widetilde{E}\left[\zeta_2\right] + \cdots + y_{n_2}\widetilde{E}\left[\zeta_{n_2}\right] \\
&\quad + p_1\widehat{E}\left[\xi_1\right] + p_2\widehat{E}\left[\xi_2\right] + \cdots + p_{n_3}\widehat{E}\left[\xi_{n_3}\right] \\
&\quad + q_1\widehat{E}\left[\eta_1\right] + q_2\widehat{E}\left[\eta_2\right] + \cdots + q_{n_4}\widehat{E}\left[\eta_{n_4}\right] \\
&= \mathbf{x}^{\mathrm{T}}\boldsymbol{\Delta} + \mathbf{y}^{\mathrm{T}}\mathbf{Z} + \mathbf{p}^{\mathrm{T}}\boldsymbol{\Xi} + \mathbf{q}^{\mathrm{T}}\mathbf{N},
\end{aligned} \tag{66}$$

where

$$\boldsymbol{\Delta} = \left(\tilde{E}[\delta_1], \tilde{E}[\delta_2], \cdots, \tilde{E}[\delta_{n_1}]\right)^T, \boldsymbol{Z} = \left(\tilde{E}[\zeta_1], \tilde{E}[\zeta_2], \cdots, \tilde{E}[\zeta_{n_2}]\right)^T,$$
$$\boldsymbol{\Xi} = \left(\hat{E}[\xi_1], \hat{E}[\xi_2], \cdots, \hat{E}[\xi_{n_3}]\right)^T, \boldsymbol{N} = \left(\hat{E}[\eta_1], \hat{E}[\eta_2], \cdots, \hat{E}[\eta_{n_4}]\right)^T.$$

Several drawbacks exist in utilizing variance as a measure of investment risk. Firstly, risk is quantified by the variance of the anticipated portfolio return rates, encompassing return rates exceeding the target rates as risk. However, investors typically view higher return rates positively, rendering this inclusion unnecessary. Secondly, the reliance on variance to gauge risk presupposes a normal distribution for the probability distribution of stock return rates, which is unrealistic in actual stock markets. In addition, another assumption that underlies the use of variance to evaluate risk is that investors are risk-averse and possess ample information. However, K.J. Arrow and J.W. Pratt have shown that risk aversion does not universally represent risk preferences [36,37].

In view of the aforementioned limitations of employing variance as a risk measurement tool, this paper adopts the lower variance to measure risk, which can better symbolize an investor's intuitive feeling of risk and definition of risk. For an uncertain random variable X, define the lower variance X^- as follows:

$$X^- = \begin{cases} X, X \le 0, \\ 0, X > 0. \end{cases} \tag{67}$$

Considering the case of $R \le h$, the lower variance of portfolio can be expressed as

$$V^- = E\left[\left(R-h\right)^-\right]^2 = \int_0^1 \left(\Psi^{-1}\left(\alpha\right) - h\right)^2 d\alpha, \tag{68}$$

where Ψ is the chance distribution of R, and we have

$$\Psi = \sum_{i=1}^{n_1} x_i \widetilde{E}\left[\delta_i\left(t\right)\right] + \sum_{j=1}^{n_2} y_j \widetilde{E}\left[\zeta_j\left(t\right)\right] + \sum_{k=1}^{n_3} p_k \widehat{E}\left[\xi_k\left(t\right)\right] + \sum_{m=1}^{n_4} q_m \widehat{E}\left[\eta_m\left(t\right)\right]. \tag{69}$$

In Equation (69), $\widetilde{E}\left[\delta_i\left(t\right)\right]$, $\widetilde{E}\left[\zeta_j\left(t\right)\right]$, $\widehat{E}\left[\xi_k\left(t\right)\right]$, and $\widehat{E}\left[\eta_m\left(t\right)\right]$ indicate that the return rates satisfy the expected value of UDE, UTSA, SDE, and RTSA, respectively. According to Equations (33), (43), (57), and (65) the following equations can be obtained, respectively.

$$\begin{aligned}
&\widetilde{E}\left[\delta_i\left(t\right)\right] = \int_0^1 \left(\frac{\exp\left(e_1 t + \sigma_1 \Phi^{-1}\left(\alpha\right) t\right)}{\exp\left(e_2 (t-1) + \sigma_2 \Phi^{-1}\left(\alpha\right)(t-1)\right)} - 1 \right) d\alpha, \quad i = 1, 2, \cdots, n_1, \\
&\widetilde{E}\left[\zeta_j\left(t\right)\right] = a_0^* + \sum_{i=1}^{g} a_i^* X_{t-i}, \quad j = 1, 2, \cdots, n_2, \\
&\widehat{E}\left[\xi_k\left(t\right)\right] = \frac{\exp(\mu_1 t)}{\exp\left(\mu_2 (t-1)\right)} - 1, \quad k = 1, 2, \cdots, n_3, \\
&\widehat{E}\left[\eta_m\left(t\right)\right] = d + \sum_{i=1}^{p} \phi_i Y_{t-i}, \quad m = 1, 2, \cdots, n_4.
\end{aligned} \tag{70}$$

In order to align the model more closely with the actual dynamics of the stock market, we incorporate liquidity constraints, upper and lower limit constraints, expected return rate constraints, and risk constraints into this uncertain random mean-lower variance model. The liquidity risk associated with stocks pertains to the potential inability to trade them in a timely and cost-effective manner due to inadequate market liquidity. In general, a higher level of liquidity is associated with a lower level of relative risk. Therefore, the liquidity of the stock represents a primary indicator of interest to indicators in the context of actual investment. In this paper, stock liquidity is described by turnover rate, which refers to the frequency of stock turnover in the market within a certain period of time. This is one of the indicators reflecting the strength of stock liquidity, which can be expressed by the following formula:

$$\sum_{i=1}^{n_1}\tau_i x_i+\sum_{j=1}^{n_2}\chi_j y_j+\sum_{k=1}^{n_3}\lambda_k p_k+\sum_{m=1}^{n_4}\gamma_m q_m \ge a, \tag{71}$$

where τ_i, λ_k denote the turnover rate of the i th, k th, stock whose price meets UDE, SDE, respectively and χ_j, γ_m denote the turnover rate of the j th, m th stock whose return rates meet UTSA, RTSA, respectively.

The upper and lower limits mean that the proportion of investment made by an investor in any stock can neither be negative nor exceed c, which can be expressed as

$$\begin{aligned}&0\le x_i\le c,\quad i=1,2,\cdots,n_1,\\&0\le y_j\le c,\quad j=1,2,\cdots,n_2,\\&0\le p_k\le c,\quad k=1,2,\cdots,n_3,\\&0\le q_m\le c,\quad m=1,2,\cdots,n_4,\end{aligned} \tag{72}$$

where c is the maximum value of the investor's proportion of investment in each stock.

The objective of investors is to obtain higher investment return rates while bearing lower investment risks. This can be expressed as the following equation:

$$\max \quad R=E\left[\mathbf{x}^{\mathrm{T}}\boldsymbol{\delta}+\mathbf{y}^{\mathrm{T}}\boldsymbol{\zeta}+\mathbf{p}^{\mathrm{T}}\boldsymbol{\xi}+\mathbf{q}^{\mathrm{T}}\boldsymbol{\eta}\right] \tag{73}$$

$$\min \quad V^{-}=E\left[\left(R-h\right)^{-}\right]^{2} \tag{74}$$

Additionally, it should be noted that different investors have varying degrees of acceptance of investment risks. Consequently, a risk preference factor ι is introduced. In accordance with Equations (66)–(72) and the objective functions (73) and (74). Model (75) can be incorporated into the following equation:

$$\begin{cases} \min \quad M = (1-\iota)\dfrac{R_{max} - R}{R_{max}} + \iota\dfrac{V^{-} - V_{min}^{-}}{V_{min}^{-}} \\ \text{s.t.} \quad R \geq h, \\ \qquad V^{-} \leq b, \\ \qquad \sum_{i=1}^{n_1} \tau_i x_i + \sum_{j=1}^{n_2} \chi_j y_j + \sum_{k=1}^{n_3} \lambda_k p_k + \sum_{m=1}^{n_4} \gamma_m q_m \geq a, \\ \qquad \sum_{i=1}^{n_1} x_i + \sum_{j=1}^{n_2} y_j + \sum_{k=1}^{n_3} p_k + \sum_{m=1}^{n_4} q_m = 1, \\ \qquad 0 \leq x_i, y_j, p_k, q_m \leq c, \\ \qquad i = 1, 2, \cdots, n_1, j = 1, 2, \cdots, n_2, \\ \qquad k = 1, 2, \cdots, n_3, m = 1, 2, \cdots, n_4. \end{cases} \tag{75}$$

In problem (75), R_{max} denotes the maximum portfolio return rates when investment risk is not considered, and V_{min}^{-} denotes the minimum portfolio risk when portfolio return rates are not considered. The risk preference factor, denoted by ι, is a key in this context. The greater the value of ι, the greater the weight of the variance after normalization of problem (75), which may be interpreted as indicating a greater risk aversion on the part of the investor.

In order to solve problem (75), the values of R_{max} and V_{min}^{-} need to be obtained under the same constraints as in problem (75). Therefore, this paper puts forth two sub-models, with R and V^{-} serving as objective functions, respectively. The sub-models are formulated as follows:

$$\begin{cases} \max \quad R = E\left[\mathbf{x}^{\mathrm{T}}\boldsymbol{\delta} + \mathbf{y}^{\mathrm{T}}\boldsymbol{\zeta} + \mathbf{p}^{\mathrm{T}}\boldsymbol{\xi} + \mathbf{q}^{\mathrm{T}}\boldsymbol{\eta}\right] \\ \text{s.t.} \quad R \geq h, \\ \qquad V^{-} \leq b, \\ \qquad \sum_{i=1}^{n_1} \tau_i x_i + \sum_{j=1}^{n_2} \chi_j y_j + \sum_{k=1}^{n_3} \lambda_k p_k + \sum_{m=1}^{n_4} \gamma_m q_m \geq a, \\ \qquad \sum_{i=1}^{n_1} x_i + \sum_{j=1}^{n_2} y_j + \sum_{k=1}^{n_3} p_k + \sum_{m=1}^{n_4} q_m = 1, \\ \qquad 0 \leq x_i, y_j, p_k, q_m \leq c, \\ \qquad i = 1, 2, \cdots, n_1, j = 1, 2, \cdots, n_2, \\ \qquad k = 1, 2, \cdots, n_3, m = 1, 2, \cdots, n_4. \end{cases} \tag{76}$$

and

$$
\begin{cases}
\min \quad V^- = E\left[\left(R-h\right)^-\right]^2 \\
\text{s.t.} \quad R \ge h, \\
\qquad V^- \le b, \\
\qquad \sum_{i=1}^{n_1} \tau_i x_i + \sum_{j=1}^{n_2} \chi_j y_j + \sum_{k=1}^{n_3} \lambda_k p_k + \sum_{m=1}^{n_4} \gamma_m q_m \ge a, \\
\qquad \sum_{i=1}^{n_1} x_i + \sum_{j=1}^{n_2} y_j + \sum_{k=1}^{n_3} p_k + \sum_{m=1}^{n_4} q_m = 1, \\
\qquad 0 \le x_i, y_j, p_k, q_m \le c, \\
\qquad i = 1,2,\cdots,n_1, j = 1,2,\cdots,n_2, \\
\qquad k = 1,2,\cdots,n_3, m = 1,2,\cdots,n_4.
\end{cases}
\tag{77}
$$

5. Problem solving

GA is an optimization method inspired by biological evolution and is typically employed to address complex optimization problems [25]. The algorithm simulates the genetic and evolutionary processes observed in nature, with the objective of identifying the optimal or near-optimal solution to a given problem. This is achieved by simulating the natural selection mechanism, in which the fittest survive and reproduce.

The five principal elements involved in the GA are as follows: parameter coding, setting of the initial population, design of the fitness function, design of genetic operation, and setting of the control parameters. The operation process of the GA is characterized by a typical iterative process, comprising the following essential elements and basic steps:

The GA typically employs a binary coding method, whereby the decision variables are represented as binary strings. The length of the binary coding string, denoted by L, is dependent on the required precision. The binary coding strings for each decision variable are then concatenated to form a chromosome. The range of intervals is represented by the closed interval $[H_r, H_l]$, the length of the encoding is denoted by δ, and the current string $O_i = s_1 s_2 \cdots s_L$ corresponds to the decimal H_i. The integer part of the encoding represents the quotient when dividing by two, while the fractional part represents the product when multiplying by two. The precision of this representation is denoted by Equation (78). The corresponding decoding equation is presented in Equation (79).

$$
\delta = \frac{H_r - H_l}{2^L - 1}. \tag{78}
$$

$$
H_i = H_r + \sum_{i=1}^{L} s_i \cdot \frac{H_l - H_r}{2^L - 1}. \tag{79}
$$

Following the initialization of the population or subsequent perturbations, it is necessary to assess the adaptation of individuals. The assessment of adaptation is based on the objective function pursued. In this context, it is possible to consider the objective function to be solved as its hardness function. In the event that the objective function is a maximum problem, then

$$F(O_i) = f(X_i). \tag{80}$$

Conversely, in the event that the objective function is a minimization problem, then

$$F(O_i) = W - f(X_i), \tag{81}$$

where W is a constant we introduced to satisfy the non-negative value of the fitness function.

The evaluation function corresponding to each chromosome is then defined as the ratio of the fitness of a particular chromosome to the sum of the fitness values of all chromosomes. The evaluation function of an individual O_i , $Eval(O_i)$, is shown in Equation (82).

$$Eval(O_i) = \frac{F(O_i)}{\sum_{i=1}^{N} F(O_i)}. \tag{82}$$

Subsequently, the selection operation is performed using the roulette wheel method. This entails that the probability of each individual entering the next generation is equal to the ratio of its fitness value to the sum of the fitness values of the individuals in the entire population. The formula is presented in Equation (83). A random variable r is generated in the interval $[0,1]$, and the individual variable O_k is selected if $P_{k-1} < r \leq P_k$.

$$P_k = \frac{Eval(O_k)}{\sum_{i=1}^{k} Eval(O_i)}. \tag{83}$$

In the above process, two parent individuals, designated as O_1 and O_2, are selected from the current population. The crossover point is then set at a specific position, designated as M_1, at which two offspring individuals, designated as O_1' and O_2', can be generated. The following equation can be used to calculate the values of O_1' and O_2':

$$O_1' = O_1[1:M_1] + O_2[M_1:],$$
$$O_2' = O_2[1:M_1] + O_1[M_1:],$$

where $O_1[1{:}M_1]$ represents the gene segment of the parent O_1 before the crossover point, while $O_2[M_1{:}]$ represents the gene segment of the parent O_2 after the crossover point.

The individual O_i to be mutated can be determined by roulette selection. Once a specific mutation site M_2 has been selected, the mutated individual O_i'' can be generated by

$$O_i'' = O_i.$$

The mutation operation of the gene at the specified point is then performed, for example, by setting $O_i^{''}[M_2] = G$. The new gene value G can be either a randomly chosen one or one generated according to a specific rule.

The complete algorithmic process is illustrated in Algorithm 1.

Algorithm 1 Solution Algorithms for Optimization

Step 1. Encoding.
Step 2. Generate random initialization of N chromosomes.
Step 3. Calculate target values for all chromosomes.
Step 4. Calculate the fitness of each chromosome based on the target value.
Step 5. Selection of chromosomes by means of roulette.
Step 6. Update chromosomes with crossover algorithms.
Step 7. Update chromosomes with mutation algorithms.
Step 8. Repeat steps 3 to 7 until the termination condition is satisfied.
Step 9. Output the best chromosome as the optimal solution.

6. Numerical simulation

In this section, a numerical simulation is designed to predict daily stock return rates and the optimal portfolio. In order to illustrate the models proposed in this paper, we selected some stocks from the Shenzhen Stock Exchange and the Shanghai Stock Exchange. The daily return rates and prices from 4 May 2023 to 30 April 2024 are collected for each stock, resulting in a total of 242 data points. Finally, each of the two stock exchanges has four stocks that meet the four predictive models discussed in Section 3, with two stocks meeting each of the UDE, UTSA, SDE, and RTSA models ($n_1 + n_2 + n_3 + n_4 = 8$). All the data on daily stock return rates and prices used in this paper are obtained from the CHOICE database (https://choice.eastmoney.com/dataservice).

The UDE equation is used to predict the price of stock 300528.SZ. Firstly, the parameters $e = -0.0871$ and $\sigma = 11.6035$ are obtained by using the method of moments estimation, and the UDE obtained after parameter estimation is

$$dX_t = eX_t dt + \sigma X_t dC_t \tag{84}$$

$$dX_t = 0.0871 X_t dt + 11.6035 X_t dC_t. \tag{85}$$

Subsequently, a hypothesis test is performed on the residuals, i.e., whether the residuals conform to a linear uncertain distribution. From Equation (40), the number of points belonging to the rejection domain is 12, which is less than $n * \alpha = 242 * 0.05 = 12.1$. This shows that Equation (85) can fit the stock price well. Based on Equation (86), the stock price on 6 May 2024 is obtained, and then based on the stock price on 30 April 2024, the daily return rate on 6 May 2024 is calculated using the formula (32) for the simple return rates, resulting in a daily return rate of 1.71%.

$$\tilde{E}[X_1(t)] = \int_0^1 1\, 0.94\, exp(-0.087t + 11.6035\Phi^{-1}(\alpha)t)\, d\alpha, \tag{86}$$

The prediction process for stock 000651.SZ is analogous to that of 300528.SZ.

The parameter $e = 0.1975$ and the standard deviation $\sigma = 3.3732$ are used to construct the corresponding UDE, which is given by

$$dX_t = 0.1975X_t dt + 3.3732X_t dC_t. \tag{87}$$

The hypothesis testing of the residuals indicates that the number of points falling in the rejection domain is 12, which is less than the number of points required to reject the null hypothesis,$n * \alpha = 242 * 0.05 = 12.1$. This suggests that the equation can predict the stock price well. Based on Equation (88), the stock price on 6 May 2024 is calculated, and then based on the simple return rates, the return rate on 6 May 2024 is obtained as 0.35%.

$$\widetilde{\boldsymbol{E}}[\boldsymbol{X_2(t)}] = \int_0^1 3\,5.42\,\boldsymbol{exp}\left(\boldsymbol{0.1975t + 3.3732\Phi^{-1}(\alpha)t}\right) \boldsymbol{d\alpha}, \tag{88}$$

The AR model is employed to forecast the daily return rates of stock 600886. SH. The LB test on the residuals initially indicates that the series is a white noise series. The *p*-value of the PP test is 0.0375, which is less than 0.05, indicating that the series is a smooth series. Subsequently, a fixed-order AR model based on the AIC criterion is employed to fit the series, thereby obtaining an AR (2) model. The resulting coefficients are as follows: $a_0 = 0$, $\epsilon_t = -0.0011$, $a_1 = 0.7828$, $a_2 = 0.4273$. The original Equation (41) for the uncertain time series fit is as follows:

$$X_t\text{=}0.7828X_{t-1}\text{+}0.4273X_{t-2}\text{-}0.0011. \tag{89}$$

Then the prediction equation is obtained by least square estimation method:

$$\tilde{E}[\zeta_1(t)] = 0.7251X_{t-1} + 0.2748X_{t-2}, \tag{90}$$

where $a_0^* = 0$, $a_1^* = 0.7251$, and $a_2^* = 0.2748$. Using the aforementioned prediction Equation (90), the daily return rate on 6 May 2024 is calculated to be 1.7095%.

The prediction process for stock 600674.SH is analogous to that of 600886.SH. The residuals are initially subjected to the LB test, which indicates that the series is a white noise series. The *p*-value of the PP test is 0.0272, which is below the 0.05 level of significance, indicating that the series is a smooth series. The series is fitted using the AR model, according to the fixed order AIC criterion, in order to obtain the AR (2) model. The resulting coefficients are as follows: $a_0 = 0$, $\grave{o}_t = -0.0021$, $a_1 = 0.7832$ and $a_2 = 0.3539$. The original equation is fitted to the uncertain time series, resulting in the following equation:

$$X_t\text{=}0.7832X_{t-1}\text{+}0.3539X_{t-2}\text{-}0.0021. \tag{91}$$

Then the prediction equation is obtained by least square estimation method:

$$\tilde{E}\left[\zeta_1(t)\right] = 0.7255X_{t-1} + 0.2744X_{t-2}, \tag{92}$$

where $a_0^* = 0$, $a_1^* = 0.7255$ and $a_2^* = 0.2744$. Through Equation (92), the daily return rate of the stock on 6 May 2024 is calculated to be 3.3364%.

The SDE model is used to predict the price of stock 000786.SZ and thus predict its daily return rates. Firstly, the parameters $\mu = 2.3212$ and $\sigma = 1.4724$ are obtained by maximum likelihood estimation to obtain the SDE after parameter estimation:

$$dS_t = \mu S_t dt + \sigma S_t dW_t \tag{93}$$

$$dS_t = 2.3212 S_t dt + 1.4724 S_t dW_t. \tag{94}$$

Hypothesis testing is conducted on the parameters μ and σ using Matlab's own function chi2gof, resulting in a *p*-value of 0.0187, which exceeds the significance level of 0.01. The stock price on 6 May 2024 is firstly obtained from formula

$$\hat{E}[S_1(t)] = 26.34\, exp(2.3212t) \tag{95}$$

and then the formula (56) is used to obtain the daily return rates on 6 May 2024 as 2.35%.

The stocks 603689.SH and 000786.SZ are analogous, as they are both estimated using the maximum likelihood estimation method to obtain the parameters $\mu = 2.9116$ and $\sigma = 1.2326$. These parameters correspond to the prediction of the stock price of the SDE, which is given by

$$dS_t = 2.9116 S_t dt + 1.2326 S_t dW_t. \tag{96}$$

The hypothesis testing of the parameters μ and σ using Matlab's own function chi2gof yielded a p-value of 0.0331, which is greater than the significance level of 0.01. The stock price on 6 May 2024 is determined by the following formula

$$\hat{E}[S_2(t)] = 56.58\, exp(2.9116t). \tag{97}$$

The simple return rate is then used to determine the daily return rate on 6 May 2024, which is calculated to be 2.95%.

The RStudio software is employed to forecast the daily return rates of 688089.SH and 000752.SZ, utilising the RTSA model. In the case of stock 688089.SH, the initial step is to verify the smoothness of the series using the ADF test. In all three Equations (19)–(21) the *p*-values are less than 0.05, indicating that the series is smooth. This is true for the following scenarios: no-drift-no-trend, with-drift-no-trend, and with-drift-and-trend. Subsequently, the LB test is conducted by Equation (22), resulting in *p*-values of 6.324×10^{-3} and 9.813×10^{-3} for periods 6 and 12, respectively. These values are less than 0.05, indicating that the series is not a white noise series.

The series is then fitted using the ARIMA model with the BIC criterion to obtain the optimal model as ARIMA(1,0,0). The coefficient of this model is $\phi_1 = 0.2588$. Consequently, the model equation is as follows:

$$Y_t = 0.2588 Y_{t-1} + \varepsilon_t, \tag{98}$$

where Y_t is the daily return rates series, and ε_t is the error term at moment t. Then, the prediction equation is given by:

$$\hat{E}\left[\eta_1(t)\right] = 0.2588 Y_{t-1}. \tag{99}$$

In accordance with the aforementioned prediction equation, a daily return rate of 3.0478% is predicted for 6 May 2024.

As with the stock 688089.SH, we first perform an ADF test on 000752.SZ. In all three cases, the p-value is less than 0.05, indicating that the series is smooth. Subsequently, the LB test is performed, and the p-values for periods 6 and 12 are found to be 2.985×10^{-9} and 4.736×10^{-8}, respectively, which are less than 0.05. This suggests that the series is not a white noise series.

Subsequently, the series is fitted using the ARIMA model, with the optimal model selected using the BIC criterion to obtain an ARIMA$(1,0,1)$ model. The coefficients of this model are $\phi_1 = 0.7440$ and $\theta_1 = 0.4930$. Consequently, the model equation is as follows:

$$Y_t = 0.7440Y_{t-1} + \varepsilon_t - 0.4930\varepsilon_{t-1}, \tag{100}$$

where Y_t is the daily return rate series, and ε_t is the error term at moment t. Then, the prediction equation is given by:

$$\hat{E}[\eta_2(t)] = 0.7440Y_{t-1}. \tag{101}$$

In accordance with the aforementioned prediction equation, a daily return rate of 1.4520% is predicted for 6 May 2024.

For the above, the results of hypothesis testing are presented in **Table 3**, while the prediction results are shown in **Table 4**.

Table 3. Hypothesis testing passed.

No.	Code	UDE	UTSA	SDE	RTSA
1	300528.SZ	√	×	×	×
2	000651.SZ	√	×	×	×
3	600886.SH	×	√	×	×
4	600674.SH	×	√	×	×
5	000786.SZ	×	×	√	×
6	603689.SH	×	×	√	×
7	688089.SH	×	×	×	√
8	000752.SZ	×	×	×	√

Lower turnover rate is denoted as $a = 0.01$, maximum risk tolerance as $b = 0.05$, maximum investment proportion for each stock as $c = 0.25$, target return rates as $h = 0.04$, and risk preference factor as $\iota = 0.75$. In this context, GA is employed to solve two sub-models, model (76) and (77), as well as the optimization model (75). The results obtained are presented in **Table 5**. The solutions for model (76) and model (77) are $R_{max} = 2.76\%$ and $V_{min}^{-} = 1.553 \times 10^{-4}$, respectively, which are used as normalization standards. Subsequently, these values are substituted into model (75) to obtain the solution for the optimization model:

$$x=(0.0764,0.2495,0.1767,0.0509,0.0187,0.1715,0.0137,0.2427).$$

Table 4. The data of eight stocks.

	No.	Return rates	Expect value (%)
UDE	1	μ_1	1.7100
	2	μ_2	0.3500
UTSA	3	ν_1	1.7095
	4	ν_2	3.3364
SDE	5	ξ_1	2.3500
	6	ξ_2	2.9500
RTSA	7	η_1	3.0478
	8	η_2	1.4520

According to model (75), the optimal value (OV) is $M = 1.8521$, accompanied by lower variances of 5.5994×10^{-4} and return rate of 1.6337%.

Table 5. The investment proportions in model (75) and sub-models (76) and (77).

	x	**y**	**p**	**q**	**OV**
Model (75)	(0.0764, 0.2495)	(0.1767, 0.0509)	(0.0187, 0.1715)	(0.0137, 0.2427)	$M = 1.8521$
Model (76)	(0.0024, 0.0025)	(0.2406, 0.2488)	(0.0073, 0.2488)	(0.2484, 0.0023)	$R_{max} = 2.76\%$
Model (77)	(0.0032, 0.0030)	(0.2373, 0.2496)	(0.0073, 0.2492)	(0.2461, 0.0047)	$V_{min}^{-} = 1.553\times10^{-}$

Furthermore, the investment proportions and OVs are calculated for varying values of ι, as illustrated in **Table 6**, along with the corresponding values of R and V^{-}, as shown in **Table 7**. As evident in **Figure 1**, the values of ι, resulted in distinct OVs. In **Figure 2** (a) and (b), the dashed lines represent $R_{max} = 2.76\%$ and $V_{min}^{-} = 1.553 \times 10^{-4}$. As ι increases, the changes in R and V^{-} are not substantial, but the difference in multiples between V^{-} and V_{min}^{-} was much larger than the multiples between R_{max} and R. Consequently, when normalized, $\frac{V^{-}-V_{min}^{-}}{V_{min}^{-}}$ became significantly larger than $\frac{R_{max}-R}{R_{max}}$. Therefore, with an increasing ι, the weight of $\frac{V^{-}-V_{min}^{-}}{V_{min}^{-}}$ became more pronounced, leading to an increase in the value of M.

Table 6. The investment proportions and OV for different ι.

	x	**y**	**p**	**q**	**OV**
$\iota = 0.1$	(0.0040, 0.0020)	(0.2328, 0.2485)	(0.0131, 0.2484)	(0.2471, 0.0051)	0.0012
$\iota = 0.2$	(0.0184, 0.2476)	(0.1836, 0.1232)	(0.0809, 0.1828)	(0.0144, 0.1490)	0.1361
$\iota = 0.3$	(0.0168, 0.2144)	(0.1894, 0.0352)	(0.0993, 0.1894)	(0.0056, 0.2499)	0.4423

Table 6. (*Continued*).

	x	y	p	q	OV
$\iota=0.4$	(0.0522, 0.2251)	(0.2281, 0.1023)	(0.1152, 0.0795)	(0.0293, 0.1684)	0.6956
$\iota=0.5$	(0.0138, 0.2410)	(0.2152, 0.1090)	(0.0751, 0.2150)	(0.0023, 0.1287)	0.8308
$\iota=0.6$	(0.1288, 0.2428)	(0.1981, 0.0280)	(0.1479, 0.0515)	(0.0163, 0.1866)	1.5306
$\iota=0.7$	(0.1749, 0.2475)	(0.2034, 0.0291)	(0.1071, 0.1370)	(0.0138, 0.0872)	1.6587
$\iota=0.8$	(0.2373, 0.2486)	(0.2472, 0.0045)	(0.0071, 0.0028)	(0.0050, 0.2476)	2.7690
$\iota=0.9$	(0.0783, 0.2350)	(0.2086, 0.0594)	(0.1230, 0.1090)	(0.0023, 0.1843)	2.2430

Table 7. R and V^- for different ι.

ι	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
$R(\%)$	2.7587	1.8327	1.7170	1.7381	1.8439	1.5575	1.6549	1.3299	1.6564
$V^-(\times 10^{-4})$	1.5408	4.6970	5.2120	5.1160	4.6489	5.9659	5.4994	7.1296	5.4924

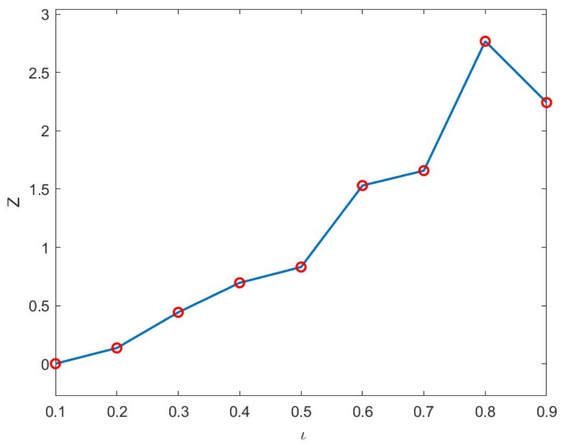
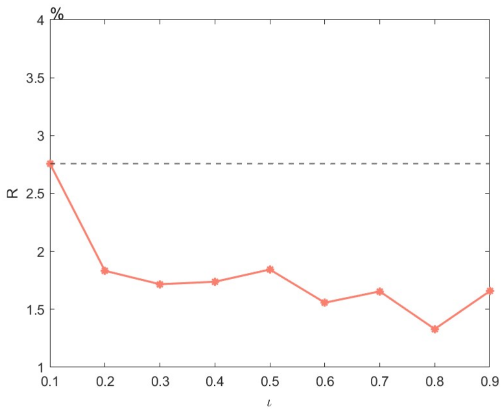
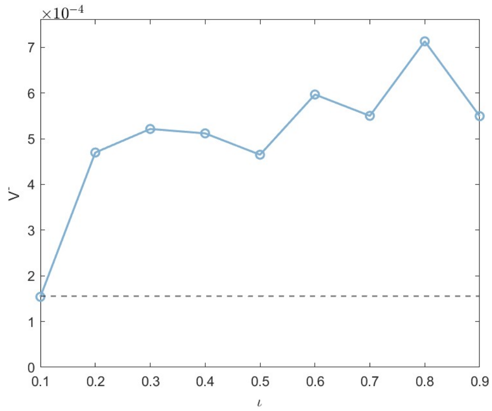


Figure 1. The OV for different ι.



(a) R for different ι.



(b) V^- for different ι.

Figure 2. R and V^- for different ι.

7. Conclusion

This paper discussed the portfolio problem in an uncertain random environment. Stock return rates were characterized using uncertain random variables. To enhance the accuracy of stock return rate predictions, this paper employed four methods to predict stock return rates: UDE, UTSA, SDE, and RTSA. Additionally, the lower variance was utilized to measure investment risk, which aligned more closely with the definition of risk and the mindset of investors than measuring risk with variance. A risk preference factor ι was introduced, leading to the proposal of an uncertain random mean-lower variance portfolio model. The results of numerical experiments demonstrated the effectiveness of the method for solving portfolio return rate maximization.

The model is being constructed based on the last five years of market data. However, this dataset may be subject to sample selection bias, as it fails to take into account changes in the economic cycle. Furthermore, the accuracy and completeness of the data are pivotal factors influencing the model's validity, particularly when confronted with the absence of data or anomalous volatility. Furthermore, market volatility and policy changes represent external factors that must be taken into account. These factors not only interfere with the normal functioning of the market, but they may also cause the historical performance on which the model is based to fail to reflect future trends. Although the model performs well on the training set, it may not be able to cope well with future market changes or extreme situations. These are some unresolved issues.

The portfolio model proposed in this paper has both theoretical value and the potential for wide-ranging applications in several fields. In economics, the model can be utilized for portfolio optimization and risk management. In engineering, it can be employed to assess the efficacy of project investments. Furthermore, in data science, the model can be integrated with machine learning algorithms for feature selection and predictive analysis. Future research could investigate the potential of machine learning techniques to enhance model performance on an ongoing basis. This not only has implications for theoretical research but also offers substantial support for investment decisions in practice.

Author contributions: Conceptualization, YS (Yanrui Su), YS (Yanjiao Song) and CL; methodology, YS (Yanrui Su), YS (Yanjiao Song) and CL; software, YS (Yanrui Su), YS (Yanjiao Song) and CL; validation, YS (Yanrui Su), YS (Yanjiao Song) and CL; formal analysis, YS (Yanrui Su), YS (Yanjiao Song) and CL; investigation, YS (Yanrui Su), YS (Yanjiao Song) and CL; resources, YS (Yanrui Su), YS (Yanjiao Song) and CL; data curation, YS (Yanrui Su); writing—original draft preparation, YS (Yanrui Su), YS (Yanjiao Song) and CL; writing—review and editing, YS (Yanrui Su), YS (Yanjiao Song) and CL; visualization, YS (Yanrui Su) and CL; supervision, YS (Yanrui Su) and YS (Yanjiao Song); project administration, YS (Yanrui Su), YS (Yanjiao Song) and CL; funding acquisition, YS (Yanrui Su), YS (Yanjiao Song) and CL. All authors have read and agreed to the published version of the manuscript.

Acknowledgments: The authors thank all the teachers at the School of Mathematics and Statistics, Nanjing University of Science and Technology, China, with special thanks to Professor Yuanguo Zhu for taking the time to give us a lot of guidance.

Conflict of interest: The authors declare no conflict of interest.

References

1. Markowitz H. Portfolio selection*. The Journal of Finance. 1952; 7(1): 77-91. doi: 10.1111/j.1540-6261.1952.tb01525.x
2. Jin Y, Zhang W. Robust portfolio optimization with higher moments. Mathematical Finance, 2021, 31(4): 1132-1165.
3. Yang J, Li Y. Stochastic dominance and portfolio optimization. Review of Financial Studies, 2024, 37(1): 50-78.
4. Konno H, Suzuki K ichi. A mean-variance-skewness portfolio optimization model. Journal of the Operations Research Society of Japan. 1995; 38(2): 173-187. doi: 10.15807/jorsj.38.173
5. Huang X. Portfolio selection with a new definition of risk. European Journal of Operational Research. 2008; 186(1): 351-357. doi: 10.1016/j.ejor.2007.01.045
6. Krejić N, Kumaresan M, Rožnjik A. VaR optimal portfolio with transaction costs. Applied Mathematics and Computation. 2011; 218(8): 4626-4637. doi: 10.1016/j.amc.2011.10.047
7. Li B, Shu Y. The skewness for uncertain random variable and application to portfolio selection problem. Journal of Industrial & Management Optimization. 2022; 18(1): 457. doi: 10.3934/jimo.2020163
8. Li J, Liu X. GARCH model with fractional Brownian motion correction. Statistics and Decision, 2021, 1(5): 29-33.
9. Konno H, Shirakawa H, Yamazaki H. A mean-absolute deviation-skewness portfolio optimization model. Annals of Operations Research. 1993; 45(1): 205-220. doi: 10.1007/bf02282050
10. Farrar D. The Investment decision under Uncertainty. Prentice Hall, Englewood Cliffs, New Jersey, 1962.
11. Li X, He L. Higher-order moments and risk management in portfolio optimization. Quantitative Finance, 2023, 23(1): 57-76.
12. Wang Z, Liu Q. Incorporating higher-order moments into portfolio construction under uncertainty. Review of Financial Studies, 2024, 37(3): 987-1020.
13. Zadeh L. Fuzzy sets. Information and Control, 1965, 8(3): 338-353.
14. Deng X, Li R. A portfolio selection model with borrowing constraint based on possibility theory. Applied Soft Computing. 2012; 12(2): 754-758. doi: 10.1016/j.asoc.2011.10.017
15. Pahade JK, Jha M. Credibilistic variance and skewness of trapezoidal fuzzy variable and mean–variance–skewness model for portfolio selection. Results in Applied Mathematics. 2021; 11: 100159. doi: 10.1016/j.rinam.2021.100159
16. Liu B. Uncertainty Theory, 2nd ed Springer-Verlag, Berlin, 2007.
17. Liu B. Uncertainty Theory: A Branch of Mathematics for Modeling Human Uncertainty. Springer-Verlag, Berlin, 2010.
18. Yu L, Wang L. Uncertainty Modeling in Portfolio Optimization: A Review. European Journal of Operational Research, 2021, 293(2): 645-659.
19. Cai X, Zhu H. Robust Portfolio Optimization under Uncertainty with Conditional Value-at-Risk. Operations Research Letters, 2022; 50(1): 82-90.
20. Liu S, Zhou X. Dynamic Portfolio Optimization with Uncertainty: A Stochastic Programming Approach. Mathematical Finance, 2023; 33(3): 678-705.
21. Liu Y. Uncertain random variables: a mixture of uncertainty and randomness. Soft Computing. 2012; 17(4): 625-634. doi: 10.1007/s00500-012-0935-0
22. Qin Z. Mean-variance model for portfolio optimization problem in the simultaneous presence of random and uncertain returns. European Journal of Operational Research. 2015; 245(2): 480-488. doi:

10.1016/j.ejor.2015.03.017
23. Mehlawat MK, Gupta P, Khan AZ. Portfolio optimization using higher moments in an uncertain random environment. Information Sciences. 2021; 567: 348-374. doi: 10.1016/j.ins.2021.03.019
24. Treanja, S. LU-Optimality Conditions in Optimization Problems with Mechanical Work Objective Functionals. IEEE Transactions on Neural Networks and Learning Systems, 2021.
25. Holland J. Adaptation in natural and artificial systems. University of Michigan Press, New York, 1975.
26. Kaya O, Schildbach J, Schneider S, Darius R. Robo-advice-a true innovation in asset management, Deutsche Bank Research, 2017.
27. DeMiguel V, Garlappi L, Uppal R. Optimal Versus Naive Diversification: How Inefficient is the 1/NPortfolio Strategy? Review of Financial Studies. 2007; 22(5): 1915-1953. doi: 10.1093/rfs/hhm075
28. Kearney C, Liu S. Textual sentiment in finance: A survey of methods and models. International Review of Financial Analysis. 2014; 33: 171-185. doi: 10.1016/j.irfa.2014.02.006
29. Yao K, Chen X. A numerical method for solving uncertain differential equations. Journal of Intelligent & Fuzzy Systems. 2013; 25(3): 825-832. doi: 10.3233/ifs-120688
30. Liu B. Uncertainty Theory, 4th ed Springer Berlin Heidelberg; 2015. doi: 10.1007/978-3-662-44354-5
31. Ye T, Liu B. Uncertain hypothesis test with application to uncertain regression analysis. Fuzzy Optimization and Decision Making. 2021; 21(2): 157-174. doi: 10.1007/s10700-021-09365-w
32. Phillips PCB, Perron P. Testing for a unit root in time series regression. Biometrika. 1988; 75(2): 335-346. doi: 10.1093/biomet/75.2.335
33. Akhter MF, Hassan D, Abbas S. Predictive ARIMA Model for coronal index solar cyclic data. Astronomy and Computing. 2020; 32: 100403. doi: 10.1016/j.ascom.2020.100403
34. Evans L. An Introduction to Stochastic Differential Equations. 2013. doi: 10.1090/mbk/082
35. Harlow WV, Rao RKS. Asset Pricing in a Generalized Mean-Lower Partial Moment Framework: Theory and Evidence. The Journal of Financial and Quantitative Analysis. 1989; 24(3): 285. doi: 10.2307/2330813
36. Yang X, Liu B. Uncertain time series analysis with imprecise observations. Fuzzy Optimization and Decision Making. 2018; 18(3): 263-278. doi: 10.1007/s10700-018-9298-z
37. Yao K, Liu B. Parameter estimation in uncertain differential equations. Fuzzy Optimization and Decision Making. 2019; 19(1): 1-12. doi: 10.1007/s10700-019-09310-y