## OPTIMIZATION



# A new uncertain enhanced index tracking model with higher-order moment of the downside

Tingting Yang<sup>1,2</sup> · Xiaoxia Huang<sup>2</sup> · Kwon Ryong Hong<sup>2,3</sup>

Accepted: 16 April 2023 / Published online: 19 May 2023 © The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2023

#### Abstract

Enhanced index tracking (EIT) problem is concerned with selecting a tracking portfolio to beat the benchmark on return while having the minimum tracking error. This paper addresses the EIT problem based on uncertainty theory where stock returns are treated as uncertain variables instead of random variables. Under the framework of uncertainty theory, the paper proposes a new uncertain EIT model where the higher-order moment of the downside is used as the tracking error measure, as higher-order moment makes the model more widely applicable and the downside risk is in line with investors' perception of risk. Besides, some realistic constraints are considered in the new uncertain EIT model. Then, the properties of the proposed model are discussed. To solve the model, we proposed, which is a nonlinear integer programming problem, a meta-heuristic algorithm presented. The efficiency of the algorithm and the applications of the proposed model are illustrated through numerical experiments.

Keywords Portfolio selection  $\cdot$  Enhanced index tracking optimization model  $\cdot$  Uncertain programming  $\cdot$  Higher-order moment

# **1** Introduction

The enhanced index tracking strategy is to select a tracking portfolio to beat the benchmark on return while having the minimum tracking error. In recent years, the enhanced index funds perform well and their scale grow rapidly (Jorion 2002; Filippi et al. 2016), so the EIT strategy is highly praised in the investment industry. In financial academia, the EIT problem has also attracted scholars' attention. One central research question involves the definition of the tracking error and the corresponding construction of the EIT model. In an important

 Xiaoxia Huang hxiaoxia@manage.ustb.edu.cn
 Tingting Yang oyanglin@163.com
 Kwon Ryong Hong hongkwonryong@163.com

- <sup>1</sup> Wujinglian School of Economics, Changzhou University, Changzhou 213164, China
- <sup>2</sup> School of Economics and Management, University of Science and Technology Beijing, Beijing 100083, China
- <sup>3</sup> Institute of Natural Sciences, Kim Il Sung University, Pyongyang, Democratic People's Republic of Korea

work, Roll (1992) first uses the variance of the difference between tracking portfolio returns and benchmark returns (TEV) as the tracking error measure and proposes the TEV criterion to build the EIT model. Then, many scholars have developed this approach further by, e.g., adding a constraint on the tracking portfolio variance that investors can take to reduce risk (Jorion 2003).

However, variance as the tracking error measure has been challenged, as it implies that investors weigh the underperformance and overperformance of their tracking portfolios relative to the benchmark equally. Actually, in the EIT strategy, investors welcome the overperformance and do not want the underperformance. So the symmetric tracking error measures are no longer suitable. For this reason, some scholars begin to use the downside risk measure. The downside risk measure is consistent with the investor's perception of risk because most investors understand risk as the potential underachievement of a target return by an asset (Mao 1970). Among many downside risk measures, the absolute downside deviation has become a popular one in the EIT problem. For example, Rudolf et al. (1999) use the absolute downside deviation as the tracking error measure to propose a linear EIT model and compare it with EIT models using other tracking error measures. Lejeune (2012) provides a game theoretical formulation for the EIT problem in which the minimum excess return over all allowable probability distributions is maximized subject to the absolute downside deviation tracking error constraint. Gaustaroba et al. (2016) put forward an EIT model in which they maximize the ratio of the upside and downside deviations between the tracking portfolio and the benchmark returns. In addition to the absolute downside deviation tracking error measure, the higher-order moment of the downside is also used as a tracking error measure in the literature. Higher-order moment contains more information, and it can make the model more widely applicable (Maringer and Parpas 2009). At present, few studies focus on the EIT problems involving the higher-order moment of the downside as the tracking error measure. To the best of our knowledge, we only find that Yu et al. (2006) have employed the higherorder moment of the downside as a tracking error measure in their model.

Realistic constraints are often considered when constructing the EIT models. For example, Beasley et al. (2003) take into account the transaction costs and the number of stocks that can be purchased in their tracking model. Strub and Baumann (2018) present a mixed-integer linear programming formulation with constraints of the capital budget, transaction costs, and the number of stocks that can be purchased. Besides, Andriosopoulos and Nomikos (2014) consider the realistic constraints of cardinality and min./max.weight when reproducing the performance of a geometric average Spot Energy Index. SantAnna et al. (2017) consider the realistic constraint of cardinality when investigating the out-of-sample performance of cointegration and correlation methods for the EIT strategies. When the EIT models contain realistic constraints, it is difficult to give the analytical solution of the proposed model. So scholars use algorithms to solve their EIT models. Guastaroba and Speranza (2012) present a heuristic framework called Kernel Search to solve their mixed-integer linear programming formulations for the EIT problem. Filippi et al. (2016) modify the Kernel Search heuristic to solve the bi-objective programming they proposed for the EIT problem.

In the above studies, probability theory is used as a main mathematical tool. Though probability theory is a powerful tool for helping people deal with indeterminate parameters, its use is conditional upon there being enough valid data so that obtained distributions from the data are close enough to the real frequencies. Yet there exist situations in reality where people have few data or the data are invalid due to unexpected events in society. For example, the outbreak of COVID-19 has led to a series of unexpected events so that historical data cannot reflect the future effectively. Besides, some empirical tests show that in many cases observed data in the financial market are not random variables. By using the two-sample Kolmogorov–Smirnov test, Ye and Liu (2022a) show US Dollar to Chinese Yuan (USD-CNY) exchange rates cannot be treated as random variables because residuals from different parts are neither from the same population nor white noise in the sense of probability theory. Similar tests also show that Alibaba stock prices (Liu and Liu 2022) and GDP (Ye and Liu 2022b) are not suitable to be treated as random variables. Therefore, there does exist a situation where probability theory is not applicable in reality.

To handle this problem, Liu (2007) proposes a new theory, i.e., uncertainty theory. Uncertainty theory is a new branch of mathematics and is developed based on four axioms. Nowadays, uncertainty theory has been applied in solving various optimization problems (Liu 2010; Gao 2011). Particularly, Huang (2010) is the first to use uncertainty theory to study portfolio selection systematically. Subsequently, scholars have studied many portfolio problems based on uncertainty theory. Wang and Huang (2019) discuss the effect of options on the portfolio performance and find that portfolios with options gain higher expected returns than those without options. Huang and Yang (2020) study how the background risk affects investment decisions. They give properties of the portfolio efficient frontier when stock and background asset returns all take normal uncertainty distributions and tell how background risk affects portfolio selection decisions. To reflect different attitudes toward risk in one portfolio investment, Huang and Di (2020) apply mental accounts to an uncertain optimization model. Besides, Ning et al. (2015) consider triangular entropy in the uncertain portfolio selection problem. Qin et al. (2016) study the uncertain portfolio adjusting problem. Zhang (2019), and Chen et al. (2019) focus on multi-period uncertain portfolio selection problem. Besides, scholars further discuss the performances of uncertain portfolios. Studies show that the returns or the expected returns of uncertain portfolios are higher than those of traditional portfolios in some situations. For example, Huang and Yang (2020) present evidence that the return of the portfolio got from the uncertain mean-variance model is greater than that of traditional Markowitz's mean-variance model in volatile markets. Xue et al. (2019) find that the expected return of the aggregate portfolio when security return rates are uncertain variables is greater than the expected return of the aggregate portfolio when security return rates are random variables in the case of facing invalid data. It can be seen from the above examples that the uncertain portfolio selection models have some advantages. Yet so far, no work deals with the uncertain EIT problem. How to build the uncertain EIT model? How does the uncertain EIT model perform? The lack of research on these issues motivates us to do the study.

The rest of the paper is organized as follows. In Sect. 2, we provide the necessary knowledge of uncertainty theory for easy understanding of our paper. In Sect. 3, we propose an uncertain EIT model with higher-order moment of the downside and discuss its properties. In Sect. 4, we present a meta-heuristic algorithm for the proposed model. In Sect. 5,

we report the experimental results. Section 6 concludes the paper.

# 2 Preliminaries

In this section, we review some necessary knowledge about uncertainty theory.

**Definition 1** (Liu 2007) Let  $\Gamma$  be a nonempty set, and L a  $\sigma$ algebra over  $\Gamma$ . Each element  $\Lambda \in L$  is called an event. A set function M{ $\Lambda$ } is called an uncertain measure if it satisfies the following four axioms: (i) Normality, (ii) Duality, (iii) Subadditivity, (iv) Product Axiom.

**Definition 2** (Liu 2007) An uncertain variable is a function  $\xi$  from an uncertainty space ( $\Gamma$ , L, M) to the set of real numbers such that { $\xi \in B$ } is an event for any Borel set *B* of real numbers.

An uncertain variable  $\xi$  is called normal if it has a normal uncertainty distributions

$$\Phi(t) = \left(1 + \exp\left(\frac{\pi(e-t)}{\sqrt{3}\sigma}\right)\right)^{-1}, \quad t \in \Re$$
(1)

denoted by  $\xi \sim \mathcal{N}(e, \sigma)$  where *e* and  $\sigma$  are real numbers and  $\sigma > 0$ .

Let  $\xi$  and  $\eta$  be uncertain variables. We say  $\xi > \eta$  if  $\xi(\gamma) > \eta(\gamma)$  for almost all  $\gamma \in \Gamma$ .

**Definition 3** (Liu 2007) For any real number *t*, the uncertainty distribution  $\Phi$  of an uncertain variable  $\xi$  is defined by

$$\Phi(t) = \mathbf{M}\{\xi \le t\}. \tag{2}$$

An uncertainty distribution  $\Phi(t)$  is called regular if it is a continuous and strictly increasing function with respect to *t* at which  $0 < \Phi(t) < 1$ , and  $\lim_{t \to -\infty} \Phi(t) = 0$ ,  $\lim_{t \to +\infty} \Phi(t) = 1$ . It is seen that normal uncertainty distribution is regular.

A regular uncertainty distribution  $\Phi$  has an inverse function  $\Phi^{-1}(\alpha)$  which is called the inverse uncertainty distribution.

The operational law is given as follows:

**Theorem 1** (Liu 2010) Let  $\xi_1, \xi_2, \ldots, \xi_n$  be independent uncertain variables with regular uncertainty distributions  $\Phi_1, \Phi_2, \ldots, \Phi_n$ , respectively. If  $f(\xi_1, \xi_2, \ldots, \xi_n)$  is strictly decreasing with respect to  $\xi_1, \xi_2, \ldots, \xi_m$  and strictly increasing with respect to  $\xi_{m+1}, \xi_{m+2}, \ldots, \xi_n$ , then  $\xi = f(\xi_1, \xi_2, \ldots, \xi_n)$  is an uncertain variable with inverse uncertainty distribution function

$$\Psi^{-1}(\alpha) = f(\Phi_1^{-1}(1-\alpha), \dots, \Phi_m^{-1}(1-\alpha), \\ \Phi_{m+1}^{-1}(\alpha), \dots, \Phi_n^{-1}(\alpha)), \ \alpha < 1.$$
(3)

According to the operational law, the following theorem can be obtained:

**Theorem 2** (Liu 2010) Let  $\xi_1$  and  $\xi_2$  be independent normal uncertain variables  $\mathcal{N}(e_1, \sigma_1)$  and  $\mathcal{N}(e_2, \sigma_2)$ , respectively. Then, the sum  $\xi_1 + \xi_2$  is also a normal uncertain variable, *i.e.*,

$$\mathcal{N}(e_1, \sigma_1) + \mathcal{N}(e_2, \sigma_2) = \mathcal{N}(e_1 + e_2, \sigma_1 + \sigma_2).$$
(4)

The product of a normal uncertain variable  $\mathcal{N}(e, \sigma)$  and a scalar number k > 0 is also a normal uncertain variable, *i.e.*,

$$k \cdot \mathcal{N}(e, \sigma) = \mathcal{N}(ke, k\sigma). \tag{5}$$

The expected value and variance of an uncertain variable are defined as follows:

**Definition 4** (Liu 2007) Let  $\xi$  be an uncertain variable. Then, the expected value of  $\xi$  is defined by

$$E[\xi] = \int_0^\infty \mathbf{M}\{\xi \ge r\} \mathrm{d}r - \int_{-\infty}^0 \mathbf{M}\{\xi \le r\} \mathrm{d}r \tag{6}$$

provided that at least one of the two integrals is finite.

**Definition 5** (Liu 2007) Let  $\xi$  be an uncertain variable with finite expected value *e*. Then, the variance of  $\xi$  is defined by

$$V[\xi] = E[(\xi - e)^2].$$
(7)

Since operations of the uncertain variables are mainly in the form of inverse uncertainty distributions, Liu (2010) and Yao (2015) presented following formulas to calculate the expected value and variance of an uncertain variable via inverse uncertainty distributions.

$$E[\xi] = \int_0^1 \Phi^{-1}(\alpha) \mathrm{d}\alpha. \tag{8}$$

$$V[\xi] = \int_0^1 \left( \Phi^{-1}(\alpha) - e \right)^2 \mathrm{d}\alpha.$$
(9)

For a normal uncertain variable  $\xi \sim \mathcal{N}(e, \sigma)$ , we can calculate that its expected value  $E[\xi] = e$  and variance  $V[\xi] = \sigma^2$ .

The moment of an uncertain variable is defined as follows:

**Definition 6** (Liu 2010) Let  $\xi$  be an uncertain variable and let k be a positive integer. Then,  $E[\xi^k]$  is call the k - th moment of  $\xi$ .

# 3 The uncertain EIT model with higher-order moment of the downside

In this section, we will provide an uncertain EIT model in which the tracking error is measured by the higher-order moment of the downside.

## 3.1 The assumptions and notations

Suppose there are *n* different stocks in the asset universe. Short-selling is not allowed. Let  $X_I$  denote the benchmark with the uncertain return  $r_I$ . Let  $n \times 1$  vector  $X_P$  denote the tracking portfolio with the uncertain return  $r_P$  which is independent of  $r_I$ . For example,  $r_P$  represents the return of a portfolio from Shanghai Stock Exchange and  $r_I$  represents NASDAQ-100 Index return, and they can be converted into independent variables by using the factor method in Huang and Zhao (2014). Next we introduce the rest of the notations used in this paper. For convenience, we show them in Table 1.

## 3.2 Higher-order moment of the downside

The goal of the EIT strategy is to track a benchmark index closely while providing some excess returns. Generally speaking, the tracking portfolio's return is not exactly the same as the benchmark's. There are positive or negative deviations between them. Investors welcome the positive deviation and do not want the negative deviation. So in our model, we will control the negative deviation.

Remember that  $r_{\rm P}$  and  $r_{\rm I}$  represent the uncertain returns of tracking portfolio and benchmark, respectively. Then  $(r_{\rm P} - r_{\rm I})^-$  represents the negative deviation of  $r_{\rm P}$  with respect to  $r_{\rm I}$ . For simplicity, we write  $(r_{\rm P} - r_{\rm I})^- = min(r_{\rm P} - r_{\rm I}, 0)$ . According to Definition 6, the *m*-th moment of  $|(r_{\rm P} - r_{\rm I})^-|$  is  $E[|(r_{\rm P} - r_{\rm I})^-|^m], m = 1, 2, \ldots$  Let  $\Psi$  denote the uncertainty distributions of  $r_{\rm P} - r_{\rm I}$ . Then according to Definition 4, we can have

$$E[|(r_{\rm P} - r_{\rm I})^{-}|^{m}] = \int_{0}^{+\infty} M\{|(r_{\rm P} - r_{\rm I})^{-}|^{m} \ge t\}dt$$
  
$$-\int_{-\infty}^{0} M\{|(r_{\rm P} - r_{\rm I})^{-}|^{m} \le t\}dt$$
  
$$= \int_{0}^{+\infty} M\{|(r_{\rm P} - r_{\rm I})^{-}|^{m} \ge t\}dt$$
  
$$= \int_{0}^{+\infty} M\{(r_{\rm P} - r_{\rm I}) \le -\sqrt[m]{t}\}dt$$
  
$$= \int_{0}^{+\infty} \Psi\left(-\sqrt[m]{t}\right)dt$$
  
$$= m\int_{-\infty}^{0} (-t)^{m-1}\Psi(t)dt.$$
  
(10)

Equation (10) provides a method for calculating the higher order moment of an uncertain variable.

## 3.3 The uncertain model

The uncertain EIT model with higher-order moment of the downside is designed to maximize the expected excess return over the benchmark subject with a fixed level of tracking error. The expected excess return is

$$E[r_{\rm P}] - E[r_{\rm I}] = E\left[\sum_{i=1}^{n} x_i \xi_i\right] - E[r_{\rm I}].$$
(11)

The tracking error is measured by the higher-order moment of the negative deviation of  $r_P$  with respect to  $r_I$ , i.e.,

$$E[|(r_{\rm P} - r_{\rm I})^{-}|^{m}].$$
(12)

So the uncertain EIT model with higher-order moment of the downside is as follows:

$$\max E\left[\sum_{i=1}^{n} x_i \xi_i\right] - E[r_{\mathrm{I}}]$$
  
subject to:
$$E\left[\left|\left(\sum_{i=1}^{n} x_i \xi_i - r_{\mathrm{I}}\right)^{-}\right|^{m}\right] \le D_m$$
(13)

$$\sum_{i=1}^{n} x_i = 1$$
  
$$x_i \ge 0, i = 1, 2, ..., n,$$

where  $D_m$  is the investors' tolerance level of the tracking error. By varying the value of *m*, model (13) can be changed into different uncertain EIT models that respond to different risk preferences of the investors.

Next we extend model (13) to a general form by including some realistic constraints. (i) The minimum transaction lots constraint. Because some securities markets have a minimum transaction lots limit, investors can only buy an integral multiple of the minimum transaction lots instead of any number of shares. To incorporate minimum transaction lots constraint, let  $\omega_i$  denote the number of transaction lots invested in stock *i* and  $N_i$  represent the number of units of stock *i* required as the minimum quantity. So  $\omega_i \cdot N_i$  represent the total quantity to be invested on stock *i*.

(ii) The budget constraint. In real investment, investors usually set budgets. Let  $P_i$  denote the unit market price of stock *i*. Thus,  $\sum_{i=1}^{n} \omega_i \cdot N_i \cdot P_i$  represents the total amount of money invested in *n* stocks.

(iii) The limits on holdings and cardinality constraint. The first one is to make sure that each stock is allowed to be

Table 1Notations used in thepaper	Notations	Meanings					
	X <sub>P</sub>	The tracking portfolio					
	$x_i$	The weight on stock $i$ in the tracking portfolio $X_P$					
	rp	The uncertain return of the tracking portfolio $X_P$					
	$r_{\mathrm{I}}$	The uncertain return of the benchmark index $X_I$					
	ξi	The uncertain return of stock <i>i</i>					
	$\omega_i$	The number of transaction lots invested in stock i					
	$N_i$	The number of units of stock $i$ required as the minimum quantity					
	$P_i$	The price per share of stock <i>i</i>					
	$l_i$	The minimum proportion allocated to stock <i>i</i> if it is held					
	ui	The maximum proportion allocated to stock $i$ if it is held					
	Q	The number of stocks that investors wish to hold in the tracking portfolio					
		$1 \le Q \le n$					
	Zi	The binary variable. If stock <i>i</i> is included in the portfolio, then $z_i = 1$					
		otherwise, $z_i = 0$					
	В	The amount of investment after paying transaction costs					

invested in specific ranges, i.e.,  $0 \le l_i \le x_i \le u_i \le 1$ . The second one ensures that only in a specified number of stocks the investment can be made, i.e.,  $\sum_{i=1}^{n} z_i = Q$  where  $z_i \in$  $\{0, 1\}$ . If  $z_i = 1$ , then the stock *i* is chosen for investment; otherwise not.

Based on the aforementioned discussions, the uncertain EIT model with higher-order moment of the downside considering realistic constraints can be formulated as follows:

$$\begin{aligned}
\max E\left[\sum_{i=1}^{n} x_i \xi_i\right] - E[r_I] \\
\text{subject to:} \\
E\left[\left|\left(\sum_{i=1}^{n} x_i \xi_i - r_I\right)^{-}\right|^{m}\right] \le D_m \\
\sum_{i=1}^{n} \omega_i \cdot N_i \cdot P_i \le B \\
x_i = \frac{\omega_i \cdot N_i \cdot P_i}{\sum_{i=1}^{n} \omega_i \cdot N_i \cdot P_i} \\
\sum_{i=1}^{n} z_i = Q \\
l_i \cdot z_i \le x_i \le u_i \cdot z_i \\
z_i \in \{0, 1\} \\
\omega_i \text{ are integers, } i = 1, 2, \dots, n.
\end{aligned}$$
(14)

#### 3.4 Discussion on the proposed model

In this section, we compare our model (14) with other uncertain EIT models. First, the comparison is conducted between models with higher-order and lower-order moments. In model (14), when  $m \ge 3$ , tracking error is measured by the higher-order moment of the downside. Below we give an uncertain EIT model in which the tracking error is measured by the lower order moment of the downside, i.e., model (15):

$$\max E\left[\sum_{i=1}^{n} x_i \xi_i\right] - E[r_1]$$
  
subject to:  
$$E\left[\left|\left(\sum_{i=1}^{n} x_i \xi_i - r_1\right)^{-}\right|\right] \le D$$
  
$$\sum_{i=1}^{n} \omega_i \cdot N_i \cdot P_i \le B$$
  
$$x_i = \frac{\omega_i \cdot N_i \cdot P_i}{\sum_{i=1}^{n} \omega_i \cdot N_i \cdot P_i}$$
  
$$\sum_{i=1}^{n} z_i = Q$$
  
$$l_i \cdot z_i \le x_i \le u_i \cdot z_i$$
  
$$z_i \in \{0, 1\}$$
  
$$\omega_i \text{ are integers, } i = 1, 2, ..., n.$$

We compare model (14) with model (15), and get the following result.

Theorem 3 With the optimal portfolio produced by model (15), either the expected return is no bigger or the tracking error is higher than that with the optimal portfolio produced by model (14).

**Proof** Let  $Obj_{14}$  and  $Obj_{15}$  denote the objective functions of models (14) and (15), respectively, the corresponding optimal solutions are  $\mathbf{X}_{14}$  and  $\mathbf{X}_{15}$ . (i) $\mathbf{X}_{15}$  does not satisfy the constraints in model (14), so  $\mathbf{X}_{15}$  is infeasible for model (14). This means that the tracking error is higher than that of model (14). (ii)Therefore, we get  $Obj_{15}(\mathbf{X}_{15}) \leq Obj_{14}(\mathbf{X}_{14})$ . This means that the expected return is no bigger than that of model (14).

Theorem 3 implies that the expected return of the model with higher order moment may be greater than that of the model with lower order moment, which gives the advantage of higher-order moment as the tracking error measure in the EIT problems.

Then the comparison is conducted between models with and without realistic constraints. We compare models (13) and (14), and obtain the following result.

**Theorem 4** The optimal objective value of the uncertain EIT model with realistic constraints is smaller than that of the model without realistic constraints.

**Proof** Suppose  $X_{13}$  and  $X_{14}$  are the optimal solutions of models (13) and (14), respectively, and their objective functions are  $Obj_{13}$  and  $Obj_{14}$ . Obviously,  $X_{14}$  is a feasible solution of model (13) but  $X_{13}$  is not necessarily a feasible solution of model (14). So  $X_{13} \neq X_{14}$ . Since  $X_{13}$  is an optimal solution of model (13) and  $X_{14}$  is a feasible one, we get  $Obj_{13}(X_{14}) < Obj_{13}(X_{13})$ . Since the objective functions of model (13) and model (14) are the same,  $Obj_{13}(X_{14}) = Obj_{14}(X_{14})$ . Therefore,  $Obj_{14}(X_{14}) < Obj_{13}(X_{13})$ . Thus, Theorem 4 is proved.

Though the optimal objective value of model (13) is relatively big, there do exist some realistic constraints in the real investment. Therefore, it is very necessary to consider realistic constraints in modeling.

## 3.5 The deterministic form

To solve model (14), we give the deterministic form of the model below.

**Theorem 5** Suppose stock returns  $\xi_i$  and the benchmark return  $r_1$  have regular uncertainty distributions  $\Phi_i$ , i = 1, 2, ..., n, and  $\Phi_I$ , respectively. Then model (14) is equivalent to the following model:

$$\max \int_{0}^{1} \sum_{i=1}^{n} x_{i} \Phi_{i}^{-1}(\alpha) d\alpha - \int_{0}^{1} \Phi_{I}^{-1}(\alpha) d\alpha$$
  
subject to:  

$$m \int_{-\infty}^{0} (-t)^{m-1} \Psi(t) dt \leq D_{m}$$
  

$$\Psi^{-1}(\alpha) = \sum_{i=1}^{n} x_{i} \Phi_{i}^{-1}(\alpha) - \Phi_{I}^{-1}(1-\alpha)$$
  

$$\sum_{i=1}^{n} \omega_{i} \cdot N_{i} \cdot P_{i} \leq B$$
  

$$x_{i} = \frac{\omega_{i} \cdot N_{i} \cdot P_{i}}{\sum_{i=1}^{n} \omega_{i} \cdot N_{i} \cdot P_{i}}$$
  

$$\sum_{i=1}^{n} z_{i} = Q$$
  

$$l_{i} \cdot z_{i} \leq x_{i} \leq u_{i} \cdot z_{i}$$
  

$$z_{i} \in \{0, 1\}$$
  

$$\omega_{i} \ are \ integers, \ i = 1, 2, ..., n.$$
  
(16)

**Proof** (i) Since  $x_i > 0$  and  $\xi_i$  have regular uncertainty distributions  $\Phi_i$ , according to Theorem 1, the inverse uncertainty distribution of  $\sum_{i=1}^{n} x_i \xi_i$  is  $\sum_{i=1}^{n} x_i \Phi_i^{-1}(\alpha)$ . Then according to Eq. (8), we have

$$E\left[\sum_{i=1}^{n} x_i \xi_i\right] = \int_0^1 \sum_{i=1}^{n} x_i \Phi_i^{-1}(\alpha) d\alpha$$
$$E[r_I] = \int_0^1 \Phi_I^{-1}(\alpha) d\alpha.$$

(ii) Let  $\eta = \sum_{i=1}^{n} x_i \xi_i - r_I$  and  $\Psi$  denote the uncertainty distribution of  $\eta$ . Since we suppose that  $\xi_i$  is independent of  $r_I$ , according to Theorem 1, we can have the inverse uncertainty distribution of  $\eta$  is  $\Psi^{-1}(\alpha) = \sum_{i=1}^{n} x_i \Phi_i^{-1}(\alpha) - \Phi_I^{-1}(1-\alpha)$ . Then according to Eq. (10), we have

$$E\left[\left|\left(\sum_{i=1}^{n} x_i \xi_i - r_{\mathrm{I}}\right)^{-}\right|^{m}\right] = m \int_{-\infty}^{0} (-t)^{m-1} \Psi(t) \mathrm{d}t$$

where  $\Psi^{-1}(\alpha) = \sum_{i=1}^{n} x_i \Phi_i^{-1}(\alpha) - \Phi_I^{-1}(1-\alpha).$ 

When stock returns all take normal uncertainty distributions, we further give the deterministic of model (16).

**Theorem 6** Suppose stock returns  $\xi_i$  and the benchmark return  $r_1$  take normal uncertainty distributions, i.e.,  $\xi_i \sim$ 

 $\mathcal{N}(e_i, \sigma_i)$  and  $r_{\rm I} \sim \mathcal{N}(e_I, \sigma_I)$ , respectively. Then model (16) can be transformed into the following form:

$$\begin{cases} \max \sum_{i=1}^{n} x_{i}e_{i} - e_{I} \\ subject to: \\ m \int_{-\infty}^{0} (-t)^{m-1} \left( 1 + exp\left(\frac{\pi(\sum_{i=1}^{n} x_{i}e_{i} - e_{I} - t)}{\sqrt{3}(\sum_{i=1}^{n} x_{i}\sigma_{i} + \sigma_{I})} \right) \right)^{-1} dt \leq D_{m} \qquad (I) \\ \sum_{i=1}^{n} \omega_{i} \cdot N_{i} \cdot P_{i} \leq B \qquad (II) \end{cases}$$

$$x_i = \frac{\omega_i \cdot N_i \cdot P_i}{\sum_{i=1}^n \omega_i \cdot N_i \cdot P_i}$$
(III)

$$\sum_{i=1}^{n} z_i = Q \tag{IV}$$

$$l_i \cdot z_i \le x_i \le u_i \cdot z_i \tag{V}$$

$$z_i \in \{0, 1\}$$

$$U_{\omega_i} \text{ are integers, } i = 1, 2, \dots, n.$$
 (VII)

**Proof** (i) As  $x_i > 0$  and  $\xi_i$  take normal uncertainty distributions, according to Theorem 2, we can proof that  $\sum_{i=1}^{n} x_i \xi_i$  takes a normal uncertainty distributions, i.e.,

$$\sum_{i=1}^n x_i \xi_i \sim \mathcal{N}\left(\sum_{i=1}^n x_i e_i, \sum_{i=1}^n x_i \sigma_i\right).$$

For a normal uncertain variable  $\xi \sim \mathcal{N}(e, \sigma)$ , we can calculated that its expected value  $E[\xi] = e$ . So we have  $E\left[\sum_{i=1}^{n} x_i \xi_i\right] = \sum_{i=1}^{n} x_i e_i$  and  $E[r_I] = e_I$ .

(ii) Similarly, according to Theorem 2, we can prove that  $\sum_{i=1}^{n} x_i \xi_i - r_{\rm I}$  also takes a normal uncertainty distributions, i.e.,

$$\sum_{i=1}^n x_i \xi_i - r_I \sim \mathcal{N}\left(\sum_{i=1}^n x_i e_i - e_I, \sum_{i=1}^n x_i \sigma_i + \sigma_I\right).$$

So we can get the uncertainty distribution of  $\sum_{i=1}^{n} x_i \xi_i - r_1$ . Then according to Eq. (10), we have

$$E\left[\left|\left(\sum_{i=1}^{n} x_i \xi_i - r_I\right)^{-}\right|^{m}\right]$$
  
=  $m \int_{-\infty}^{0} (-t)^{m-1} \left(1 + exp\left(\frac{\pi(\sum_{i=1}^{n} x_i e_i - e_I - t)}{\sqrt{3}(\sum_{i=1}^{n} x_i \sigma_i + \sigma_I)}\right)\right)^{-1} \mathrm{d}t.$ 

#### 

(VI)

(17)

# 4 Solving method based on discrete artificial bee colony algorithm

Model (17) is an NP-hard optimization problem that is difficult to obtain an optimal solution with accurate algorithms. Due to this difficulty, meta-heuristic algorithms have become popular in this context because they can obtain an acceptable optimal solution within an allowable time for the NP-hard optimization problem. As one of the meta-heuristic algorithms, the artificial bee colony (ABC) algorithm (Karaboga 2005; Karaboga and Basturk 2007) is simple to implement and has strong robustness. The literature shows that the ABC algorithm performs well not only in continuous optimization problems but also in many discrete optimization problems such as economic maintenance planning problem (Chen et al. 2020), flow shop scheduling problem (Xuan et al. 2019) and clustering problem (Masdari et al. 2019). Therefore, in this paper, we will use the discrete ABC algorithm for solving the proposed optimization model (17).

#### 4.1 Solution feasibility analysis of proposed model

In this section, we conduct the solution feasibility analysis of model (17). The constraints mentioned below are all from model (17).

First, we give "parameter conditions." From constraint (III) we get  $\sum_{i=1}^{n} x_i = 1$ . Then by summing constraint (V) with respect to *i*, we have  $\sum_{i=1}^{n} l_i z_i \leq 1 \leq \sum_{i=1}^{n} u_i z_i$ . Obviously, if  $\sum_{i=1}^{n} l_i z_i > 1$  or  $\sum_{i=1}^{n} u_i z_i < 1$ , then the feasible solution does not exist. Thus, "parameter conditions" are as follows:

(i) 
$$\sum_{i=1}^{n} l_i z_i \le 1$$
.  
(ii)  $\sum_{i=1}^{n} u_i z_i \ge 1$ .

Then, we give "solution generating conditions." From constraint (IV) we know that the number of  $z_i$  which equals 1 value is Q and the number of  $z_i$  which equals 0 value is (n - Q). From constraints (III) and (V), we can get that if  $z_i = 0$ , then  $\omega_i = 0$ . The upper limit of  $w_i$  is obtained from constraints (II), (III), and (V). So "solution generating conditions" are as follows:

(i) 
$$\sum_{i=1}^{n} z_i = Q, z_i \in \{0, 1\}.$$
  
(ii) If  $z_i = 0$ , then  $w_i = 0$ .  
(iii) If  $z_i = 1$ , then  $w_i \leq \frac{Bu_i}{N_i P_i}$ ,  $w_i$  are positive integers.

If  $z_i$  and  $w_i$  are generated so that the solution generating conditions are satisfied, the optimization problem to be solved is to obtain a solution that maximizes the objective function under constraints (I), (II), and (V). We improve the exploitation ability of the algorithm by ensuring that the above three

V	$Z_1$	$Z_2$		$Z_n$
Λ	$W_1$	W <sub>2</sub>		W <sub>n</sub>

Fig. 1 Solution coding structure

constraints are satisfied when generating and updating solutions in our discrete ABC algorithm.

Algorithm 1 Random generation of solution						
<b>Input:</b> problem parameters $(n, Q, B, N_i, P_i, l_i, u_i)$ <b>Output:</b> a generated solution ( <b>X</b> ) 01: select randomly $Q$ stocks $i, i \in I / * I$ is a set of $Q$ distinct stocks $*/$ 02: for $i = 1 : n / *$ Determination of $z_i$ and $w_i$ 03: if $i \in I$ then according to the solution generating 04: $z_i = 1$ conditions $*/$ 05: $w_i$ is equal to a random positive integer not larger than $\frac{Bu_i}{N_i P_i}$ 06: else 07: $z_i = 0$ 08: $w_i = 0$ 09: end if 10: end for 11: generate <b>X</b>						

#### 4.2 Discrete ABC algorithm for proposed model

The ABC algorithm is a meta-heuristic algorithm inspired by the behavior of bees. In the ABC algorithm, bees are divided into three classes: employed bees, onlooker bees, and scouts. The employed bees use the previous food source information to find new food sources and share the information with onlooker bees who wait in the hive and look for new food sources according to the information shared by employed bees. The task of scouts is to find a new valuable food source. They randomly look for a food source near the hive. The goal of the whole colony is to find the nectar source with the largest amount of nectar.

Combining the basic principles of the ABC algorithm with the uncertain EIT model, the main steps of the algorithm for solving the proposed model (17) are as follows.

#### 4.2.1 Solution coding method

The optimization problem to be solved is described by the binary variables  $z_i$  and the integer variables  $w_i$ , i = 1, 2, ..., n. In the algorithm, a solution **X** is coded as a  $2 \times n$ matrix in which the first row consists of  $z_i$  and the second row consists of  $w_i$  as shown in Fig. 1.

#### 4.2.3 Crossover and Mutation

4.2.2 Population initialization

shown in Algorithm 1.

The *SN* initial solutions to compose the population are randomly generated while satisfying the solution generating

conditions in Sect. 4.1. The pseudo-code for generating each solution  $X_j$ , j = 1, 2, ..., SN, in the initialization phase is

In the employed bee phase and the onlooker bee phase, the solution is updated through crossover and mutation. In the algorithm, we implement two-point crossover and two-point mutation and make sure that the solution generating conditions are satisfied. In the two-point crossover process, first, we find the positions where the numbers of 1 value are the same from 1 to that position in  $\mathbf{X}_i$  and  $\mathbf{X}_k$ . Let H be the set of such positions. Now we give an example in which the parameter settings are n=9 and Q=5. As shown in Fig. 2, there are two solutions  $\mathbf{X}_i$  and  $\mathbf{X}_k$ . At the position i = 3, both  $\mathbf{X}_i$ and  $\mathbf{X}_k$  contain two 1's in (1:3) slices, so i = 3 becomes an element of H. After constructing H in the same way, we select two random positions  $i_1$  and  $i_2$  in H and get that the slices  $(i_1 + 1 : i_2)$  of  $\mathbf{X}_i$  and  $\mathbf{X}_k$  contain the same number of 1's. Then we replace the slices  $(i_1 + 1 : i_2)$  of  $\mathbf{X}_i$  and  $\mathbf{X}_k$ with each other to generate new solutions  $X_1^{new}$  and  $X_2^{new}$  as shown in Fig. 2.

The two-point mutation method is as follows. Let the set of positions where  $z_i$  takes 1 value be  $H_1$  and the set of positions where  $z_i$  takes 0 value be  $H_2$  in the solution  $\mathbf{X}_j$ . First, we select randomly the positions  $i_1$  and  $i_2$  from sets  $H_1$  and  $H_2$ , respectively. Then we invert the values of  $z_{i_1}$  and  $z_{i_2}$  at the  $i_1$  and  $i_2$  positions. We give the  $w_{i_1}$  value as 0 and  $w_{i_2}$  as a random positive integer not larger than  $\frac{Bu_{i_2}}{N_{i_2}P_{i_2}}$ . Repeat this mutation process  $p_m \times n$  times as shown in Fig. 3. From the

new solutions obtained by crossover and mutation, the best solution is selected according to the tournament selection rule. Algorithm 2 shows the process of generating a new solution by crossover and mutation. (I), (II), and (V) need to be considered when calculating the fitness value.

The probability that the onlooker bee goes to the jth food source is determined by the following equation:

#### Algorithm 2

Generation of new solution by Crossover and Mutation

**Input:** solution  $(X_i)$  and its neighborhood  $(X_k)$ **Output:** new solution  $(X_i^{new})$ 01: find the positions *i* where the numbers of 1s are /\* Two-point crossover \*/ the same in  $X_i$  and  $X_k$ ,  $i \in H / H$  is a set of those positions \*/02: choose randomly two positions,  $i_1$  and  $i_2$ , in H 03: generate  $X_1^{new}$  by replacing  $(i_1 + 1 : i_2)$  elements /\* See Fig. 2. \*/ of  $X_i$  with ones of  $X_k$ 04: generate  $X_2^{new}$  by replacing  $(i_1 + 1 : i_2)$  elements of  $X_k$  with ones of  $X_i$ 05:  $X' = X_j / *$  Mutation \*/ 06: for t = 1:  $p_{\rm m} \times n$ 07: select randomly a position  $i_1$  among 1 elements in X'08:  $z_{i_1} = 0$ 09:  $w_{i_1} = 0$ 10: select randomly a position  $i_2$  among 0 elements in  $X_i$ 11:  $z_{i_2} = 1$ 12:  $w_{i_2}$  is equal to a random positive integer not larger than  $\frac{Bu_{i_2}}{N_{i_2}P_{i_2}}$ generate  $X_{3}^{new}$  by replacing  $(i_1, i_2)$  elements of /\* See Fig. 3. \*/ 13: X' with the above 14:  $X' = X_{2}^{new}$ 5: end for 16: take out the best solution  $X_i^{new}$  by tournament selection with  $X_1^{new}$ ,  $X_2^{new}$  and  $X_3^{new}$ 

#### 4.2.4 Fitness function

The fitness values are used to evaluate the best solution in the tournament selection and to calculate the probability that the onlooker bees select a specific food source (solution) in the onlooker bee phase. The higher the fitness value, the better the solution, and the greater the probability that onlooker bees will choose the solution. The fitness fit<sub>j</sub> of solution  $X_j$  is determined by the following equations:

$$F_j = -\left(f(\mathbf{X}_j) - \sum_{l=1}^{L} g_l(\mathbf{X}_j)\right)$$
(18)

$$fit_{j} = \begin{cases} \frac{1}{1+F_{j}}, & \text{if } F_{j} \ge 0\\ 1 + \operatorname{abs}(F_{j}), & \text{if } F_{j} < 0 \end{cases}$$
(19)

where  $f(\mathbf{X}_j)$  is the objective value at  $\mathbf{X}_j$  and  $g_l(\mathbf{X}_j)$  is the *l*th constraint violation value at  $\mathbf{X}_j$  which is normalized so that max  $g_l(\mathbf{X}_j) = 1$ . *L* is the number of constraints (I), (II) and (V) of model (17), that is, L = 2n + 2. In our algorithm, all solutions in the population are generated and updated to satisfy the solution generating conditions, so only constraints

$P_j =$	$= \frac{\mathrm{In}_j}{\sum_{j=1}^{\mathrm{SN}} \mathrm{fit}_j}.$	(20)

fit

H: O			3	4		6			9
$\mathbf{v}$	0	1	1	0	0	1	0	1	1
$\Lambda_{j}$	$W_{j1}$	$W_{j2}$	W <sub>j3</sub>	$W_{j4}$	$W_{j5}$	W <sub>j6</sub>	$w_{j7}$	$W_{j8}$	<i>w</i> <sub>j9</sub>
V	1	1	0	0	1	0	1	1	0
$\boldsymbol{\Lambda}_{k}$	$W_{k1}$	$W_{k2}$	$W_{k3}$	$W_{k4}$	$W_{k5}$	$W_{k6}$	$W_{k7}$	$W_{k8}$	$W_{k9}$
V new	0	1	1	0	1	0	0	1	1
$\Lambda_1$	$W_{j1}$	$W_{j2}$	$W_{j3}$	$W_{k4}$	$W_{k5}$	$W_{k6}$	$w_{j7}$	$W_{j8}$	$W_{j9}$
$X_{2}^{new}$	1	1	0	0	0	1	1	1	0
2	$W_{k1}$	$W_{k2}$	$W_{k3}$	$W_{j4}$	$W_{j5}$	$W_{j6}$	$W_{k7}$	$W_{k8}$	$W_{k9}$

**Fig. 2** Two-point crossover when  $i_1 = 3$  and  $i_2 = 6$ 

V	0	1	1	0	0	1	0	1	1
$X_{j}$	$W_{j1}$	$W_{j2}$	$W_{j3}$	$W_{j4}$	$W_{j5}$	$W_{j6}$	$W_{j7}$	<i>W</i> <sub>j8</sub>	$W_{j9}$
TT New	0	0	1	1	0	1	1	1	0
$X_3$	$w_{j1}$	0	W <sub>j3</sub>	$\overline{W}_{j4}$	$W_{j5}$	W <sub>j6</sub>	$\overline{W}_{j7}$	<i>W</i> <sub><i>j</i>8</sub>	0

**Fig. 3** Mutation with probability  $p_{\rm m}$ 

In the algorithm, the employed bee of food source (solution) that is not updated more than *Limits* cycles becomes a scout bee. The scout bee discards the original food source and randomly searches for a new one. A new solution is generated according to Algorithm 1 to satisfy the solution generating conditions in Sect. 4.1. The pseudo-code of the total algorithm is listed in Algorithm 3. In Algorithm 3, MCN indicates the maximum number of cycles.

Table 2	Table 2         Codes and Prices of 10 stocks in the investor's asset universe									
Stock i	Stock code	Price	Stock i	Stock code	Price					
1	600929	6.61	6	601698	11.55					
2	603214	42.75	7	601330	10.75					
3	601990	13.16	8	002371	93					
4	000034	20.59	9	600841	7.41					
5	002032	77.89	10	603712	25.15					

The unit of stock price is RMB

Algorithm 3 Discrete ABC algorithm for proposed model
<b>Input:</b> common parameters ( $MCN$ , $SN$ , $Limits$ , $p_m$ ),
problem parameters $(n, Q, \text{etc})$ ,
objective function $(f(\mathbf{X}))$ ,
constraint violation function $(g_l(\mathbf{A}), l = 1,, 2n + 2)$ / Constraints (1), (1)
and $(\mathbf{V})$ of model $(1)$ $\mathbf{V}$
Output: Dest solution ( $A_{best}$ )
ora in Section 4.1 */
02: for $i = 1$ : SN /* Initialization of
population */
03: generate solution <b>X</b> by using Algorithm 1
04. end for
05: $cycle = 1$
06: while $cvcle < MCN$
07: for $i = 1:SN /*$ Employed bee phase */
08: select randomly $\mathbf{X}_k$ as neighborhood solution
09: generate new solution $\mathbf{X}_{i}^{new}$ by using Algorithm 2
with $\mathbf{X}_i$ and $\mathbf{X}_k$
10: update/maintain $\mathbf{X}_i$ by tournament selection
with $\mathbf{X}_i$ and $\mathbf{X}_i^{new}$
11: end for
12: calculate probability $\{P_s\}$ for onlookers $(s = 1 : SN)$ /* Onlooker bee phase */
13: for $j = 1:SN$
14: select $\mathbf{X}_i$ for <i>j</i> th onlooker bee by roulette selection
with probability $\{P_s\}$
15: select randomly $\mathbf{X}_k$ as neighborhood
16: generate new solution $\mathbf{X}_{i}^{new}$ by using Algorithm 2
with $\mathbf{X}_i$ and $\mathbf{X}_k$
17: update/maintain $X_i$ by tournament selection
with $\mathbf{X}_i$ and $\mathbf{X}_i^{new}$
18: end for
19: if solution $\mathbf{X}^*$ is not updated more than <i>Limits</i> , /* Scout bee phase */
20: generate new solution $\mathbf{X}^{new}$ by using Algorithm 1
and replace $\mathbf{X}^*$ with $\mathbf{X}^{hew}$
21: end if
22:  cycle = cycle + 1
23: end while
24: take out the best solution $\mathbf{X}_{best}$

**Table 3** Uncertainty annualreturns of 10 stocks in the year2020

Table 4Results of objectivevalues under different

parameters

Stock	i	D	istributio	on	Stock i	Dis	tribution		
1		Л	$\mathcal{N}(0.0560, 0.0890)$		6	$\mathcal{N}($	$\mathcal{N}(0.1760, 0.2670)$		
2		Л	(0.0890	, 0.1200)	7	$\mathcal{N}(0.1990, 0.2990)$			
3		Л	(0.1000	, 0.1500)	8	$\mathcal{N}($	0.2100, 0.3000		
4		Л	(0.1200	, 0.2100)	9	$\mathcal{N}($	0.2300, 0.3100		
5		Л	(0.1560	, 0.2400)	10	$\mathcal{N}($	0.2500, 0.3400)		
SN	MCN	LC	<i>p</i> <sub>m</sub>	Max	Min	Mean	Std		
20	3000	30	0.1	0.049819993	0.039279792	0.044999299	0.003400342		
50	10,000	50	0.1	0.053265103	0.045712645	0.050131270	0.002987591		
20	3000	30	0.2	0.056482236	0.052386886	0.053965506	0.001400843		
30	3000	30	0.2	0.056475331	0.053662966	0.055081435	0.000886998		
50	3000	30	0.2	0.056822179	0.052532265	0.055262735	0.001215924		
20	5000	30	0.2	0.056904570	0.052722858	0.054909519	0.001352166		
30	5000	30	0.2	0.056101794	0.052746888	0.054526161	0.001106524		
50	5000	30	0.2	0.057563928	0.053874579	0.055820492	0.001022906		
20	10,000	30	0.2	0.056835820	0.052537029	0.054975260	0.001244287		
30	10,000	30	0.2	0.057209390	0.054099047	0.055546643	0.001058245		
50	10,000	30	0.2	0.057446022	0.054819683	0.056078805	0.000788396		
20	3000	50	0.2	0.055979326	0.050200949	0.054170810	0.001598595		
30	3000	50	0.2	0.056502468	0.049824471	0.054070642	0.002249976		
50	3000	50	0.2	0.056377331	0.053255532	0.055173151	0.001211683		
20	5000	50	0.2	0.056075492	0.051490290	0.054140908	0.001502307		
30	5000	50	0.2	0.056647327	0.052077561	0.055367282	0.001428254		
50	5000	50	0.2	0.056978496	0.053878936	0.055729771	0.001031324		
20	10,000	50	0.2	0.057294428	0.052376487	0.055535992	0.001452142		
30	10,000	50	0.2	0.056482289	0.054397561	0.055686587	0.000810668		
50	10,000	50	0.2	0.056722580	0.053431789	0.055515924	0.000969643		
20	3000	30	0.3	0.056665311	0.052942662	0.054948114	0.001439609		
20	3000	30	0.4	0.055967714	0.051401768	0.053364756	0.001496180		
20	3000	30	0.6	0.055271474	0.048776630	0.051372146	0.001930273		

# **5 Numerical examples**

To clearly illustrate the modeling idea and the designed algorithm, we present some numerical examples. In our examples, stock returns are treated as uncertain variables. Moreover, the algorithms are run in MATLAB and all experiments are conducted on a PC with a 3.6 GHz processor and 8 GB RAM.

## 5.1 Computational results of the tracking portfolio

## 5.1.1 Data

An investor wants to track the SSE50 index whose return distribution is  $\mathcal{N}(0.130, 0.220)$ . The parameter *m* in model (17) is set at 3. According to three prospectuses of enhanced index funds coded 005,850, 009,059, and 161,037 in China's

securities market, the annual tracking error should not exceed 8%. So we set the tolerance level of the tracking error at 8% in our proposed model. The investor selects 10 stocks from Shanghai Stock Exchange and Shenzhen Stock Exchange as his/her asset universe. The investment period is set from January 2, 2020 to December 31, 2020. Stock codes and the prices of stocks on January 2, 2020, are shown in Table 2. Stock annual return distributions in the investment period are shown in Table 3. The amount of investment after paying transaction costs is B = 1,000,000. The tracking portfolio is supposed to contain 6 stocks, i.e., Q = 6 and each stock's minimum proportions are  $l_i = 5\%$ , i = 1, 2, ..., 10, respectively. The minimum transaction lots for each stocks are  $N_i = 100$ , i = 1, 2, ..., 10.

**Table 5**Optimal trackingportfolio

Stock i	$\omega_i$	Stock i	$\omega_i$
1	96	6	0
2	32	7	0
3	66	8	5
4	0	9	582
5	0	10	61

## 5.1.2 Stability of algorithm

The stability of calculation results depends on the setting of parameters, namely the size of the population (SN), the maximum number of cycles (MCN), *Limits* cycles (LC), and mutation probability ( $p_m$ ). Here, a series of experiments are conducted to determine these four parameters. We perform 20 times tests for each set of parameter settings. Table 4 gives the minimum, maximum, mean, and standard deviation of the 20 objective values under different parameters, where the best results are highlighted in boldface. According to the result, the best parameter setting is SN = 50, MCN = 10, 000, LC = 30, and  $p_m = 0.2$ . Besides, in Table 4, it can be seen that the standard deviations of the 20 optimal objective values for every set of parameters are all very small, which indicates the robustness of the proposed algorithm.

#### 5.1.3 The optimal tracking portfolio

In this section, we provide the computational result of model (17), which is conducted 20 times, and the best result is reported. Table 5 shows that the optimal tracking portfolio contains stocks 1, 2, 3, 8, 9, and 10, and the corresponding transaction lots are 96, 32, 66, 5, 582, and 61, respectively. The expected return of the optimal tracking portfolio is  $E[r_P^*] = 0.1870$ , and the variance is  $V[r_P^*] = 0.0654$ . The total investment is 918,289 RMB.

## 5.2 Sensitivity analysis

In this section, we change some parameters to test their effects on the investment decision. Note that each time only one parameter changes. Similarly, we do the experiments 20 times and select the best result.

(i) Sensitivity analysis w.r.t the parameter m. According to Eq. (10), the change of m means changing the order-moment of the downside. When changing m from 1 to 3, model (17) can be changed into three models. Figure 4 shows the relationship between the objective value and the tracking error tolerance level  $D_m$  in these three models. The abscissa represents  $D_m$  and the ordinate represents the objective value. When model (17) has no solution, the objective value is displayed as zero. As seen in Fig. 4, no matter how m changes, it

does not affect the objective value. But one thing is worthy to note. When m = 1, model (17) has a solution in the interval of  $D_1 \in [0.16, 0.2]$ ; when m = 2, model (17) has a solution in the interval of  $D_2 \in [0.08, 0.2]$ ; when m = 3, model (17) has a solution in the interval of  $D_3 \in [0.04, 0.2]$ . It can be seen that when m = 3, model (17) is solvable in a larger range of the tracking error tolerance. This is why we use the higher order moment which makes the proposed model more widely applicable. However, the increase in m also increases the complexity of model solving, so m is selected as 3 in our experiment.

(ii) Sensitivity analysis w.r.t the parameter  $D_3$ . When m = 3, the investor's tracking error tolerance level is represented by  $D_3$  in model (17). We change the value of  $D_3$  from 0.05 to 0.15. Figure 5 shows the relationship between the tracking error tolerance level and the expected return of the optimal tracking portfolio. Note that when  $D_3$  increases from 0.05 to 0.11, the expected return increases; when  $D_3$  increases from 0.11 to 0.15, the expected return remains unchanged. Figure 5 implies that when the tracking portfolio's expected return increases with an increase in the tracking error tolerance level. However, when the tracking error tolerance level exceeds a certain range, the tracking portfolio's expected return is unchanged.

(iii) Sensitivity analysis w.r.t the parameter Q. The parameter Q represents the number of stocks that the investor wants to hold in the tracking portfolio. We change the value of the parameter Q from 4 to 8. The corresponding results are shown in Table 6. In Table 6, when Q = 4, the optimal tracking portfolio's expected return and variance are  $E[r_{\rm P}^*] = 0.1904$  and  $V[r_{\rm P}^*] = 0.0664$ . As Q increases to 8, the optimal tracking portfolio's expected return and variance are are  $E[r_{\rm P}^*] = 0.1839$  and  $V[r_{\rm P}^*] = 0.0646$ . It can be seen that return and risk become small as Q increases. This is consistent with the result that diversification reduces the risk, and also reduces the return.

(iv) Sensitivity analysis w.r.t the parameter B. This parameter denotes the upper limit of the budget after paying transaction costs. We calculate the expected return and variance of the tracking portfolio with different B levels and show them in Table 7. As seen in Table 7, the tracking portfolio's expected return and variance change little when B increases. From the above experiments, it is found that the parameter B can hardly affect the optimal tracking portfolio's expected return and variance.

(v) Sensitivity analysis w.r.t the parameter  $l_i$  and  $u_i$ . These two parameters denote the minimum and maximum proportions allocated to stock *i* if it is held. Table 8 shows how  $l_i$ affect the tracking portfolio. In Table 8, when  $l_i = 0.01$ , the return and risk of the tracking portfolio are  $E[r_p^*] = 0.1898$ 



**Fig. 5** Relationship between the tracking error tolerance level and the tracking portfolio's expected return

and  $V[r_{\rm P}^*] = 0.0662$ ; when  $l_i$  increase to 0.15, the return and risk of the tracking portfolio are  $E[r_{\rm P}^*] = 0.1795$  and  $V[r_{\rm P}^*] = 0.0634$ . As the increase in the minimum proportion allocated to stock *i* if it is held, the return and risk of the tracking portfolio both decrease. Table 9 shows how  $u_i$  affect the tracking portfolio. As seen in table 9, when  $u_i$  increase from 0.2 to 0.4, the returns of the tracking portfolio increase from  $E[r_{\rm P}^*] = 0.1813$  to  $E[r_{\rm P}^*] = 0.1871$ , and the risk of the tracking portfolio increase from  $V[r_{\rm P}^*] = 0.0639$  to  $V[r_{\rm P}^*] = 0.0655$ . However, when  $u_i$  increase from 0.4 to 0.9, the return and risk of the tracking portfolio are unchanged. This is because when  $u_i \in (0.2, 0.4)$ , the constraint (V) is active; and when  $u_i \in (0.4, 0.9)$ , the constraint (V) is not active.

<b>Table 6</b> Sensitivity analysisw.r.t the parameter $Q$	Q	$E[r_{\rm P}^*]$	$V[r_{\rm P}^*]$	$\omega_1$	$\omega_2$	ω3	ω4	ω5	$\omega_6$	ω7	$\omega_8$	ω9	$\omega_{10}$
	4	0.1904	0.0664	85	48	0	0	0	0	0	0	822	28
	5	0.1887	0.0659	81	51	0	0	0	0	0	7	630	53
	6	0.1870	0.0654	79	39	68	0	0	0	0	6	529	81
	7	0.1845	0.0648	137	37	41	0	0	43	0	9	389	100
	8	0.1839	0.0646	70	41	41	0	6	40	0	6	385	85

 
 Table 9
 Sensitivity analysis
 w.r.t the parameter  $u_i$ 

<i>u</i> <sub>i</sub>	$E[r_{\rm P}^*]$	$V[r_{\rm P}^*]$
0.2	0.1813	0.0639
0.3	0.1863	0.0653
0.4	0.1871	0.0655
0.5	0.1870	0.0655
0.6	0.1870	0.0655
0.7	0.1870	0.0655
0.8	0.1870	0.0655
0.9	0.1870	0.0655

# 5.3 Applicability and effectiveness in large-scale problems

To test the applicability and effectiveness of our algorithm in large-scale problems, we do more experiments. In Sections 5.1 and 5.2, we have solved the model for 10 stocks. Next, we will solve the model for 50 stocks. Table 10 shows the uncertain returns of the 50 stocks, and Table 11 shows the results of the experiments. This problem has been solved with the restriction that there should be an investment in at least 25 stocks with at least a 2% allocation in each stock. Table 11 shows the optimal tracking portfolios for different values of tracking error tolerance in dealing with 50 stocks. When  $D_3$ increases to 0.15 or greater, the objective value converges to the near-same value. It is seen that our algorithm is effective in large-scale problems.

# 6 Conclusions

In reality, financial markets are often affected by unexpected events, which can lead to historical data failing to predict the future effectively. In this case, we employ uncertain variables to describe the stock returns.

Under uncertainty theory framework, this paper has studied the EIT problem. We have proposed an uncertain EIT model with higher order moment of the downside, in which the realistic constraints of minimum transaction lots, investment budget, and the limits on holdings and cardinality are considered simultaneously. To solve the proposed model, we have given the deterministic equivalents of the model. Then we have developed a meta-heuristic algorithm to solve the proposed nonlinear integer programming. Experiment results

Table 10Uncertainty returns ofthe 50 stocks	Stock i	Distribution	Stock i	Distribution
	1	$\mathcal{N}(0.0560, 0.0890)$	26	$\mathcal{N}(0.1690, 0.2580)$
	2	$\mathcal{N}(0.0890, 0.1200)$	27	N(0.3340, 0.4450)
	3	$\mathcal{N}(0.1000, 0.1500)$	28	$\mathcal{N}(0.2370, 0.3260)$
	4	$\mathcal{N}(0.1200, 0.2100)$	29	N(0.3530, 0.4650)
	5	$\mathcal{N}(0.1560, 0.2400)$	30	$\mathcal{N}(0.1430, 0.2200)$
	6	$\mathcal{N}(0.1760, 0.2670)$	31	$\mathcal{N}(0.0650, 0.1040)$
	7	$\mathcal{N}(0.1990, 0.2990)$	32	N(0.1230, 0.2140)
	8	$\mathcal{N}(0.2100, 0.3000)$	33	N(0.2210, 0.3090)
	9	$\mathcal{N}(0.2300, 0.3100)$	34	N(0.2340, 0.3240)
	10	$\mathcal{N}(0.2500, 0.3400)$	35	N(0.3680, 0.4710)
	11	$\mathcal{N}(0.1440, 0.2260)$	36	N(0.2610, 0.3570)
	12	$\mathcal{N}(0.1790, 0.2890)$	37	N(0.2550, 0.3450)
	13	$\mathcal{N}(0.1520, 0.2320)$	38	N(0.3880, 0.4870)
	14	$\mathcal{N}(0.2330, 0.3230)$	39	$\mathcal{N}(0.0740, 0.1100)$
	15	$\mathcal{N}(0.1350, 0.2180)$	40	N(0.3140, 0.4240)
	16	$\mathcal{N}(0.1810, 0.2900)$	41	N(0.1540, 0.2350)
	17	$\mathcal{N}(0.2170, 0.3070)$	42	N(0.3610, 0.4690)
	18	$\mathcal{N}(0.1870, 0.2980)$	43	N(0.2420, 0.3350)
	19	$\mathcal{N}(0.2320, 0.3140)$	44	$\mathcal{N}(0.0900, 0.1290)$
	20	$\mathcal{N}(0.2510, 0.3440)$	45	$\mathcal{N}(0.1850, 0.2970)$
	21	$\mathcal{N}(0.2150, 0.3050)$	46	N(0.2390, 0.3330)
	22	$\mathcal{N}(0.3990, 0.4940)$	47	N(0.3690, 0.4750)
	23	$\mathcal{N}(0.0440, 0.0750)$	48	N(0.3700, 0.4800)
	24	$\mathcal{N}(0.0880, 0.1190)$	49	N(0.3490, 0.4500)
	25	$\mathcal{N}(0.3380, 0.4470)$	50	N(0.1680, 0.2530)

Table 1	1 Opti	imal trac	cking pc	ortfolic	o in larg	ge-scal	e prob	lems f	or diffe	erent v	alues c	of track	ing errc	or tolera	nce												
$D_3$	Objec	tive valı	ie $\omega_1$	$\omega_2^{\prime}$	$\omega_3$	$\omega_4$	$\omega_5$	$\omega_6$	ĹΦ	$\omega_8$	609	$\omega_{10}$	$\omega_{11}$	$\omega_{12}$	$\omega_{13}$	$\omega_{14}$	$\omega_{15}$	$\omega_{16}$	$\omega_{17}$	$\omega_{18}$	$\omega_{19}$	$\omega_{20}$	$\omega_{21}$	$\omega_{22}$	$\omega_{23}$	$\omega_{24}$	ω25
0.08	0.0570	-	58	13	30	0	4	0	0	0	0	0	0	0	17	0	0	0	0	0	40	0	0	11	41	26	5
0.09	0.0849	-	76	13	39	0	0	0	0	б	50	0	0	0	0	0	0	0	0	0	30	0	0	9	37	24	5
0.10	0.1089	-	47	. 13	4	0	0	0	0	0	29	10	0	0	0	0	0	0	20	0	36	14	0	11	28	22	5
0.11	0.1282		31	8	0	0	0	0	0	0	60	0	0	0	20	11	0	0	0	0	74	0	47	11	0	18	5
0.12	0.1472		0	٢	19	0	0	0	0	0	0	8	0	0	0	17	0	0	0	0	38	8	71	12	0	17	4
0.13	0.1634		0	0	15	0	4	18	0	б	61	8	0	0	0	12	0	0	0	0	40	21	0	13	0	10	L
0.14	0.1799	_	0	0	0	0	0	0	19	Э	26	8	0	0	0	27	0	0	0	35	8	38	13	0	10	7	0
0.15	0.1827		0	0	0	0	0	0	27	Э	28	6	0	0	0	11	0	0	22	0	34	14	35	13	0	0	٢
0.16	0.1828	~~	0	0	0	0	0	21	24	Э	28	8	0	0	0	10	0	0	20	0	28	12	36	13	0	0	7
0.17	0.1822	•	0	0	0	0	0	0	18	4	43	8	0	0	0	12	0	0	25	0	29	6	0	13	0	0	7
	$\omega_{26}$	ω27	$\omega_{28}$	$\omega_{29}$	$\omega_{30}$	<i>w</i> 31	<i>w</i> 3	12 a	033	ω34	ω35	<i>w</i> 36	$\omega_{37}$	$\omega_{38}$	<i>w</i> 39	$\omega_{40}$	$\omega_4$	1 00.	42 $\alpha$	λ43 α.	λ44 C	045 0	ω46	ω47	$\omega_{48}$	$\omega_{49}$	$\omega_{50}$
	20	0	0	38	0	84	0	۲٦	33	20	б	0	33	б	75	0	75	0	0	4	) 0	0	0	19	0	47	8
	0	0	0	44	16	35	0	0	-	0	9	34	20	4	70	×	0	0	1	9 2	1 (	$\sim$	17	34	5	75	0
	0	0	0	0	0	32	0	0	-	19	٢	35	47	5	54	13	0	11	1	1	5 (	0	0	4	3	76	0
	0	43	0	69	0	73	0	C	-	0	4	17	21	5	09	15	0	17	2	1	7 (	0	0	47	4	62	8
	0	0	0	36	0	0	0	(7)	8	0	7	42	19	9	40	19	0	1(	) 1	6 1	3 (	0	26	47	9	76	0
	0	28	3	67	0	0	23	-	5	7	19	50	5	0	0	0	0	11	1	0		0	0	49	5	75	0
	0	50	0	LL	0	0	0	1	9	11	٢	17	0	9	0	22	0	1	3	5 0	<u> </u>	<u> </u>	19	52	9	LL	0
	0	52	Э	78	0	0	0	0	~	15	٢	20	19	9	0	23	0	1	3	0	-	0	17	54	9	78	0
	0	53	0	76	0	0	0	0	~	11	٢	34	18	9	0	22	0	1.	3 1	5 0		0	0	53	9	LL	0
	22	53	3	LL	0	0	0	C	~	10	٢	19	19	9	0	22	0	1:	3	5 0	_	0	0	53	9	LL	0

show that (i) higher-order moment makes the proposed model more widely applicable; (ii) the proposed model and the corresponding algorithm are effective.

Author Contributions TY involved in conceptualization, methodology, writing—original draft, visualization. XH took part in methodology, writing—review and editing, validation, supervision, funding acquisition. KRH involved in software, visualization, validation, writing—review and editing.

**Funding** This work was supported by Fundamental Research Funds for the Central Universities No. FRF-MP-20-12.

**Data availability** Enquiries about data availability should be directed to the authors.

#### **Declarations**

**Conflict of interest** The authors declare that they have no conflict of interest.

**Ethical approval and Informed Consent** This article does not contain any studies with human participants or animals performed by any of the authors.

# References

- Andriosopoulos K, Nomikos N (2014) Performance replication of the spot energy index with optimal equity portfolio selection: evidence from the UK, US and Brazilian markets. Eur J Oper Res 234(2):571–582
- Beasley JE, Meade N, Chang TJ (2003) An evolutionary heuristic for the index tracking problem. Eur J Oper Res 148(3):621–643
- Chen W, Li D, Lu S, Liu W (2019) Multi-period mean-semivariance portfolio optimization based on uncertain measure. Soft Comput 23(15):6231–6247
- Chen Z, Zhang L, Tian G, Nasr EA (2020) Economic maintenance planning of complex systems based on discrete artificial bee colony algorithm. IEEE Access 8:108062–108071
- Filippi C, Guastaroba G, Speranza M (2016) A heuristic framework for the bi-objective enhanced index tracking problem. Omega 65:122– 137
- Gao Y (2011) Shortest path problem with uncertain arc lengths. Comput Math Appl 62(6):2591–2600
- Gaustaroba G, Mansini R, Ogryczak W, Speranza MG (2016) Linear programming models based on omega ratio for the enhanced index tracking problem. Eur J Oper Res 251(3):938–956
- Guastaroba G, Speranza MG (2012) Kernel search: an application to the index tracking problem. Eur J Oper Res 217(1):54–68
- Huang X (2010) Portfolio analysis: from probabilistic to credibilistic and uncertain approaches. Springer-Verlag, Berlin
- Huang X, Di H (2020) Uncertain portfolio selection with mental accounts. Int J Syst Sci 51(12):2079–2090
- Huang X, Yang T (2020) How does background risk affect portfolio choice: An analysis based on uncertain mean-variance model with background risk. J Bank Financ 111:105726
- Huang X, Zhao T (2014) Mean-chance model for portfolio selection based on uncertain measure. Insur: Math Econ 59:243–250
- Jorion P (2002) Enhanced index funds and tracking error optimization. Unpublished Paper, Graduate School of Management, University of California at Irvine
- Jorion P (2003) Portfolio optimization with tracking-error constraints. Financ Anal J 59(5):70–82

- Karaboga D (2005) An idea based on honey bee swarm for numerical optimization. Erciyes University, Engineering Faculty, Computer Engineering Department, Tech. rep
- Karaboga D, Basturk B (2007) A powerful and efficient algorithm for numerical function optimization: artificial bee colony (abc) algorithm. J Glob Optim 39:459–471
- Lejeune MA (2012) Game theoretical approach for reliable enhanced indexation. Decis Anal 9(2):146–155
- Liu B (2007) Uncertainty theory. Springer-Verlag, Berlin
- Liu B (2010) Uncertainty theory: a branch of mathematics for modeling human uncertainty. Springer-Verlag, Berlin
- Liu Y, Liu B (2022) Residual analysis and parameter estimation of uncertain differential equations. Fuzzy Optim Decis Mak 23:1–18
- Mao JC (1970) Survey of capital budgeting: theory and practice. J Financ 25(2):349–360
- Maringer D, Parpas P (2009) Global optimization of higher order moments in portfolio selection. J Glob Optim 43(2):219–230
- Masdari M, Barshande S, Ozdemir S (2019) Cdabc: chaotic discrete artificial bee colony algorithm for multi-level clustering in largescale wsns. J Supercomput 75:7174–7208
- Ning Y, Ke H, Fu Z (2015) Triangular entropy of uncertain variables with application to portfolio selection. Soft Comput 19(8):2203– 2209
- Qin Z, Kar S, Zheng H (2016) Uncertain portfolio adjusting model using semiabsolute deviation. Soft Comput 20(2):717–725
- Roll R (1992) A mean/variance analysis of tracking error. J Portf Manag 18(4):13–22
- Rudolf M, Wolter HJ, Zimmermann H (1999) A linear model for tracking error minimization. J Bank Financ 23(1):85–103
- Sant'Anna LR, Filomena TP, Caldeira JF (2017) Index tracking and enhanced indexing using cointegration and correlation with endogenous portfolio selection. Q Rev Econ Financ 65:146–157
- Strub O, Baumann P (2018) Optimal construction and rebalancing of index-tracking portfolios. Eur J Oper Res 264(1):370–387
- Wang X, Huang X (2019) A risk index to model uncertain portfolio investment with options. Econ Model 80:284–293
- Xuan H, Zhang H, Li B (2019) An improved discrete artificial bee colony algorithm for flexible flowshop scheduling with step deteriorating jobs and sequence-dependent setup times. Math Probl Eng 2019:1–13
- Xue L, Di H, Zhao X, Zhang Z (2019) Uncertain portfolio selection with mental accounts and realistic constraints. J Comput Appl Math 39:42–52
- Yao K (2015) A formula to calculate the variance of uncertain variable. Soft Comput 19:2947–2953
- Ye T, Liu B (2022) Uncertain hypothesis test for uncertain differential equations. Fuzzy Optim Decis Mak 22:1–17
- Ye T, Liu B (2022) Uncertain significance test for regression coefficients with application to regional economic analysis. Commun Stat-Theory Methods 52:1–18
- Yu L, Zhang S, Zhou XY (2006) A downside risk analysis based on financial index tracking models. In: Stochastic finance, Springer, pp 213–236
- Zhang P (2019) Multiperiod mean absolute deviation uncertain portfolio selection with real constraints. Soft Comput 23(13):5081–5098

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.