

## Research Article

# Two-Stage Robust Optimization Model for Uncertainty Investment Portfolio Problems

Dongqing Luan,<sup>1</sup> Chuming Wang ,<sup>1</sup> Zhong Wu,<sup>2</sup> and Zhijie Xia<sup>1</sup>

<sup>1</sup>School of Management, Shanghai University of Engineering Science, Shanghai 201620, China

<sup>2</sup>School of Management, Shanghai University of International Business and Economics, Shanghai 201620, China

Correspondence should be addressed to Chuming Wang; wang\_chuming@163.com

Received 12 October 2021; Revised 7 November 2021; Accepted 23 November 2021; Published 7 December 2021

Academic Editor: Alfred Peris

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

Investment portfolio can provide investors with a more robust financial management plan, but the uncertainty of its parameters is a key factor affecting performance. This paper conducts research on investment portfolios and constructs a two-stage mixed integer programming (TS-MIP) model, which comprehensively considers the five dimensions of profit, diversity, skewness, information entropy, and conditional value at risk. But the deterministic TS-MIP model cannot cope with the uncertainty. Therefore, this paper constructs a two-stage robust optimization (TS-RO) model by introducing robust optimization theory. In case experiments, data crawler technology is used to obtain actual data from real websites, and a variety of methods are used to verify the effectiveness of the proposed model in dealing with uncertainty. The comparison of models found that, compared with the traditional equal weight model, the investment benefits of the TS-MIP model and the TS-RO model proposed have been improved. Among them, the Sharpe ratio, Sortino ratio, and Treynor ratio have the largest increase of 19.30%, 8.25%, and 7.34%, respectively.

## 1. Introduction

Investment portfolio means that investors invest their own assets in stocks, funds, bonds, and other securities at the same time, in order to achieve higher or more stable profits under lower risk conditions. In the early research, many scholars constructed the mean-variance model and the portfolio effective boundary model, which opened a new era of investment portfolio. Among them, since Markowitz founded the portfolio theory, the development of portfolio theory has been extremely rapid [1,2]. Risks persist due to poor risk management and uncertainty in the market environment. Portfolio theory is one of the important financial theories that discuss the profits and risks in the investment process. In the management of risk, many scholars believe that the research of investment portfolio can effectively measure uncertainty and avoid risk. The investment portfolio mainly embodies the idea of investment diversification. By screening different types of risk asset projects, the

purpose of reducing investment risks is achieved. The investment portfolio optimization method provides a powerful tool for risk management [3,4]. Using the model method of investment portfolio optimization to find a reasonable allocation of funds can help investors balance safety and profitability and ensure that they achieve maximum profits within the acceptable risk range. Whether it is the securities and finance industry or other academic fields such as research institutes and universities, how to effectively formulate appropriate investment strategies has always been the focus of research [5,6]. It is found that the game between investment and risk is always accompanied by investors. Continuously improving investment risk management capabilities is a top priority for regulatory authorities, financial institutions, and even individual investors.

Some investors tend to avoid risks by looking for the law of risk distribution, so that they can modify their strategies in a timely manner in the face of unexpected situations. In the application process of investment portfolio optimization,

there are few or even missing historical data. How to accurately measure risk in the absence of historical data and how to reasonably estimate possible losses taking into account a variety of factors is also a huge challenge. In addition, the study found that financial risk measurement is based on a basic distribution assumption. The selection of appropriate distribution assumptions has a significant impact on the accurate measurement of risk, which in turn affects the efficiency of the overall investment plan. How to choose an appropriate distribution hypothesis is also a pain point in portfolio research. In the research on the uncertainty of portfolio risk distribution, researchers focus on how to incorporate the uncertainty of loss into the investment decision-making process and then improve the problem of portfolio optimization. In addition to the existing research on the specific distribution probability of hypothetical risks, many scholars have tried to conduct research from the risk itself. It is found that the loss is affected by many factors. In an economic environment that is rising frequently, the distribution of losses may be similar to a normal distribution. In a flat economic environment, the risk loss distribution may be similar to a certain distribution [7–9]. In a recessionary economic environment, the risk loss distribution may be similar to a certain Pareto distribution [9,10]. Under different distribution models, the choice of portfolio strategy often leads to significant differences in expected results. How to effectively screen the probability distribution of risks and how to effectively determine the investment portfolio plan are a hot topic in the research.

In addition to the research on the probability of risk distribution, many scholars have conducted multidimensional research on the inherent uncertainty by citing models in cross-domains. Garlappi et al. (2007) constructed a portfolio model by estimating the confidence interval of the expected profit. The study found that the fuzzy aversion portfolio only performed stably for a period of time [11]. Tu and Zhou (2010) proposed a method that allows Bayesian Priors to reflect the objectives of economic problems. In terms of measurement of out-of-sample loss function, the model of portfolio strategy based on the objective priori can be significantly superior to the model developed in the classical framework [12]. Bayraktar and Zhou (2017) studied the fundamental theorem of asset pricing and option hedging prices under discrete-time nondominated model uncertainty and portfolio constraints. Research has discovered the duality of superhedging prices in a market where stocks are dynamically traded and options are statically traded [13]. The research of Golosnoy and Okhrin (2008) found that there is general uncertainty in the asset allocation model, and the optimal belief degree can be determined by developing a case-based decision-making method [14]. Some scholars use the linear loss aversion function to construct the model and obtain the optimal solution. This function assumes that the marginal utility of investors' gains and losses is constant. Investor's loss aversion degree and reference profit will vary with market conditions and relative wealth. The assumption that the loss aversion coefficient and the reference point remain unchanged

does not conform to the true psychology of investors [15,16]. Huang and Qiao (2012) discussed a multiperiod securities investment portfolio selection problem and proposed an uncertain risk index adjustment model. The goal of optimal investment portfolio adjustment is to maximize the total incremental wealth under the constraints of accumulating risk index values during the control investment period and satisfying the constraints of self-raised funds in each period [17]. When analyzing the optimal investment decision of loss aversion investors, the dynamic change characteristics of loss aversion coefficient and reference point should be considered. At present, when building loss aversion models, scholars assume that the expected profit of assets is determined, and the average profit and profit distribution of risky assets in the market are both uncertain. Using the uncertainty of asset profit distribution for reference, the assumption of unknown profit distribution is introduced into robust optimization to describe the fuzzy aversion characteristics of investors, which is more in line with the real psychology of investors. Uncertainties are particularly prominent, and the economy has been affected. In this case, reasonable investment decisions are particularly important. For multiperiod investments, the experience and wisdom of investors are extremely important, where investing, profits, risks, and liquidity are often involved. These three are interrelated and restrict each other. This relationship is in the relationship between profit and risk. The performance is more prominent in time. High risk means high returns, and risk and profit are often proportional, but investors hope to obtain higher profits with less risk, which poses a big problem for investors and researchers. Therefore, it is still necessary to continue to study multistage investment under uncertain conditions.

In recent years, with the steady optimization theory and methods being gradually accepted and recognized by scholars, the uncertainty of using robust optimization to study the portfolio has become particularly novel. The robust optimization theory simply expresses the uncertain parameters in a specific interval and can flexibly control the degree of uncertainty. The robust optimization theory proposed by Bertsimas and Sim (2004) and Ben Tal et al. (2007) to deal with the uncertainty in the model is now a commonly used research method [18,19]. Huang et al. (2010) studied how to use robust optimization to analyze uncertain parameter problems and proved the effectiveness of the robust optimization model [20]. Dai and Wen (2014) proposed a robust optimization theory for minimizing the CVaR of the investment portfolio under the general affine data perturbation uncertainty set. Their research uses numerical experiments with real market data to illustrate the practicality of robust optimization models [21]. In venture capital, investors pay more attention to the right-leaning loss distribution risk, and then it is of practical significance to introduce information entropy into risk measurement research in the financial field. Through the review of the above literature, diversification of investment demand and uncertainty of risk preference are also hot issues in investment portfolio research in recent years. Many scholars have

conducted in-depth investigations on the uncertainty of risk appetite. Information entropy can be used to measure the degree of deviation from uniform distribution. The deviation corresponds to a point that reflects the uncertainty of the probability distribution over the entire range of values. At the same time, information entropy can express more information about the distribution, more accurately reflect the risk of loss distribution, and then serve as an ideal means of measuring risk. It can be found from the above-mentioned literature that robust optimization, CAaR, and information entropy, as tools to measure stock investment risk, can be well applied in the symmetric and asymmetric distribution of probability. However, although most of the existing algorithms have played a certain role in improving the accuracy of the model solution, the convergence speed is often slow due to the complexity of the solution algorithm.

In the research of portfolio optimization, the mean-risk model usually assumes that the expected profit on risky assets is subject to a known probability distribution and that the number of transactions is infinitely divided, which is not consistent with the actual situation. In this paper, a two-stage robust portfolio optimization model is constructed, and the optimal solution in the worst case is found in the uncertain set. Based on the analysis and comparison of the above related literature, the research in this paper mainly includes the following innovations:

- (i) This paper summarizes the research results of related scholars, analyzes the classical portfolio problems in depth, and proposes a deterministic TS-MIP model considers the risk and diversity of investment problems.
- (ii) On the basis of deterministic model, the research content is further extended to uncertainty, and the mechanism of influence of uncertainty parameters on portfolio efficiency is analyzed in depth.
- (iii) In the research of dealing with uncertainty interference, this paper introduces the theoretical method of robust optimization, constructs a TS-RO model, and applies it to the research field of investment portfolio.
- (iv) In the numerical case analysis part, this paper uses data crawling technology to obtain the latest real data from real-time websites as the input part of the model and uses real data for a large number of case studies and parameter comparisons, to obtain practical investment advice.

The rest of this framework is as follows: Section 2 lists related basic theories and methods. In the third section, based on the basic deterministic TS-MIP model, we analyze the problem under uncertain parameters and further construct the TS-RO model. Section 4 obtains relevant data through data crawling technology and constructs numerical cases for analysis. Section 5 compares the key parameters of the model and puts forward some management investment recommendations.

## 2. Basic Theory and Model

**2.1. Information Entropy.** Information entropy can be defined as the possibility of discrete random events [22–24], and the specific expression is shown in

$$S = - \sum_{i=1}^n p_i \ln p_i. \quad (1)$$

Among them, the function of  $p_i$  is to measure the probability of occurrence of the situation  $i$ . In the current stage of theoretical research, information entropy has been widely used to measure the uncertainty of information, and there is a positive correlation between entropy and uncertainty. In addition, the information entropy function also has the following three special properties:

- (i) Continuity:  $S$  is a continuous function of  $p_i$ , ( $p_i \in [0, 1]$ )
- (ii) Nonnegativity:  $S(p_1, p_1, p_1, \dots, p_n) \geq 0$
- (iii) Extreme value: if and only if the condition is  $p_1 = p_2 = \dots = p_n$ , the information entropy achieves the maximum value

**2.2. Value-Added Entropy and Actual Entropy.** Value-added entropy is a kind of generalized entropy [25,26], which can be defined as

$$H = \sum_{i=1}^W p(x_i) \log \sum_{i=1}^N x_i r_i. \quad (2)$$

Suppose there are  $N$  types of stocks or securities in the market, and the  $k$  type contains  $n_i$  possible prices ( $i = 1, 2, 3, \dots, N$ ). The number of all possible value combinations is  $W = n_1 \times n_2 \times n_3 \times \dots \times n_N$ .  $\mathbf{X} = (x_1, x_2, x_3, \dots, x_k; i = 1, 2, \dots, W)$  refers to the investment ratio of the  $k$  securities, and  $\mathbf{R} = (r_1, r_2, r_3, \dots, r_k)$  is the output ratio in the  $i$  price combination. In addition,  $R_i = 1 + r_i$  and  $r_i$  represent the rate of profit, assuming that short selling is not allowed, that is,  $r_k \geq 0, k = 1, 2, \dots, N$ . The above formula (2) can be further transformed into

$$H = \log \prod_{i=1}^W R_i^{p(x_i)}, \quad (3)$$

$$R_i = \sum_{k=1}^N q_k R_{ik}.$$

**Actual entropy:** in the actual financial market, investment transactions require costs, and an investment portfolio that ignores transaction costs will often have a great impact on actual profits and may make investments invalid. Introduce typical transaction costs into the portfolio and define linear transaction costs. However, linear transaction costs rarely appear in investment costs, so the improved transaction cost function  $C(\mathbf{X}) = [c(x_1), c(x_2), c(x_3), \dots, c(x_n)]$  can be expressed as

$$c(x_i) = \begin{cases} r\sqrt{x_i}, & 0 \leq x_i \leq a, \\ r(kx_i^2 + hx_i + q), & a \leq x_i \leq b, 0 \leq k \leq 1, \\ rx_i^3, & b \leq x_i \leq 1. \end{cases} \quad (4)$$

This paper draws on the improved typical transaction cost function and establishes a dual-objective portfolio model with the largest profit expectation and the smallest risk expectation. Set  $\mathbf{X} = (x_1, x_2, x_3, \dots, x_n)$  as the investment ratio in  $n$  stocks,  $\mathbf{R} = (r_1, r_2, r_3, \dots, r_n)$  as the expected rate of profit, and  $M(\mathbf{X})$  as the actual rate of profit of stocks and securities, and the portfolio model based on the CVaR measurement is expressed as

$$\Pi = \sum_{i=1}^n r_i x_i - \sum_{i=1}^n c_i x_i. \quad (5)$$

Among them,  $f(\mathbf{X}, \bar{Y}_j) = -X' \bar{Y}_j$ ,  $x_i \geq 0$  means that short selling is not allowed, if and only if  $\omega(\mathbf{X}) > \omega(\mathbf{Y})$  and  $\bar{F}_\beta(\mathbf{X}, \delta, \mu) < \bar{F}_\beta(\mathbf{Y}, \delta, \mu)$ . Actual entropy can reflect the speed of appreciation and the cumulative profit of funds. Compared with the mean function that cannot directly reflect the cumulative profit of the fund, the combination of actual entropy and actual entropy can be more effective and more appropriate to reflect the investment profit.

2.3. CVaR. VaR can usually be defined as value at risk [27,28], that is, the maximum loss of an investment portfolio in a normal market in a specific period under a specific confidence level of  $\alpha$ . Set  $f(\mathbf{X}, \mathbf{Y}): \mathbf{R}^m \times \mathbf{R}^n \rightarrow \mathbf{R}^m$  as the loss function of the investment portfolio, and the specific

meaning is the expected loss of the investment portfolio. Among them,  $\mathbf{X}$  is the investment decision vector, which specifically means the proportion of risky securities in the investment portfolio.  $\mathbf{Y}$  is a 0–1 variable of value at risk, which represents the profit rate of securities stocks. Then, there is  $f(\mathbf{X}, \mathbf{Y}) = -X' \mathbf{Y}$ . When  $\mathbf{X}$  is fixed,  $f(\mathbf{X}, \mathbf{Y})$  is a function of  $\mathbf{Y}$ . Assuming that  $\mathbf{Y}$  is a continuous random variable, the probability density function of  $\mathbf{Y}$  is  $\varphi(\mathbf{Y})$ , then  $\forall \delta \in \mathbf{R}$ , and the distribution function is  $\psi(\mathbf{X}, \delta) = \int_{f(\mathbf{X}, \mathbf{Y}) \geq \delta} \varphi(\mathbf{Y}) d\mathbf{Y}$ , which represents the probability that the loss  $f(\mathbf{X}, \mathbf{Y})$  does not exceed  $\delta$ ,  $\delta$  is the upper limit of risk, and  $\psi(\mathbf{X}, \delta)$  is the nonincreasing right continuous function of  $\delta$ . Assuming that  $\beta$  ( $0 \leq \beta \leq 1$ ) is used to represent the confidence level,  $\mathbf{X}$  of VaR is expressed as

$$\text{VaR}_\beta(\mathbf{X}) = \min\{\delta \in \mathbf{R}: \psi(\mathbf{X}, \delta) \geq \beta\}. \quad (6)$$

However, because VaR is only a quintile within a certain confidence level, the left-tail risk beyond this quintile cannot be measured by VaR, which leads to failure in assessing the overall risk of the portfolio and makes investors may ignore huge losses in some small probability events. At the same time, VaR cannot measure the risks in some extreme situations, such as financial crises. Based on this, VaR under certain CVaR was proposed. CVaR can be defined as the conditional average value of losses exceeding under a certain confidence level, which is used to measure the average excess level of losses [29,30]. According to the research of related scholars, CVaR is more sensitive to the shape of the loss distribution at the tail of the distribution [31,32]. According to the definition of CVaR, we can get

$$\text{CVaR}(\mathbf{X}) = E[f(\mathbf{X}, \mathbf{Y}) | f(\mathbf{X}, \mathbf{Y}) \geq \text{VaR}_\beta(\mathbf{X})] = (1 - \beta)^{-1} \int_{f(\mathbf{X}, \mathbf{Y}) \geq \text{VaR}_\beta(\mathbf{X})} f(\mathbf{X}, \mathbf{Y}) \varphi(\mathbf{Y}) d\mathbf{Y}. \quad (7)$$

It is found that VaR must be used in the process of solving equation (7), and the mathematical expression is extremely complicated, and it is difficult to calculate the result. In order to solve this problem, Rockafellar and other scholars constructed an auxiliary function  $F_\beta(\mathbf{X}, \delta)$  and effectively connected CVaR and VaR [33,34]. The auxiliary function is

$$F_\beta(\mathbf{X}, \delta) = \delta + (1 - \beta)^{-1} \int_{\mathbf{Y} \in \mathbf{R}^n} \max[f(\mathbf{X}, \mathbf{Y}) - \delta, 0] \varphi(\mathbf{Y}) d\mathbf{Y}, \quad (8)$$

where  $\text{VaR}_\beta(\mathbf{X}) = \min F_\beta(\mathbf{X}, \delta)$ . Due to the diversity and complexity of the market environment, the loss caused is also uncertain, and the probability density function  $\varphi(\mathbf{Y})$  is not easy to calculate. Therefore, the method of using historical data to simulate calculations is selected, and historical data is used to predict the specific future random vector. Assuming that the random variable  $\mathbf{Y}$  in the past trading day represents  $m$  scenarios, the value of each scenario is  $\bar{Y}_j$ ,  $j = 1, 2, \dots, m$ , and the probability of any one scenario is  $1/m$ . So, it exists that

$$\bar{F}_\beta(\mathbf{X}, \delta) = \delta + m^{-1} (1 - \beta)^{-1} \sum_{j=1}^m \{\max[f(\mathbf{X}, \bar{Y}_j) - \delta, 0]\}. \quad (9)$$

Because of the existence of function max in equation (9), and because function  $\bar{F}_\beta(\mathbf{X}, \delta)$  is usually not continuous and differentiable, general algorithms cannot be solved. Here, the aggregation function proposed by related scholars is used to smooth the max function [35,36]. For any  $\mu > 0$ , the above function can be expressed as

$$\bar{\bar{F}}_\beta(\mathbf{X}, \delta, \mu) = \delta + m^{-1} (1 - \beta)^{-1} \sum_{j=1}^m \mu \ln \left[ \exp \frac{f(\mathbf{X}, \bar{Y}_j) - \delta}{\mu} + 1 \right]. \quad (10)$$

2.4. Two-Stage Mixed Integer Programming (TS-MIP) Model. A diversified investment portfolio has lower volatility risk than personal assets, and increasing the diversification of the investment portfolio can help investors reduce the variance of the investment portfolio. Not only

can the risk in the investment portfolio be effectively and accurately measured, but also the investment portfolio can evaluate diversity. According to investment portfolio theory, diversity  $G$  can be used to reduce nonsystematic risks. Specifically, the greater the value, the more diversified the investment portfolio. Based on the above theory, in the model proposed in this paper, CVaR is used to adjust the direct function to determine the new objective function to measure the risk of the investment portfolio. The value-added entropy combined with the real entropy function  $M$  is used to measure the statistical profit of investment. Skewness function is used to measure investment increment. Based on the above analysis and combined investment portfolio strategy research, this paper proposes the following TS-MIP model, the structure of which is shown in Figure 1. The problem of outer layer maximization benefit planning mainly analyzes the maximization of income and the maximization of investment diversity and the maximization of skewness. Corresponding to it is the problem of internal minimization of risk, which mainly includes the minimization of information entropy and the minimization of investment risk. The investment risk here is measured by the CVaR.

The stage of maximizing the benefit of the outer layer is the overall stage, and the specific form is as follows:

$$\begin{aligned}
\max \quad & Z = \{\alpha(R'X) + \beta(S'X) + \gamma E_\delta \min \tilde{F}_\beta(X, Y)\} \\
\text{s.t.} \quad & \sum_{i=1}^M \alpha r_i x_i - \sum_{i=1}^M c_i x_i \geq \Pi^{\min}, \quad \forall r_i \in \mathbf{R}, i = 1, 2, 3, \dots, M, \\
& \sum_{i=1}^M x_i = 1, \quad \forall x_i \in \mathbf{X}, \\
& \sum_{i=1}^M \beta x_i \ln x_i \leq G^{\max}, \quad i = 1, 2, 3, \dots, M, \\
& \sum_{i=1}^M s_i x_i \leq U^{\max}, \quad \forall s_i \in \mathbf{S}, i = 1, 2, 3, \dots, M, \\
& \alpha, \beta \in \{0, 1\}, \\
& x_i \in [0, 1], \quad i = 1, 2, 3, \dots, M.
\end{aligned} \tag{11}$$

The objective function is combinatorial optimization problem, which maximizes the profit under the condition of the lowest risk. The first line of constraint indicates the lower bound of profit. The second line of constraint indicates the maximum investment ratio, short selling is not allowed, and there is no additional capital inflow. The third line of constraints indicates the upper limit of investment diversity. The fourth line of constraints indicates the upper limit of the investment increment. The fifth and sixth lines of constraints are related variable constraints; the purpose is to ensure the feasibility. The inner minimization problem is second stage, and the form is as follows:

$$\begin{aligned}
& E_\delta \min \tilde{F}_\lambda(X, Y, K, \gamma, \delta, \mu), \\
\text{s.t.} \quad & - \sum_{i=1}^M \sum_{j=1}^N x_i p_j \ln \left| \frac{r(\text{CVaR})_i}{\sum_{i=1}^M r(\text{CVaR})_i} \right| \geq K^{\min}, \\
& \delta + \frac{1}{m(1-\lambda)} \sum_{i=1}^M \sum_{j=1}^N \mu \ln \left\{ \exp \frac{\gamma[f(X, Y) - \delta]}{\mu} + 1 \right\} \leq r_{\text{CVaR}}^{\max}, \\
& f(X, Y) = \sum_{i=1}^M \sum_{j=1}^N x_i y_j, \quad \forall x_i \in \mathbf{X}, \forall y_j \in \mathbf{Y}, \\
& \sum_{i=1}^M x_i = 1, \quad \forall x_i \in \mathbf{X}, \\
& \sum_{i=1}^M \sum_{j=1}^N x_i p_j = 1, \quad p_j \geq 0, \forall x_i \in \mathbf{X}, \\
& \gamma \in \{0, 1\}, \\
& x_i \in [0, 1], \quad \forall x_i \in \mathbf{X}, i = 1, 2, 3, \dots, M, \\
& y_j \in [0, 1], \quad y_j \in \mathbf{Y}, j = 1, 2, 3, \dots, N.
\end{aligned} \tag{12}$$

The specific meaning: the objective function is to minimize information entropy, the risk. The first line of constraints represents that the lowest value of information entropy is higher than the limit threshold. The second line of constraint represents that the risk that investors can bear is lower than the upper bound of the maximum tolerance. The third line of constraints describes the calculation strategy for cumulative risk. The fourth line of constraints indicates the maximum investment ratio and no remittance of external funds. The fifth line of constraints indicates that the probability of all scenarios is 1. The sixth, seventh, and eighth lines of constraints are related variable constraints. In order to calculate CVaR with  $r_{\text{VaR}}$ , assuming that the rate of profit on investment shows a normal distribution,  $r_{\text{VaR}}$  can be expressed as  $r_{\text{VaR}} = r_q + \sigma_q z_\alpha$  according to the law of large numbers, where  $r_q$  measures the maximum loss of the investment portfolio. Then, we obtained that

$$r(\text{CVaR}) = \frac{1}{1-\alpha} \partial(z_\alpha) \sigma_q - r_q. \tag{13}$$

Among them,  $\partial(z_\alpha)$  is the probability density function of the standard normal distribution, which can be expressed as

$$\partial(z_\alpha) = \frac{1}{\sqrt{2\pi}} e^{-(z_\alpha^2/2)}. \tag{14}$$

In addition, because  $p_i$  cannot directly predict and is extremely difficult to calculate, this paper uses a computer program to assist in the solution of the model. In order to introduce the two-stage optimization model and its application in the risk-considering investment portfolio problem,

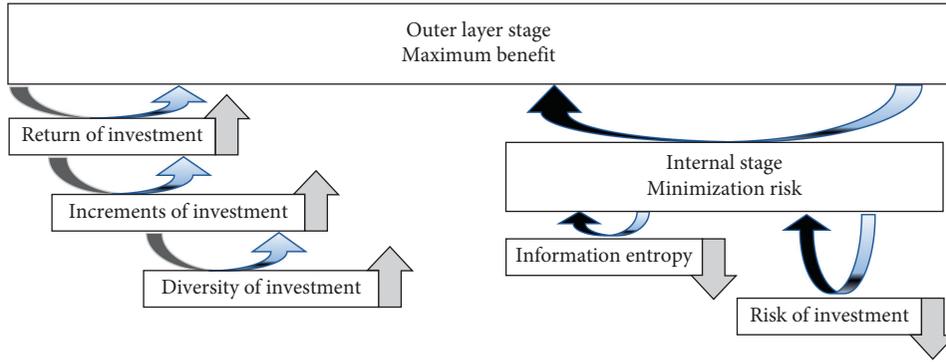


FIGURE 1: The theoretical framework of the TS-MIP model.

the relevant decision variables and parameters are summarized in Table 1.

### 3. Two-Stage Robust Optimization (TS-RO) Model with Uncertainty

**3.1. Portfolio Optimization Problem with Uncertainty.** The real capital market environment is full of instability and randomness, and it is extremely difficult or even impossible to obtain the accuracy of data parameters. In the research of dealing with uncertain information, many scholars have cited the theory of robust optimization. Robust optimization model can effectively provide an effective measure of uncertainty. Studies have found that it plays an extremely restrictive role in dealing with uncertain information parameters. In this section, by citing the theory of robust optimization, a two-stage robust optimization model for solving portfolio problems under uncertain conditions is constructed [37]. Based on the analysis of real securities and stock markets, this paper takes the two-stage basic model as the research object to explore the impact of uncertainty in depth. In the actual market analysis process, the upper bound of the expected risk probability is not always certain. In other words, in a complex capital market environment, investors' risk attitudes are not always consistent. As investors' risk attitudes change, the original assumptions of expected risk probability as deterministic will not reflect the reality. Therefore, this section extends the expected risk probability from certainty to uncertainty. According to the relevant theory of robust optimization, the assumptions are as follows. Set the expected risk probability that  $\delta_i \rightarrow \tilde{\delta}_i$  is  $\tilde{\delta}_i = \delta_i^0 + \hat{\delta}_i$ , where  $\delta_i^0$  is the nominal demand,  $\hat{\delta}_i = \varepsilon \delta_i^0$  and  $\varepsilon$  are the disturbance items [38]. Based on the above problem description and feasibility assumptions, this paper builds the TS-RO model. The expected risk of uncertainty in the model is  $\tilde{\delta}_i$ , and the uncertainty set is the most commonly used ellipsoid set  $E_E$  [39]. The  $l_2$  norm is used to define the uncertain ellipsoid sets  $\{\cup_E\}$ ,  $\{\cup_E | \varepsilon_j: \|\varepsilon_j\|_2 \leq \Omega \iff \varepsilon_j \sqrt{\sum_j \varepsilon_j^2} \leq \Omega\}$ , and  $\Omega_j$  as adjustable safety parameters, as well as the spherical diameter of the uncertain set. It means that at most  $\Omega_j$  disturbance items

deviate from the nominal set  $\{\cup_E\} = \{\tilde{\delta}_j \in \mathbb{R}, \sum_{\mathcal{F}} [(\tilde{\delta}_j - \delta_j^0)/\hat{\delta}_j]^2 \leq \Omega_j^2\}$  [40,41].

**Theorem 1.** For the given condition that  $\tilde{\delta}_i$  is closed and convex, when the parameter  $\varepsilon = 0$  is uncertain, the constraint adjustment of the second stage part of the TS-RO model under uncertain conditions is equivalent to TS-MIP model, that is,  $\max\{\tilde{\delta}_i + H(\mathbf{X}, \mathbf{Y}, \tilde{\delta}_i)\} \leq r_{CVaR}^{\max} \iff \sup\{\tilde{\delta}_i + [\varepsilon]_{\varepsilon=0} H(\mathbf{X}, \mathbf{Y}, \varepsilon, \tilde{\delta}_i)\} \leq r_{CVaR}^{\max}$ , in which, in order to simplify the expression, set  $H(\mathbf{X}, \mathbf{Y}) = (1/m(1-\lambda)) \sum_{i=1}^M \sum_{j=1}^N \mu \ln\{\exp(\gamma[f(\mathbf{X}, \mathbf{Y}) - \delta]/\mu) + 1\}$ .

*Proof.* The conventional ellipsoid uncertainty set  $\{\cup_E\}$  is  $\{c_i \in \mathbb{R}^n: c_i = \bar{c}_i + \Delta\xi, \xi \leq \Omega\}$ , where  $\Delta = \sum^{1/2}$ , and its constraints  $\max c_i^T X \leq B$  can be translated as  $\max\{c_i^T X: (c_i - \bar{c}_i)^T \Sigma^{-1} (c_i - \bar{c}_i) \in \Omega^2\}$ . For  $\forall \Sigma$  is positive and closed, it is a convex problem. Therefore, convex problem can be solved by Karush-Kuhn-Tucher condition.  $\min \mathcal{F}(c_i^*) = -c_i^{*T} X$  with s.t.  $g(c_i^*) = (c_i^* - \bar{c}_i)^T \Sigma^{-1} (c_i^* - \bar{c}_i) - \Omega^2 \leq 0$ . Therefore, the meaningful solution to deal with uncertain problems is its robust feasible solution. In all robust feasible solutions, how to interpret the value of the target (and possibly uncertain) still needs deeper analysis. When applied to the target problem, it is natural to quantify the quality of the solution through the guaranteed value  $\sup_{E_E} E[\tilde{\delta}_i: c]$  (i.e., the maximum value) with the help of the "worst case-oriented" concept of robust theory. Thus, the best possible robust feasible solution is the one that solves the TS-MIP model  $\max\{\tilde{\delta}_i + H(\mathbf{X}, \mathbf{Y}, \tilde{\delta}_i)\} \leq r_{CVaR}^{\max}$ , or which is the same, the optimization problem  $\sup\{\tilde{\delta}_i + [\varepsilon]_{\varepsilon=0} H(\mathbf{X}, \mathbf{Y}, \varepsilon, \tilde{\delta}_i)\} \leq r_{CVaR}^{\max}$  in the TS-RO model. The latter problem is called the robust counterpart of the original TS-MIP problem. Above all, Theorem 1 can be proved.  $\square$

**3.2. Two-Stage Robust Optimization (TS-RO) Model with Uncertainty.** Based on the above analysis and combined investment portfolio strategy research, this paper proposes the following TS-RO model. Under uncertain conditions, the expression form of the outer maximization stage is the same as the TS-MIP model, and the specific form is as follows:

$$\begin{aligned}
 \sup_{\varepsilon \in E_E} \mathcal{Z} &= \left\{ \alpha(R' \mathbf{X}) + \beta(S' \mathbf{X}) + \gamma_\delta \mathbb{E} \inf_{\varepsilon \in E_E} \tilde{F}_\beta(\mathbf{X}, \mathbf{Y}, \tilde{\delta}_j) \right\}, \\
 \text{s.t. } \sum_{i=1}^M \alpha r_i x_i - \sum_{i=1}^M c_i x_i &\geq \Pi^{\min}, \quad \forall r_i \in \mathbf{R}, \quad i = 1, 2, 3, \dots, M, \\
 \sum_{i=1}^M x_i &= 1, \quad \forall x_i \in \mathbf{X}, \\
 \sum_{i=1}^M \beta x_i \ln x_i &\leq G^{\max}, \quad i = 1, 2, 3, \dots, M, \\
 \sum_{i=1}^M s_i x_i &\leq U^{\max}, \quad \forall s_i \in \mathbf{S}, i = 1, 2, 3, \dots, M, \\
 \alpha, \beta &\in \{0, 1\}, \\
 x_i &\in [0, 1], \quad i = 1, 2, 3, \dots, M.
 \end{aligned} \tag{15}$$

Since the risk attitude of investors is considered, the expression form of the second stage inner minimization

problem is significantly different from the basic model. The specific form is as follows:

$$\begin{aligned}
 &\mathbb{E}_\delta \inf_{\varepsilon \in E_E} \tilde{F}_\lambda(\mathbf{X}, \mathbf{Y}, K, \gamma, \tilde{\delta}_j, \mu), \\
 \text{s.t. } &-\sum_{i=1}^M \sum_{j=1}^N x_i p_j \ln \left| \frac{r(\text{CVaR})_i}{\sum_{i=1}^M r(\text{CVaR})_i} \right| \geq K^{\min}, \\
 &\delta_j^0 + \Omega_j \left\{ \frac{1}{m(1-\lambda)} \sum_{i=1}^M \sum_{j=1}^N \mu \ln \left\{ \exp \frac{\gamma[f(\mathbf{X}, \mathbf{Y}) - \hat{\delta}_j]}{\mu} + 1 \right\} \right\} \leq r_{\text{CVaR}}^{\max}, \\
 &f(\mathbf{X}, \mathbf{Y}) = \sum_{i=1}^M \sum_{j=1}^N x_i y_j, \quad \forall x_i \in \mathbf{X}, \forall y_j \in \mathbf{Y}, \\
 &\sum_{i=1}^M x_i = 1, \quad \forall x_i \in \mathbf{X}, \\
 &\sum_{i=1}^M \sum_{j=1}^N x_i p_j = 1, \quad p_j \geq 0, \forall x_i \in \mathbf{X}, \\
 &\gamma \in \{0, 1\}, \\
 &x_i \in [0, 1], \quad \forall x_i \in \mathbf{X}, i = 1, 2, 3, \dots, M, \\
 &y_j \in [0, 1], \quad y_j \in \mathbf{Y}, j = 1, 2, 3, \dots, N.
 \end{aligned} \tag{16}$$

It is worth noting that, in the inner minimization problem with uncertainty, the objective of the objective function is rewritten to solve the infimum problem due to the heterogeneity of investors' risk attitudes. The meaning of the second term of the constraint is the risk probability scenario under the worst scenario. When the disturbance term of the uncertain parameter fluctuates within the range

of the ellipsoid set, the adjustable safety parameter reflects the decision under the worst scenario. Its purpose is to solve the most stable investment income scenario, that is, the most stable income. In the process of calculation, the planning problem is more difficult due to the solution of multiple quadratic functions. Therefore, in the follow-up study, in the program involving a large number of calculations, computer

TABLE 1: Description of parameters and variables.

Annotation	Description
$x_i$	Continuous variables, $x_i \in [0, 1]$ $\mathbf{X} = \{x_1, x_2, \dots, x_i\}$ , represent the investment ratio;
$y_j$	Continuous variables, $y_j \in [0, 1]$ $\mathbf{Y} = \{y_1, y_2, \dots, y_j\}$ , represent the risk ratio;
$\alpha$	0-1 variable, $\alpha \in \{0, 1\}$ , indicates the existence or nonexistence of the return;
$\beta$	0-1 variable, $\beta \in \{0, 1\}$ , indicates whether the skewness be evaluated;
$\gamma$	0-1 variable, $\gamma \in \{0, 1\}$ , indicates whether the risk measure is selected or not;
$P_i$	Measure the probability of the $i$ th combination, $i = 1, 2, \dots, M$ ;
$r(\text{CVaR})_i$	Measure the CVaR combination of the investment ratio;
$R$	Measure the output ratio of in portfolio, $R = 1 + r$ ;
$r$	The profit rate;
$s_i$	Indicative stock skewness, $s_i \in \mathbf{S}$ ;
$G^{\max}$	Indicates the maximum number of combinations that can be selected;
$U^{\max}$	Indicates the maximum acceptable skewness;
$K^{\min}$	Indicates the lowest acceptable information entropy;
$r_{\text{CVaR}}^{\max}$	Indicate the maximum acceptable risk value;
$r_q$	Measure the maximum loss of the portfolio;
$\sigma_q$	Standard deviation of portfolio;
$z_\alpha$	The quintile of normal distribution at the confidence level of $\alpha$ ;
$\delta$	Indicates acceptable expected risk value;
$M$	$i = 1, 2, \dots, M$ ;
$N$	$i = 1, 2, \dots, N$ ;

software and existing commercial software are used to solve the problem. The package is used for solving calculations to ensure the accuracy of the obvious results.

#### 4. Numerical Experiment

Suppose that an investor conducts investment activities on a stock exchange and selects 10 securities stocks as alternatives through multiple comparisons in multiple securities markets. After rigorous data comparison and advantage analysis, the best investment plan combination is screened out in order to achieve the purpose of maximizing the comprehensive benefits. In the process of selecting securities and stocks, the following factors and principles need to be considered comprehensively. First, the ultimate goal of investors is to maximize profits, but, at the same time, they must also take into account their risk tolerance. In other words, investors are pursuing the maximization of profits on the basis of certain risk expectations. Second, the investor's investment assets are fixed, and the initial investment proceeds will be used again for investment. Third, the investment process is rational. When investors face the risk of resisting the initial threshold, the investment quota will no longer be added. According to the above basic elements and principles, this paper first uses data crawling technology to obtain the relevant database, then uses the computer to simulate and check the model, and finally obtains investment management recommendations and research enlightenments through a large number of comparisons.

**4.1. Data Collection.** In this section, we use actual financial securities market data to conduct simulation experiments to verify the effectiveness of the model. In the data acquisition stage, this section uses python to write the technical framework, *Python* 3.6, requests, math, time, pandas, matplotlib, pyecharts, statsmodels, scipy, jieba, and pylab.

Figure 2 describes the calculation steps of the data crawling technology. First, log in to the target web page, browse the relevant web page information (Eastern Fortune.com), and search for the online information of the target web page. Secondly, analyze the web page structure, view the source code of the web page, and click on the next page to observe the change of the url in the search bar of the browser. Thirdly, use the dictionary method to read the data directly, and save the read data to the preset file database. Next, use a loop strategy to crawl the data step by page, jump to the next page after completing the crawling of one page of data, and repeat the same data collection process. Finally, when the acquired data is stored in the database, the crawling process is terminated.

This section obtains historical data of historical stocks from a real website (Eastern Fortune.com). The data is obtained from the period from July 24, 2020, to December 20, 2020, about 1,000 statistical data generated by 100 exchanges of 10 stocks on the next day. Since the trading of stocks is only carried out on the trading day, the data sequence is sorted by time, but the time sequence is not continuous. In order to facilitate statistical calculations, this section sorts them according to the order of the trading days, which are 1 to 100 trading days. Table 2 describes the initial part of the data form of the obtained stock information. Before the numerical simulation, we conducted a simple statistical analysis of the historical profits of 10 stocks. Table 2 lists the average profit, standard deviation, skewness, and kurtosis of a single stock.

The relevant parameter settings in this section are as follows: when investors sell securities in the securities market, the ratio of securities lending margin shall not be less than 50%, so  $k = 0.5$ . In the process of stock trading, the transaction cost includes commission, stamp duty, and transfer fee, which is generally 0.3% of the transaction amount. Due to the existence of transaction fees, stock

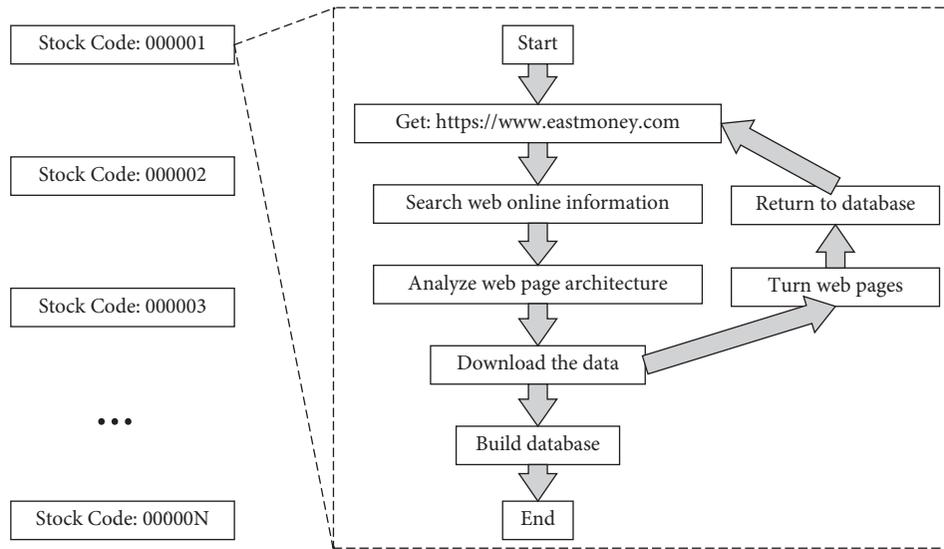


FIGURE 2: Data crawling process steps.

TABLE 2: Original stock data table (partial data).

Code	Date	Closing point	Opening.P	Highest.P	Lowest.P	Ups and downs	Quote	Volume
000001.SH	0724	3196.7684	3310.6449	3319.1268	3184.9645	128.3418	-3.8598	427054026
000001.SH	0723	3325.1102	3306.1489	3336.3002	3257.8269	-8.0533	-0.2416	407042499
000001.SH	0722	3333.1635	3315.1816	3381.9757	3311.7862	12.2688	0.3694	393335335
000001.SH	0721	3320.8947	3330.5467	3336.6791	3300.5719	6.7457	0.2035	359252029
000001.SH	0720	3314.149	3243.9105	3314.149	3220.684	100.0203	3.1119	418555306
000001.SH	0717	3214.1287	3214.4022	3252.7786	3181.2745	4.0301	0.1255	359652413
000001.SH	0716	3210.0986	3356.359	3373.5317	3209.7272	151.2058	-4.4984	490613123
000001.SH	0715	3361.3044	3422.0778	3432.4506	3345.7488	-53.3142	-1.5614	492030542
000001.SH	0714	3414.6186	3435.0237	3451.2224	3366.0828	-28.6677	-0.8326	543211492
000001.SH	0713	3443.2863	3379.3867	3458.7914	3369.0378	59.9641	1.7723	558006875
000001.SH	0710	3383.3222	3418.9347	3433.1085	3372.5066	-67.2713	-1.9496	556120011
...	...	...	...	...	...	...	...	...

trading should not be too frequent. This paper studies wind as three levels; one is based on a 30-day cycle, another is based on a 60-day cycle, and the third is based on a 90-day cycle. Analyze the relevance and influence of various indicators of stock trading. The profit rate of the stock is fitted according to historical parameters, and the fitted value of the historical data is used as the profit rate for analysis. On the assumption of the confidence interval, this paper, respectively, assumes 85%, 90%, 95%, and 99% confidence intervals and analyzes the parameters of each confidence interval.

Figure 3 depicts the volatility of the profit rate of the selected 10 stocks 100 times. It can be seen that the fluctuation of stocks is relatively large, which shows that the volatility of investment income is relatively large. It can also be found that there is no obvious law for the fluctuation of stocks, so there are no algorithms and tools that can accurately predict stock investment profits. Overall, the volatility of individual stocks is affected by objective economic factors, which is manifested in the overall change trend of the macro. From the above figure, it can be seen that the volatility of stocks has a certain time regularity; that is, there are overall ups and downs. The trend of. Therefore, in the investment process, it is necessary to take into account the

various factors of the investment, as well as to make a reasonable plan in combination with objective conditions and risk tolerance.

Figure 4 depicts the overall (100 trading days) average amount of fluctuations in 10 stocks. Only from the perspective of the average value, among the 10 stocks, the average value of volatility of 8 stocks is higher than 0, and the average value of volatility of 2 stocks is lower than 0. This shows that the overall stock market at this stage shows a good development trend. Stockholders showed a positive attitude towards the improvement of economic conditions and actively invested funds in the stock market.

4.2. Model Algorithm Design. In this paper, based on the Matlab (R2016a) platform, the algorithm framework is designed, and the algorithms of the two-stage basic model and the TS-RO model under uncertain conditions are constructed, respectively. Solve by calling the solver Gurobi (9-1-0). In order to ensure the rigor and scientificity of the experiment, all algorithms are operated in the same computing environment (Windows 10, Intel (R) Core (TM) i5-8300H CPU @2.3 GHz, RAM 8 GB, 512 G SSD). The

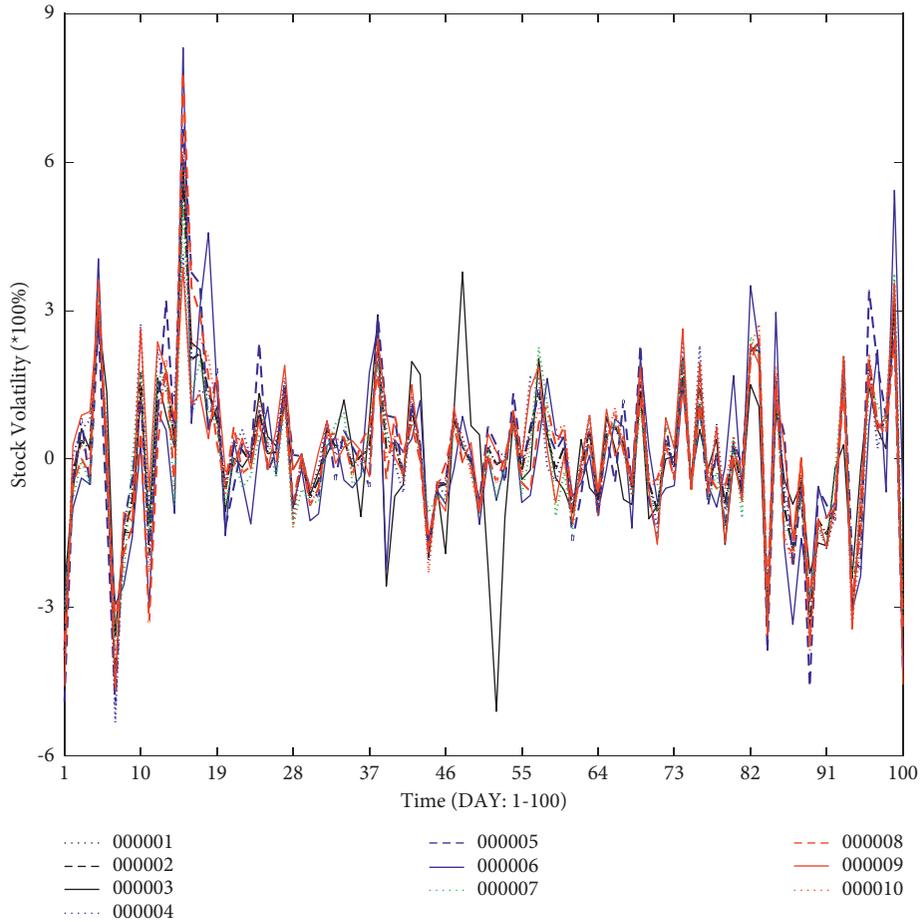


FIGURE 3: Volatility of stocks (100 Trades).

algorithm framework and calculation steps of the model are shown in Algorithm 1.

### 5. Sensitivity Analysis

**5.1. The Computational Efficiency of the Model.** In the same computing environment and the same computing scale, this section analyzes the factors affecting the running time of the model. After fixing the relevant parameters, the investment cycle is the only variable. We divide the time period (L/DAY) into 30, 60, and 90 trading days and compare the differences in the running time of the model. The calculation results are shown in Table 3. The analysis found that the length of the time period has a positive correlation with the calculation time; that is, the longer the time, the longer the calculation time of the model. Secondly, the calculation time of different models is slightly different. Among them, the calculation time of the equal-weight model is the shortest, the calculation time of the TS-MIP model is in the middle, and the calculation time of the TS-RO model is the longest. The increase in computing time reflects the increase in computing volume, which means that the computing volume and complexity of the TS-RO model are also the largest.

**5.2. Autocorrelation Analysis.** The long memory of financial securities asset profits refers to the significant autocorrelation of distant time intervals in the time series of profits; that is, historical events will continue to affect the future. Since stock securities data has obvious long memory, historical data information can be used to predict future profits and other related parameters. In this section, the parameters obtained are normalized, and the correlation effects can be compared more clearly. In this section, using the autocorrelation coefficient method, research and analysis found that the income of the industrial average index has significant long memory, as shown in Figures 5 and 6. The autocorrelation function is a measure of correlation, that is, a measure of similarity, which is the correlation between the function and the function itself. When there are periodic components in the function, the maximum value of the autocorrelation function can well reflect this periodicity. Cross-correlation is the similarity between two functions. When both functions have the same periodic component, the maximum value can also reflect this periodic component. From the perspective of linear space, correlation operation is actually an inner product operation, and the inner product of two vectors in linear space represents the projection of one vector to another vector, indicating the similarity of the

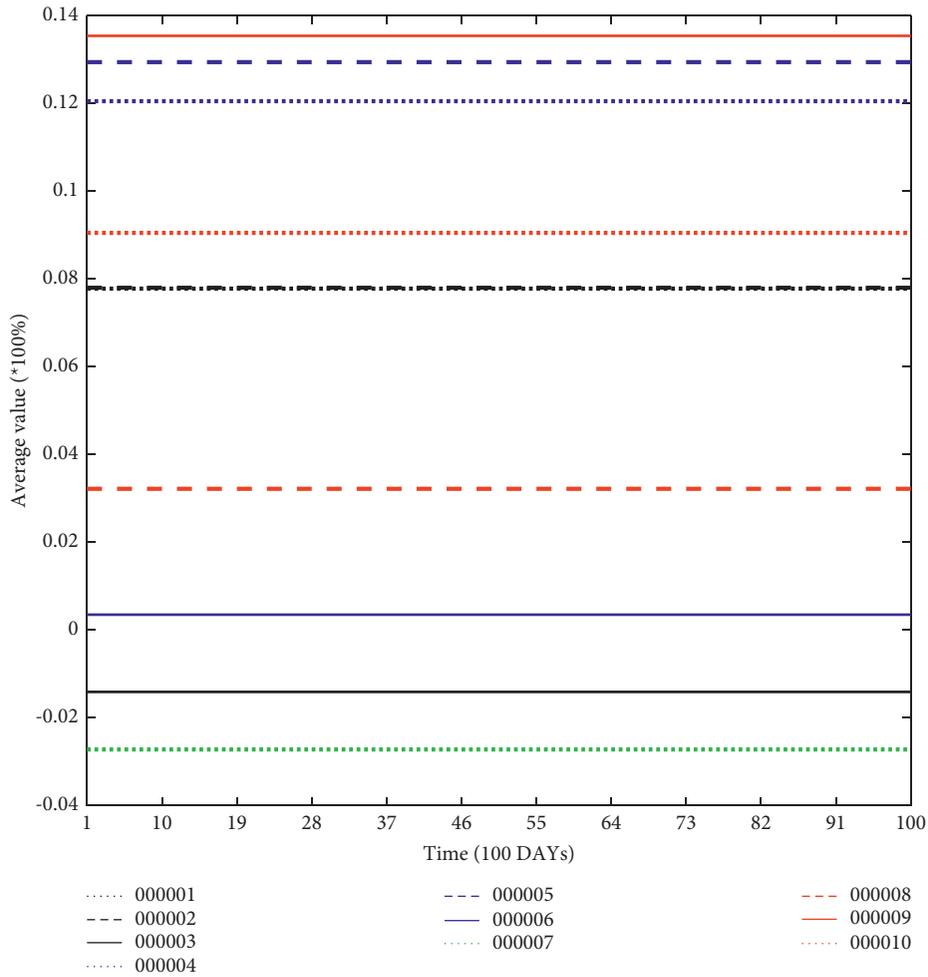


FIGURE 4: Average value of stock volatility.

```

(i) Input Measurement data
(ii) Initialization: Input the initial value; Setting parameter value boundary;
(iii) Step 1:
(iv)   For TS-MIP model: Define variable type; Set probability function;
(v)   else TS-RO model
(vi)   Define the variable type; Generate random number; Set the safety parameters;
(vii)  End
(viii) Step 2: While
(ix)   Input  $\{\varepsilon, \xi\} \in \{\cup_{B_j}, \cup_{B_j}\}$ , do step 3; Otherwise: re-step 1;
(x)   End while
(xi)  Step 3: Initialize
(xii) Input Parameter constraints; Variable constraints;
(xiii) End
(xiv) Step 4: Do for: Set the solution environment; Solved by Gurobi;
(xv)   If met step 3: Out the result and time
(xvi)   Else: profit step 2;
(xvii) End do
(xviii) Profit  $\{X^*\}$ 
    
```

ALGORITHM 1: Specific steps of models algorithm.

TABLE 3: CPU calculation time of three models under different time windows.

Period (L/DAY)	30 (s)	60 (s)	90 (s)
Equal weight method	1.983	3.781	8.093
TS-MIP model	2.081	4.234	7.781
TS-RO model	2.145	4.689	8.071

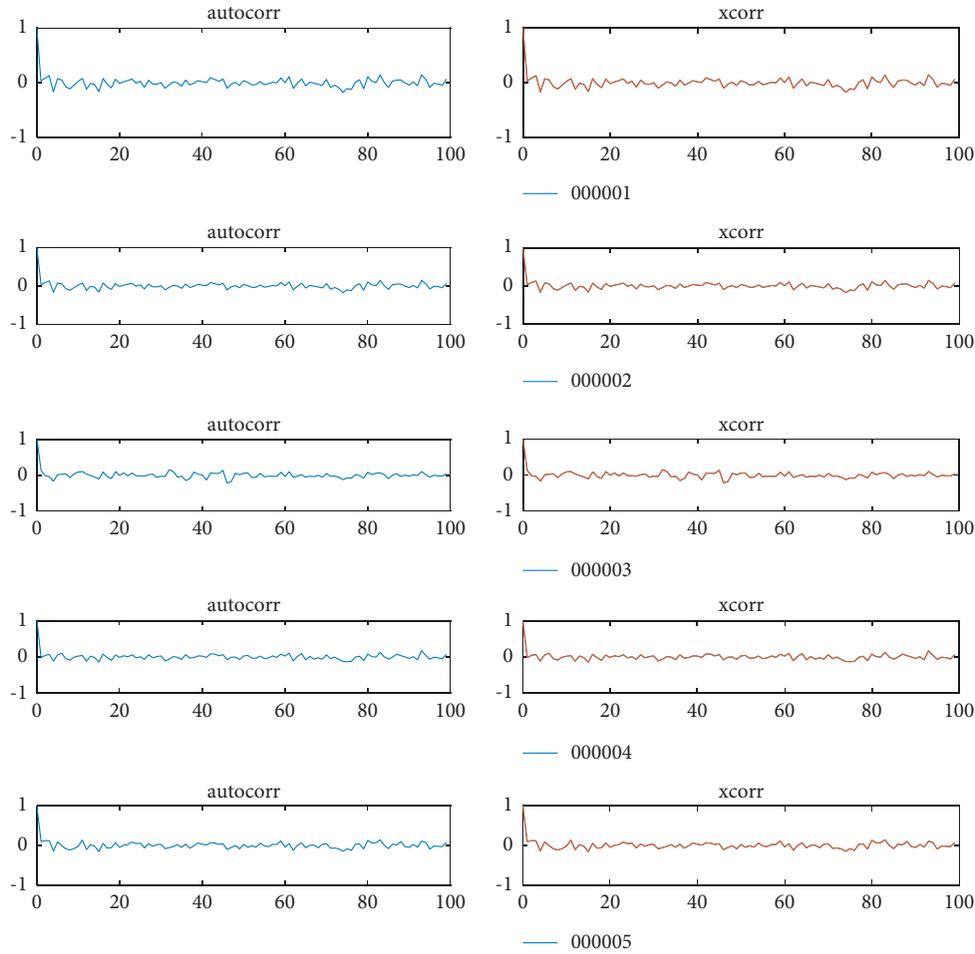


FIGURE 5: Autocorrelation analysis (Stocks 1-5).

two vectors, so the correlation operation reflects this kind of similarity.

From Figures 5 and 6, we can find that the fluctuation of the curve tends to be flat. For the autocorrelation graph test, stationary series have short-term correlation. This property indicates that, for stationary series, only recent series values usually have a more obvious influence on the current value. The farther apart the past value, the smaller the influence on the current value. As the number of delay periods increases, the autocorrelation coefficient of a stationary series will decay relatively quickly to zero and fluctuate randomly around zero, while the autocorrelation coefficient of a nonstationary series decays slowly. This is the use of autocorrelation graphs, the standard for the stationarity test.

5.3. Influence of Safety Parameters. To ensure the unbiased and scientific nature of the experiment, the parameters are

fixed. The model under the condition of deterministic information is the TS-MIP model. With the increase of security parameters, the parameters in the model such as rate of profit, diversity, skewness, information entropy, and CVaR are not affected. In the environment of uncertain risk preference, the model is the TS-RO model. The parameter diversity, skewness, and information entropy are not disturbed by uncertainty. Under the environment of uncertain risk appetite, the risk attitude of investment has changed. Here, through the method of robust optimization, the relevant risk preference range and boundary are constructed with the help of uncertain sets. The benefits of important parameters, CVaR, and total revenue are significantly affected by safety parameters.

Figure 7 describes the impact of security parameters on A: revenue and B: diversity. It can be seen from Figure 7 that, under deterministic information conditions, changes in

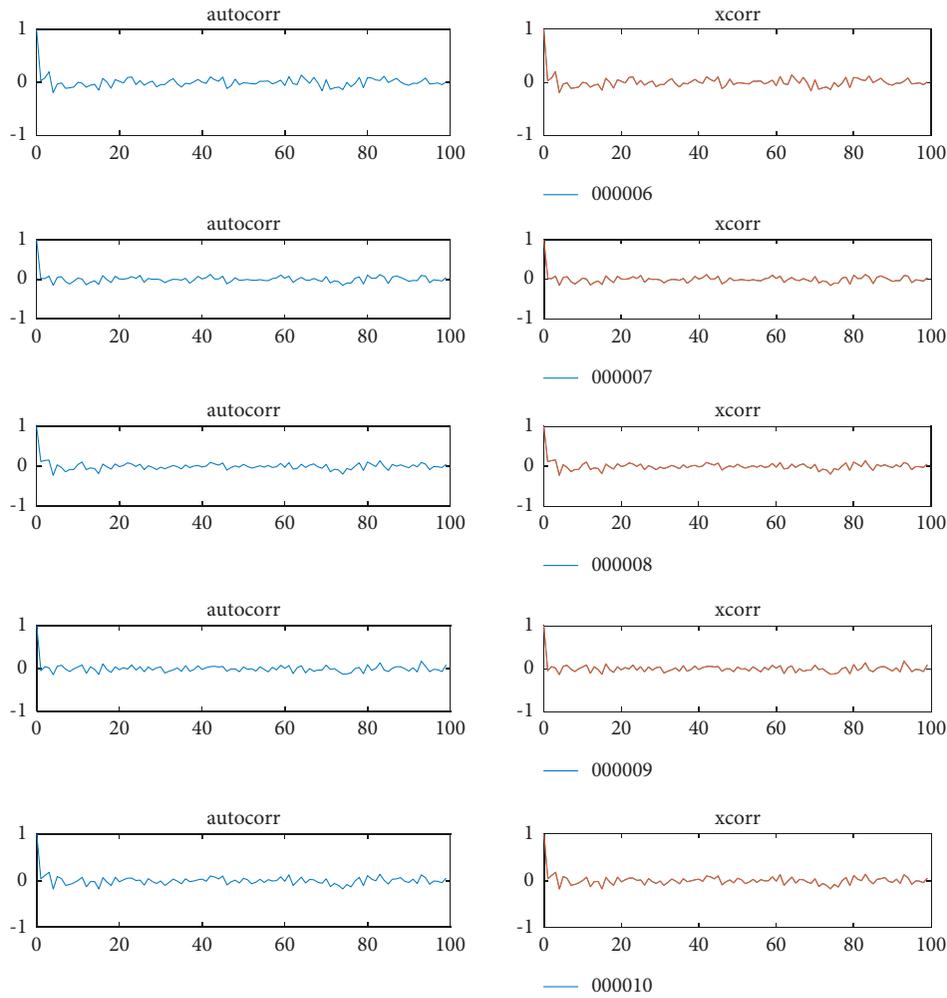


FIGURE 6: Autocorrelation analysis (Stocks 6-10).

security parameters will not have an impact on revenue and diversity. The figure shows that the influence curve of the TS-MIP model is a horizontal straight line. Of course, the deterministic TS-MIP model cannot cope with the problem of portfolio optimization under uncertain conditions. Under conditions of uncertain information, changes in security parameters have a positive impact on profits, but not on the diversity of investment portfolios. Specifically, as the level of security parameters increases, investors' investment income also shows an increasing trend. That is, when the risk is controllable, the rate of profit on investment is higher, and the investment ratio is larger.

Figure 8 describes the impact of security parameters on C: skewness and D: information entropy.

It can be seen from Figure 8 that, under deterministic information conditions, changes in security parameters will not have an impact on skewness and information entropy. The figure shows that the influence curve of the TS-MIP model is a horizontal straight line. Of course, the deterministic TS-MIP model cannot cope with the problem of portfolio optimization under uncertain conditions. Under uncertain conditions, changes in security parameters will

not have an impact on skewness and information entropy. This shows that it is not the controllability of risk that affects investor confidence, but other deeper factors.

Figure 9 describes the impact of safety parameters on E: CVaR and F: total benefits. From the change trend of the curve in Figure 9, it can be found that, under certain information conditions, changes in safety parameters will not have an impact on CVaR and overall benefits. The figure shows that the influence curve of the TS-MIP model is a horizontal straight line. Under uncertain conditions, changes in safety parameters will have a significant impact on CVaR and overall benefits. As the security parameters increase, the risk of investment shows an increasing trend, which is often referred to as a robust price, which is the price paid to resist risks. But, at the same time, the increase in safety parameters also makes the overall benefit show an increasing trend, which can increase investors' investment confidence when they choose investment.

5.4. *Testing the Effectiveness of the Portfolio.* For comparison, three financial ratios are listed below to measure the pros and

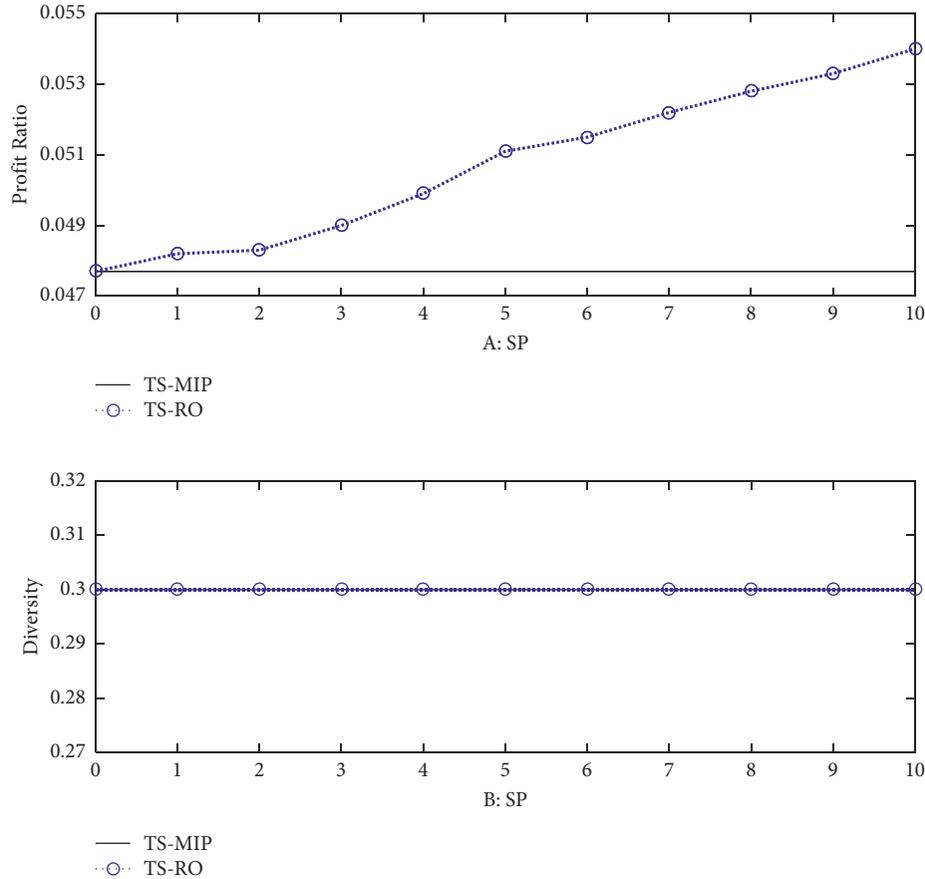


FIGURE 7: Revenue and diversity in SP.

cons of stocks and securities. The Sharpe ratio (ShR) is used to help investors understand the profit on investment related to their risk. Unlike the ShR, the Sortino ratio (SoR) is used to evaluate the excess profit of a portfolio given a downside risk. The Treynor ratio (TrR) represents a risk-adjusted profit based on systemic risk. Generally speaking, the greater the value of the Sharpe/Sortino/Treynor ratio, the greater the risk-adjusted profit, and the more attractive the risk-adjusted profit for rational investors.

ShR is an indicator proposed by William to research and measure fund performance [42]. This indicator is not only a standardized indicator for fund performance evaluation, it has been successfully applied in many fields as well. The purpose of the ShR is to calculate how much excess profit will be generated per unit of total risk assumed by the portfolio. When all assets in the portfolio are risk assets, the Sharpe ratio applies. The formula for calculating the ShR is

$$\text{ShR} = \sum_i \frac{E(Rp_i) - R_f}{\sigma p_i}, \quad i \in I. \quad (17)$$

Among them,  $E(Rp_i)$  is the expected annualized profit of the investment portfolio,  $R_f$  is the annualized risk-free interest rate, and  $\sigma p_i$  is the standard deviation of the

annualized profit of the investment portfolio. The Sharpe ratio represents how much excess profit investors can get for each additional risk point. If it is greater than 1, it means that the fund's profit is higher than the volatility risk; if it is less than 1, it means that the operational risk of the fund is greater than the rate of profit.

In this way, the Sharpe ratio of each investment portfolio can be calculated, that is, the ratio of investment profits to more risks.

The higher the Sharpe ratio, the better the portfolio. Through experimental verification, it is found (Table 4) that when  $L = 30$ , the equal weight method is the lowest (11), the TS-MIP model is in the middle ( $\text{ShR} = 0.5004$ ), and the TS-RO model is the best ( $\text{ShR} = 0.5407$ ). Compared with the equal weight method, the Sharpe value of the TS-RO model increased by 18.97%. In addition, in each model, as the time period increases ( $L = 30 \rightarrow L = 60 \rightarrow L = 90$ ), the Sharpe ratio gradually increases. This law shows that long-term investment is due to short-term investment, which can provide decision-making suggestions for investors' investment cycle planning.

SoR is the standard for quantitatively dealing with the trade-off between the profit and risk of investment plans and

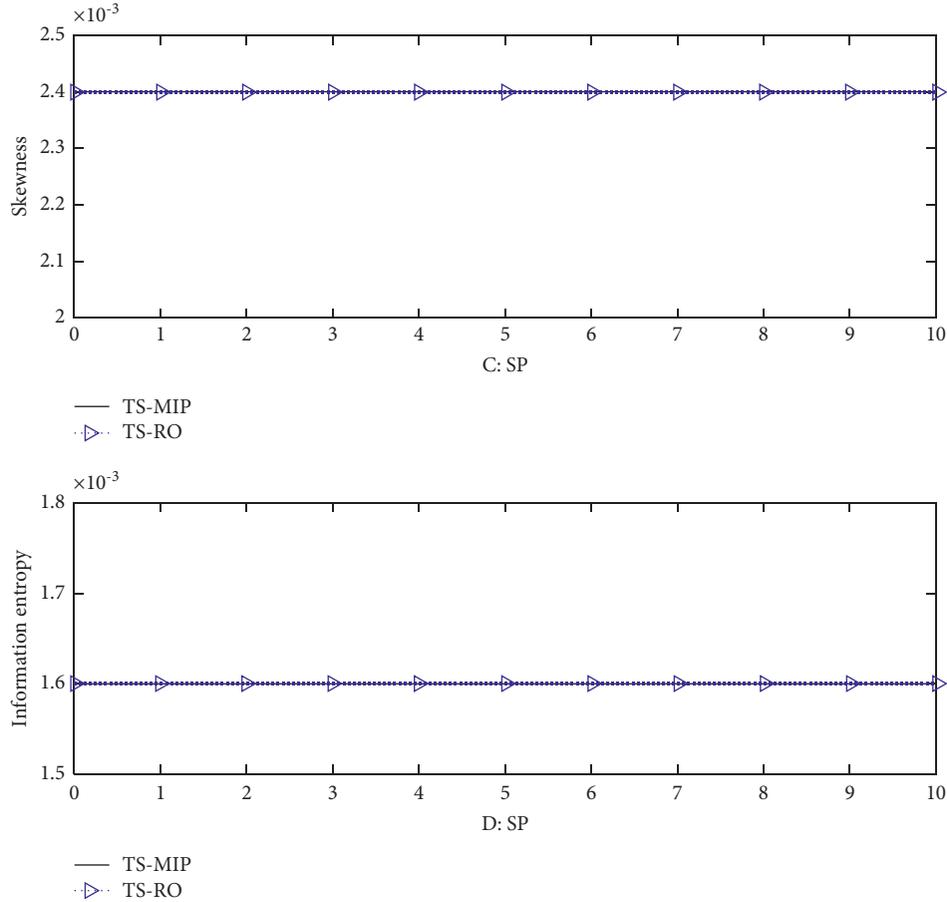


FIGURE 8: Information entropy and skewness in SP.

the final selection of investment plans. Its calculation formula is as follows:

$$SoR = \sum_i \frac{R_{pi} - r_i}{\delta_i(R, MARR)}, \tag{18}$$

$$\delta_i(R, MARR) = \sqrt{E\{[\min(R - MARR, 0)]^2\}}, \quad i \in I.$$

Among them,  $R_{pi}$  is the profit rate of the investment portfolio, which is replaced by the average value of the profit rate of the investment plan obtained by the simulation.  $r_i$  is the minimum profit required by investors of 2%.  $\delta_i(R, MARR)$  is the risk that the actual rate of profit is lower than the average value, that is, the downside risk of the rate of profit, MARR is the lowest acceptable rate of profit, and MARR can be the risk-free rate or 0. The SoR is a measure that is more in line with investors who are more sensitive to the decline in asset value. The higher the ratio, the higher the profit rate by the fund if it bears the same unit of downside risk.

The result verification in Table 5 found that, taking  $L = 60$  as an example, the equal weight method is the lowest

(SoR = 0.4758), the TS-MIP model is in the middle (SoR = 0.5012), and the TS-RO model is the best (SoR = 0.5578). Compared with the equal weight method, the SoR of the TS-RO model has an increase of 8.20%. In addition, in each model, as the time period grows ( $L = 30 \rightarrow L = 60 \rightarrow L = 90$ ), the SoR is gradually increasing, further verifying the conclusions obtained by the Sharpe Ratio. It should be pointed out that risk is the uncertainty of the result. This uncertainty includes situations where the actual rate of profit is lower than expected and situations where the actual rate of profit is higher than expected. From the perspective of project investors, it is undoubtedly beneficial for the actual rate of profit to be higher than the expected value (i.e.,

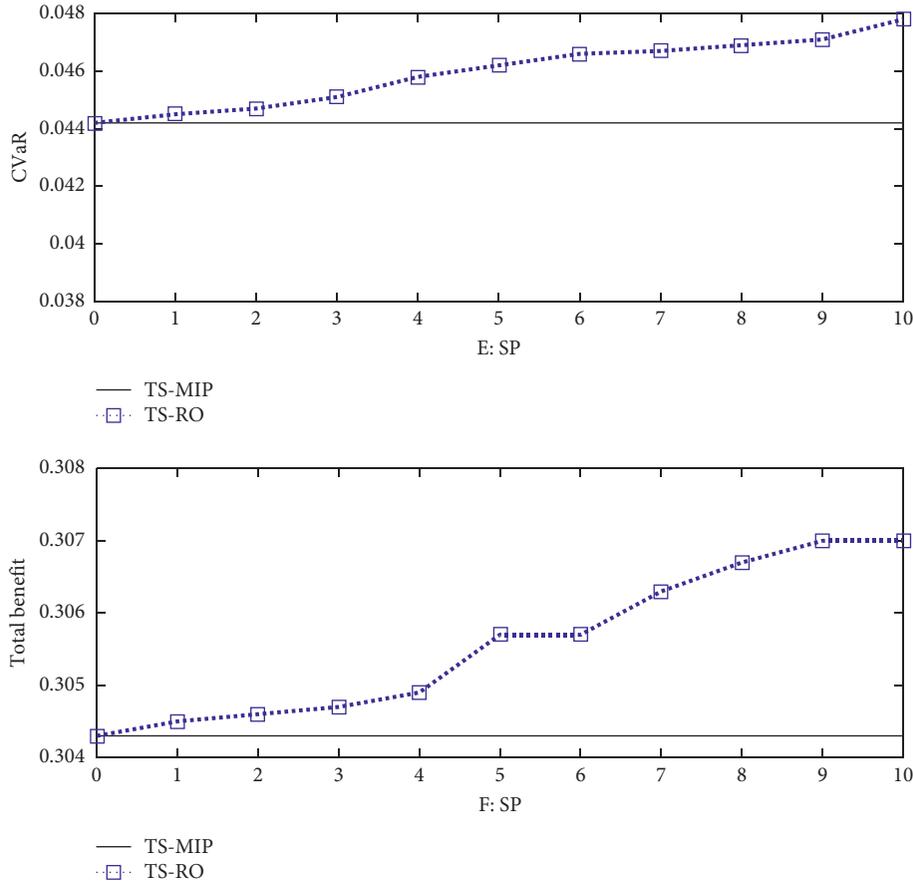


FIGURE 9: The impact of SP on CVaR and presidential benefits.

TABLE 4: Sharpe ratio under different time window lengths.

Period (L/DAY)	30	Increase	60	Increase	90	Increase
Equal weight method	0.3510	0	0.3654	0	0.3789	0
TS-MIP model	0.5004	0.1494	0.5137	0.1483	0.5245	0.1456
TS-RO model	0.5407	0.1897	0.5546	0.1892	0.5719	0.1930

TABLE 5: SoR of the model under different time window lengths.

Period (L/DAY)	30	Increase	60	Increase	90	Increase
Equal weight method	0.4514	0	0.4758	0	0.4985	0
TS-MIP model	0.4904	0.0390	0.5012	0.0254	0.5176	0.0191
TS-RO model	0.5397	0.0883	0.5578	0.0820	0.5810	0.0825

average) when making investment decisions. This uncertainty brings unexpectedly high profits to investors. Therefore, we should not pay attention to these risks in the investment but should pay attention to the risk that the actual rate of profit is lower than the expected value of the investment decision.

TrR is the ratio of the value of the fund’s profit rate exceeding the risk-free interest rate to the systemic risk. TrR measures the excess profit that the fund receives from the systemic risk of the unit. The larger the TrR, the higher the risk-adjusted profit. The TrR is based on the fact that nonsystematic risks have been completely diversified; that is, it is believed that the portfolio of assets held by the fund has

fully diversified the risks of individual stocks or industries. TrR is suitable for evaluating nonsystematic risk of completely diversified funds, such as large-cap index funds.

$$TrR = \sum_i \frac{E(R_p) - R_f}{\beta_{pi}}, \tag{19}$$

$$\beta_{pi} = \frac{Cov(r_i, r_m)}{\sigma_m^2}, \quad \forall i \in I.$$

The numerator of the Treynor ratio is still the excess profit of the asset portfolio,  $E(R_{p_i})$  is the expected

TABLE 6: The TrR of different models.

Period (L/Day)	30	Increase	60	Increase	90	Increase
Equal weight method	0.3514	0	0.3758	0	0.3985	0
TS-MIP model	0.4187	0.0673	0.4256	0.0498	0.4367	0.0382
TS-RO model	0.4260	0.0746	0.4635	0.0877	0.4719	0.0734

annualized profit of the portfolio,  $R_f$  is the annualized risk-free interest rate, and the denominator becomes systemic risk  $\beta_{pi}$ . Its economic meaning is that every unit of systemic risk is assumed. For risk,  $r_i$  represents the profit of a single stock,  $r_m$  represents the market profit, and  $\sigma_m^2$  represents the variance of the market profit.

As shown in Table 6, through experimental verification, when  $L = 90$ , the equal weight method is the lowest (TrR = 0.3985), the TS-MIP model is in the middle (TrR = 0.4367), and the TS-RO model is the best (TrR = 0.4719). Compared with the equal weight method, the TrR of the TS-RO model increased by 7.34%. In addition, in each model, as the time period increases ( $L = 30 \rightarrow L = 60 \rightarrow L = 90$ ), TrR is gradually increasing, which further verifies the conclusions obtained by ShR and TrR. Regarding the excess profit the investment portfolio can get, the greater the value of the TrR, the stronger the ability of the fund manager. The TrR applies to a fully diversified investment portfolio, at which time nonsystematic risks have been completely diversified, and only systemic risks in the market are considered [43,44].

## 6. Conclusion

Research on investment portfolio issues can provide investors with more robust investment plans, and the uncertainty in it is a key factor affecting investment performance. Aiming at the investment portfolio problem, this paper comprehensively considers the five dimensions of profit, diversity, skewness, information entropy, and conditional risk value and constructs a two-stage mixed integer programming (TS-MIP) model. However, the deterministic TS-MIP model cannot cope with the problem of uncertain information. Therefore, this paper constructs a two-stage robust optimization (TS-RO) model by introducing robust optimization theory.

In terms of case experiments, this paper uses data crawler technology to obtain actual data from real websites for analysis and verifies the effectiveness of the proposed model in dealing with uncertain problems. The comparison of models found that, compared with the traditional equal weight model, the investment benefits of the TS-MIP model and the TS-RO model proposed in this paper have been improved. Among them, the Sharpe ratio, Sortino ratio, and Treynor ratio have the largest increase of 19.30%, 8.25%, and 7.34%, respectively. Combined with the results of empirical analysis, this paper analyzes the investment decisions of investors and puts forward some suggestions to investors.

The content and contributions of this paper mainly include the following: first, in response to the uncertainty of the rate of profit, this paper uses historical data as the basis to predict the rate of profit and express the uncertainty

scenario. Second, after the analysis of the mean-variance model, this paper uses the two-stage integer programming theory to establish a portfolio model that maximizes profits, minimizes risks, and considers both profits and risks. Third, the empirical analysis of the model using the data of 10 stock indexes in my country's stock market shows that the model is feasible and effective, giving investment decision-makers a simple, clear, and intuitive guidance. Fourth, relevant suggestions are made to investors. Investors should strengthen their ability to obtain relevant information and be able to identify and analyze the information and then use relevant technologies to predict stock profits. Moreover, conservative, prudent, and aggressive investors must consider their own circumstances when investing and should not invest blindly.

## Data Availability

No data were used to support this study.

## Conflicts of Interest

The authors declare that they have no conflicts of interest.

## References

- [1] M. M. Boyer, E. P. Cowins, and W. D. Reddic, "Portfolio rebalancing behavior with operating losses and investment regulation," *International Review of Economics & Finance*, vol. 63, pp. 313–328, 2019.
- [2] K. S. Peskin, "Evaluating multi-asset strategies," *Journal of Portfolio Management*, vol. 44, no. 2, pp. 40–49, 2017.
- [3] D. Amiram and M. M. Frank, "Foreign portfolio investment and shareholder dividend taxes," *The Accounting Review*, vol. 91, no. 3, pp. 717–740, 2016.
- [4] X. Gao, T. J. Wong, L. Xia, and G. Yu, "Network-induced agency conflicts in delegated portfolio management," *The Accounting Review*, vol. 96, no. 1, pp. 171–198, 2021.
- [5] C. Lian, Y. Ma, and C. Wang, "Low interest rates and risk-taking: evidence from individual investment decisions," *Review of Financial Studies*, vol. 32, no. 6, pp. 2107–2148, 2019.
- [6] K. C. Wong, S. C. P. Yam, and J. Zeng, "Mean-risk portfolio management with bankruptcy prohibition," *Insurance: Mathematics and Economics*, vol. 85, pp. 153–172, 2019.
- [7] Y. Zhou, Y. Zhang, and M. Goh, "Choice of pricing and advertising schemes for a two-sided platform," *Managerial and Decision Economics*, vol. 42, no. 7, pp. 1865–1885, 2021.
- [8] P. C. Fishburn, "Foundations of risk measurement. I. Risk as probable loss," *Management Science*, vol. 30, no. 4, pp. 396–406, 1984.
- [9] Y. Zhou, Y. Zhang, and M. I. M. Wahab, "Optimal pricing and choice of platform advertising schemes considering across-side network effect," *Managerial and Decision Economics*, 2021.

- [10] A. Alzaatreh, F. Famoye, and C. Lee, "Weibull-Pareto distribution and its applications," *Communications in Statistics - Theory and Methods*, vol. 42, no. 9, pp. 1673–1691, 2013.
- [11] P. Armour, R. V. Burkhauser, and J. Larrimore, "Using the Pareto distribution to improve estimates of topcoded earnings," *Economic Inquiry*, vol. 54, no. 2, pp. 1263–1273, 2016.
- [12] L. Garlappi, R. Uppal, and W. Tan, "Portfolio selection with parameter and model uncertainty: a multi-prior approach," *Review of Financial Studies*, vol. 20, no. 1, pp. 41–81, 2019.
- [13] J. Tu and G. Zhou, "Incorporating economic objectives into bayesian priors: portfolio choice under parameter uncertainty," *Journal of Financial and Quantitative Analysis*, vol. 45, no. 4, pp. 959–986, 2010.
- [14] E. Bayraktar and Z. Zhou, "On arbitrage and duality under model uncertainty and portfolio constraints," *Mathematical Finance*, vol. 27, no. 4, pp. 988–1012, 2017.
- [15] V. Golosnoy and Y. Okhrin, "General uncertainty in portfolio selection: a case-based decision approach," *Journal of Economic Behavior & Organization*, vol. 67, no. 3, pp. 718–734, 2008.
- [16] N. Barberis and R. Thaler, "A survey of behavioral finance," *NBER Working Papers*, vol. 2, no. 03, pp. 1053–1128, 2002.
- [17] X. Li and Z. Qin, "Interval portfolio selection models within the framework of uncertainty theory," *Economic Modelling*, vol. 41, pp. 338–344, 2014.
- [18] X. Huang and L. Qiao, "A risk index model for multi-period uncertain portfolio selection," *Information Sciences*, vol. 217, pp. 108–116, 2012.
- [19] D. Bertsimas and M. Sim, "The price of robustness," *Operations Research*, vol. 52, no. 1, pp. 35–53, 2004.
- [20] A. Ben-Tal and A. Nemirovski, "Selected topics in robust convex optimization," *Mathematical Programming*, vol. 112, no. 1, pp. 125–158, 2007.
- [21] D. Huang, S. Zhu, F. J. Fabozzi, and M. Fukushima, "Portfolio selection under distributional uncertainty: a relative robust cvar approach," *European Journal of Operational Research*, vol. 203, no. 1, pp. 185–194, 2010.
- [22] Z. Dai and F. Wen, "Robust cvar-based portfolio optimization under a genal affine data perturbation uncertainty set," *Journal of Computational Analysis and Applications*, vol. 16, no. 1, pp. 85–92, 2006.
- [23] R. Agrawal, "Finite-sample concentration of the multinomial in relative entropy," *IEEE Transactions on Information Theory*, vol. 99, pp. 1–10, 2020.
- [24] W. Chang, "Supply Chain Movement Risk in the Sneaker Industry: An Empirical Study," *Quality & Quantity*, pp. 1–20, 2021.
- [25] M. Kamiński, "On Shannon entropy computations in selected plasticity problems," *International Journal for Numerical Methods in Engineering*, vol. 122, no. 5617, 2021.
- [26] Y. Dai, J. Hu, Z. Zhang, C. Zhang, Y. Dong, and X. Wang, "Measurement-induced entropy increment for quantifying genuine coherence," *Quantum Information Processing*, vol. 20, no. 8, pp. 1–12, 2018.
- [27] X. Liu, A. Jiang, N. Xu, and J. Xue, "Increment entropy as a measure of complexity for time series," *Entropy*, vol. 18, no. 1, pp. 1–14, 2016.
- [28] N. Khajonchotpanya, Y. Xue, and N. Rujeerapaiboon, "A revised approach for risk-averse multi-armed bandits under CVaR criterion," *Operations Research Letters*, vol. 49, no. 4, pp. 465–472, 2021.
- [29] J. Li, B. Xin, P. M. Pardalos, and J. Chen, "Solving bi-objective uncertain stochastic resource allocation problems by the cvar-based risk measure and decomposition-based multi-objective evolutionary algorithms," *Annals of Operations Research*, vol. 296, no. 1-2, pp. 639–666, 2021.
- [30] C. Yu and Y. Liu, "A personalized mean-cvar portfolio optimization model for individual investment," *Mathematical Problems in Engineering*, vol. 2021, pp. 1–12, 2021.
- [31] Y. Li and J. Ou, "Replenishment decisions for complementary components with supply capacity uncertainty under the cvar criterion," *European Journal of Operational Research*, vol. 297, no. 3, pp. 904–916, 2021.
- [32] S. Hosseini and M. Verma, "Equitable routing of rail hazardous materials shipments using cvar methodology," *Computers & Operations Research*, vol. 129, no. 1, Article ID 105222, 2021.
- [33] M. G. Mooselu, M. R. Nikoo, P. H. Bakhtiari, N. B. Rayani, and A. Izady, "Conflict resolution in the multi-stakeholder stepped spillway design under uncertainty by machine learning techniques," *Applied Soft Computing*, vol. 110, Article ID 107721, 2021.
- [34] R. T. Rockafellar and J. Sun, "Solving monotone stochastic variational inequalities and complementarity problems by progressive hedging," *Mathematical Programming*, vol. 174, no. 1-2, pp. 453–471, 2019.
- [35] D. Kong, L. Liu, and Y. Wu, "Optimal reinsurance under risk and uncertainty on orlicz hearts," *Insurance: Mathematics and Economics*, vol. 81, pp. 108–116, 2018.
- [36] X. Jiang, Y. Yang, Y. Lu, and M. Cao, "Flattened aggregate function method for nonlinear programming with many complicated constraints," *Numerical Algorithms*, vol. 86, no. 1, pp. 103–122, 2021.
- [37] G. A. Hanasusanto, D. Kuhn, S. W. Wallace, and S. Zymler, "Distributionally robust multi-item newsvendor problems with multimodal demand distributions," *Mathematical Programming*, vol. 152, no. 1-2, pp. 1–32, 2015.
- [38] Y. Shi, T. Boudouh, and G. Olivier, "A robust optimization for a home health care routing and scheduling problem with consideration of uncertain travel and service times," *Transportation Research Part E: Logistics and Transportation Review*, vol. 128, pp. 52–95, 2019.
- [39] C. Valentina, J. Qi, and L. Yang, "Robust optimization models for integrated train stop planning and timetabling with passenger demand uncertainty," *Transportation Research Part B: Methodological*, vol. 136, pp. 1–29, 2020.
- [40] A. M. Mohammad, A. Abdel, and Z. Shokri, "Robust optimization for selective newsvendor problem with uncertain demand," *Computers & Industrial Engineering*, vol. 135, pp. 838–854, 2019.
- [41] C. André and G. Marc, "Minmax regret combinatorial optimization problems with ellipsoidal uncertainty sets," *European Journal of Operational Research*, vol. 258, no. 1, pp. 58–69, 2017.
- [42] Y. Li and S. Chung, "Ride-sharing under travel time uncertainty: robust optimization and clustering approaches," *Computers & Industrial Engineering*, vol. 149, no. 9, Article ID 106601, 2020.

- [43] G. Ming and O. Hui, "Alpha Decay and Sharpe Ratio: Two Measures of Investor Performance," *Economic Modelling*, vol. 104, no. C, Article ID s0264999321001474, 2021.
- [44] C. Suresh and M. E. Srivastava, "Investigating a new methodology for ranking international mutual funds," *Journal of Economics and Finance*, vol. 18, no. 3, pp. 241–260, 1994.