Mean-chance model for portfolio selection based on uncertain measure

Xiaoxia Huang*, Tianyi Zhao

Dongling School of Economics and Management, University of Science and Technology Beijing, Beijing 100083, China

ABSTRACT

This paper discusses a portfolio selection problem in which security returns are given by experts' evaluations instead of historical data. A factor method for evaluating security returns based on experts' judgment is proposed and a mean-chance model for optimal portfolio selection is developed taking transaction costs and investors' preference on diversification and investment limitations on certain securities into account. The factor method of evaluation can make good use of experts' knowledge on the effects of economic environment and the companies' unique characteristics on security returns and incorporate the contemporary relationship of security returns in the portfolio. The use of chance of portfolio return failing to reach the threshold can help investors easily tell their tolerance toward risk and thus facilitate a decision making. To solve the proposed nonlinear programming problem, a genetic algorithm is provided. To illustrate the application of the proposed method, a numerical example is also presented.

1. Introduction

Since Markowitz (1952), quantitative research on portfolio selection has attracted many scholars, and variance has been a very popular risk measurement. In his models, Markowitz proposed that expected return could be regarded as the investment return and variance the risk because the greater the variance value, the greater deviation from the expected return and thus the less likely that investors can obtain the expected return. He proposed that for a given level of variance, an optimal portfolio was obtained when the expected return was maximized; or for a given expected return, the optimal portfolio was obtained when the variance value was minimized. Since Markowitz, numerous portfolio selection models have been developed to improve and extend the mean–variance method. For example, DeMiguel and Nogales (2009) constructed the portfolio using M- and S-robust estimators instead of the classical median and mean absolute deviation to obtain the portfolio with better stability properties than the traditional minimum variance portfolio. Soleimani et al. (2009) considered minimum transaction lots, cardinality constraints and market sector capitalization as extra constraints and used a genetic algorithm (GA) to solve the problem. Anagnostopoulos and Mamanis (2010) formulated a tri-objective model to find tradeoffs between risk, return and the number of securities in the portfolio, considering quantity and class constraints. Considering that security returns are given by experts’ evaluations rather than historical data in some situations, Huang (2012a) studied a new mean–variance and mean–semivariance methods for portfolio selection in these situations.

Though variance is a popular risk measurement, it is not intuitive. Since investors are better at stating their threshold levels for their goals and the maximum tolerable chances of failing to reach than variance, scholars have showed great interest in selecting portfolios using probability of failure to reach the threshold return level or its another version, i.e., value-at-risk (VaR), as risk measurement to control risk. Some recent examples include Tsao (2010) who developed an evolutionary multi-objective approach to construct the mean–VaR efficient frontier, Durand et al. (2011) who provided an extension of evidence supporting the empirical validity and tractability of the mean–VaR efficiency concept, and Zymler et al. (2013) who developed two tractable conservative approximations for the VaR of a derivative portfolio by evaluating the worst-case VaR over all return distributions of the derivative underliers with given first- and second-order moments. Das et al. (2010), Alexander and Baptista (2011) and Baptista (2012) have also researched the way of adding probability of failing to reach the threshold return level as an extra risk measurement into Markowitz’s mean–variance theory to help investors easily determine their risk-aversion coefficient and control their portfolio investment risk.
All these researches assumed that investors have perfect information and future security returns can be fairly well reflected by the historical data. However, since security market is complex and economic environment is changing, there are situations where security returns can hardly be reflected by the historical data. In addition, nowadays many stocks are newly listed in the market. In all these situations investors lack suitable historical data. They cannot predict the security returns according to historical data but have to invite some domain experts to evaluate their belief degree toward security returns. Kahneman and Tversky (1979) have found that human beings tend to give too much weight to unlikely events. Thus, unless further suitable observed data can be obtained to revise the belief degree, subjective probability sometimes fails to model the belief degree. So far, some theories have been proposed to deal with men’s belief degree toward an imprecise number such as possibility theory (Zadeh, 1978) and Dempster–Shafer theory (Dempster, 1967; Shafer, 1976). In 2007, an uncertainty theory was founded by Liu (2007) to deal with belief degree based on uncertain measure when little historical data are available. Nowadays, uncertainty theory has been used in solving problems in many optimization areas such as vehicle routing and project scheduling problems (Liu, 2010), shortest path problem (Gao, 2011), multi-national project selection problem (Zhang et al., 2011), facility location problem (Gao, 2012), and inventory problem (Qin and Kar, 2013), etc. Especially, uncertainty theory was first systematically introduced into portfolio selection by Huang (2010), thus producing a theory of uncertain portfolio selection. After that, uncertain mean–semivariance model (Huang, 2012a) was discussed, uncertain risk curve (Huang, 2011) and uncertain risk index (Huang, 2012b) methods were proposed, and uncertain portfolio adjustment problem (Huang and Ying, 2013) based on risk index was studied. In this paper, we will go on exploring using uncertainty theory to develop a mean-chance method using chance of failing to reach the preset threshold level as risk measurement for portfolio selection in the situation where no suitable historical data are available and security returns are given by experts’ judgments. We will first propose a factor method for evaluation of security returns based on experts’ evaluations and then develop a mean-chance model. The factor method can make good use of experts’ knowledge on the effects of economic environment and the companies’ technological and managerial uniqueness on security returns and incorporate the contemporary relationship of security returns in the portfolio. Since it is usually easy for investors to pre-give a return threshold and the tolerance toward the chance of failing to reach this threshold level, the use of chance of failing to reach a predetermined return threshold level as risk measurement can help investors easily tell their risk tolerance level and thus facilitate an easy decision making. Since in reality there usually exist transaction costs and investors are often required by law or by their own preference to make a portfolio investment with required diversification level, and they usually prefer to make investment limitations on certain securities, we will incorporate these constraints in the model.

The rest of the paper is organized as follows. In Section 2 we will first propose a factor method for evaluation of security returns based on experts’ judgments. Then we will develop a mean-chance model in Section 3 and present the deterministic equivalents of it in Section 4. Since the proposed model is a complex nonlinear programming model, we will present a GA for solving the problem in Section 5. As an illustration we will offer a numerical example in Section 6. Finally, in Section 7 we will give some concluding remarks. For better understanding of the paper, we will also briefly review some fundamentals of uncertain variables in the Appendix.

### 2. Factor method of evaluating security returns

Economic environment affects all security returns. In addition, the unique technology, management and other characteristics of a company, which distinguish itself from others, also affect the company’s security returns independently from economic environment. Therefore, we can use two-factor model to evaluate a company’s security return. One factor is economic environment factor which reflects the effect of economic environment on all security returns. It is measured in the paper by the return rate of security market index. Another factor is the company’s uniqueness factor which reflects the effect of the company’s unique technology, management and other characteristics on the company’s security returns. Since the company’s unique characteristics can result in its unique profitability from others, we compound the company’s short and long term profitability to measure the company’s uniqueness. Return on equity equals net income divided by shareholders’ equity. It measures a company’s ability to turn assets into profits. Higher values are generally favorable meaning that the company is efficient in generating income on new investment. Main business income is the income of the company’s regular and main business services. Main business operating margin reflects the company’s basic profitability. It tells the contribution of main business service profit to the total profit of the company and is a complementary indicator of the company’s profitability. When increasing rate of return on equity and increasing rate of main business operating margin indicate the company’s short term profitability, increasing rate of five-year net profit reflects the company’s long term profitability. By allocating different weights to these short and long term profitability indicators, we get the uniqueness of the company. For convenience of expression, we use $F$ to denote the factor of economic environment and $f_i$ the $i$th company’s uniqueness factor. In this paper we regard the two factors and security return rates as uncertain variables and propose that the $i$th security’s return rate $r_i$ is linearly correlated to $F$ and $f_i$, i.e., $r_i = a_i + b_F F + c f_i$. Since $f_i$ is the factor reflecting the $i$th company’s uniqueness, it is assumed that $f_i$ is independent of $F$, and $f_i$ and $f_j$ are independent of each other where $i \neq j$.

Without loss of generality, we now introduce via one security the method of evaluating the security return rate $r = a + bF + cf$ based on experts’ evaluations. The first step of evaluating the security return rates will be if $F_i/k$ is obtained. Where $i = 1, 2, \ldots, m, k = 1, 2, \ldots, p$, respectively, and $F_{i/1} \leq F_{i/2} \leq \cdots \leq F_{i/p}$. If for any $k$th value, max$_{1 \leq i \leq m} F_i/k - \min_{1 \leq i \leq m} F_i/k > \varepsilon$, the domain experts are given the sum of the results and the reasons for these results, and then are asked to provide their revised estimations of the $p$ numbers of values that the factor $F$ may take. It is believed that during the process the opinions of the experts will converge to an appropriate answer. When max$_{1 \leq i \leq m} F_i/k - \min_{1 \leq i \leq m} F_i/k \leq \varepsilon$, we calculate

$$F_k = \frac{1}{m} \sum_{i=1}^{m} F_i/k, \quad k = 1, 2, \ldots, p.$$ 

Then the data set $(F_1, F_2, \ldots, F_p)$ that the factor $F$ may take are obtained. In a similar way, the data set $(f_1, f_2, \ldots, f_p)$ that the factor $f$ may take can also be obtained.

Next, the experts are asked to evaluate what the security return rates will be if $F = F_k$ and $f = f_k, k = 1, 2, \ldots, p$. That is, from the experts the data set of

$$(y_{i/1}, F_1, f_1), (y_{i/2}, F_2, f_2), \ldots, (y_{i/p}, F_p, f_p), \quad i = 1, 2, \ldots, m$$
are obtained, where $y_i/1 \leq y_i/2 \leq \cdots \leq y_i/p$. Since the experts are regarded equally knowledgable, we allocate the same weight to each expert's estimates of security return rates and aggregate the $m$ experts' estimates as follows:

$$y_k = \frac{1}{m} \sum_{i=1}^{m} y_i/k, \quad k = 1, 2, \ldots, p.$$  

Then we get a new set of data for the security return rates as follows:

$$(y_1, F_1, f_1), (y_2, F_2, f_2), \ldots, (y_p, F_p, f_p).$$

To find the parameters $a, b,$ and $c$, we adopt the principle of least squares which says that the unknown parameters $a, b,$ and $c$ can be obtained via solution of the following minimization problem

$$\min_{a,b,c} \sum_{i=1}^{p} (y_i - a - b F_i - c f_i)^2.$$  

Let

$$Y = \begin{pmatrix} y_1 \\ \vdots \\ y_p \end{pmatrix}, \quad X = \begin{pmatrix} 1 & F_1 & f_1 \\ \vdots & \vdots & \vdots \\ 1 & F_p & f_p \end{pmatrix}, \quad \hat{\beta} = \begin{pmatrix} \hat{a} \\ \hat{b} \\ \hat{c} \end{pmatrix},$$

where $\hat{a}, \hat{b}, \hat{c}$ are least squares estimators of $a, b,$ and $c$, respectively. Then the least square estimators are given by

$$\hat{\beta} = (X'X)^{-1}X'Y$$

where $X'$ is the transpose of $X$.

Please note that in random environment, the sample data sets $(y_k, F_k, f_k)$ are obtained from historical data, where $F_k$ is the $k$th duration return rate of market index in the past, $f_k$ the company's $k$th duration uniqueness index number, and $y_k$ the observed security return rate at the same $k$th duration, $k = 1, 2, \ldots, p$, respectively. While in our situation, the sample data sets $(y_k, F_k, f_k), k = 1, 2, \ldots, p,$ are given by experts' estimations. In the process of obtaining coefficients $a, b,$ and $c$, evaluation focus is on the relationship between the pair values of the two factors and the security return rates. The experts are only required to answer the questions “if $F$ takes $F_i$ number and $f$ takes $f_i$ number, what do you think the security return rate will be?” They need not estimate the occurrence chance levels of the events being equal to or less than $F_i$ or $f_i$ values. However, in the process of obtaining the uncertainty distributions of the uncertain factors $F$ and $f$, they should also give their estimates of occurrence chance levels of the events being equal to or less than $F_i$ or $f_i$ values. The way of obtaining the uncertainty distribution of an uncertain variable can be referred to Liu (2010) and Huang (2012a). For easy understanding of the paper, we briefly introduce the process below.

Take the evaluation of the uncertainty distribution of the factor $F$ as an example. After getting the data set $(F_1, F_2, \ldots, F_p)$ that the factor $F$ may take, the experts are asked to estimate the occurrence chance levels of the events that $F$ being equal to or less than the given values. Let $\alpha_{i/k}$ denote the $i$th expert's estimate of occurrence chance of the event that $F$ being equal to or less than $F_k$. Then from $m$ experts the data sets of

$$(F_1, \alpha_{i1}), (F_2, \alpha_{i2}), \ldots, (F_p, \alpha_{ip}), \quad i = 1, 2, \ldots, m$$

are obtained, where

$$F_1 < F_2 < \cdots < F_p, \quad 0 \leq \alpha_{i1} \leq \alpha_{i2} \leq \cdots \leq \alpha_{ip} \leq 1.$$  

By allocating same weight to each expert and aggregate the $m$ experts' estimates, we get a new set of data for $F$ as follows:

$$(F_1, \alpha_1), (F_2, \alpha_2), \ldots, (F_p, \alpha_p), \quad F_1 \leq F_2 \leq \cdots \leq F_p, \quad 0 \leq \alpha_1 \leq \alpha_2 \leq \cdots \leq \alpha_p \leq 1,$$

where $\alpha_k = \frac{1}{m} \sum_{i=1}^{m} \alpha_{ik}, k = 1, 2, \ldots, p,$ respectively. If $F$ is believed to have certain uncertainty distribution, we can use least squares to obtain the parameters of the distribution. For example, if $F$ is believed to be a normal uncertain variable $\mathcal{N}(\mu, \sigma)$, we use the principle of least squares to find the parameters $\mu$ and $\sigma$. That is, we get the estimated $\hat{\mu}$ and $\hat{\sigma}$ by solving either of the following two minimization problems:

$$\min_{\mu, \sigma} \sum_{i=1}^{p} \left( \Phi(F_i|\mu, \sigma) - \alpha_i \right)^2,$$

where

$$\Phi(F_i|\mu, \sigma) = \left(1 + \exp \left(\frac{\mu - F_i}{\sqrt{3}\sigma}\right)\right)^{-1},$$

or

$$\min_{\mu, \sigma} \sum_{i=1}^{p} \left( \Phi^{-1}(\alpha_i|\mu, \sigma) - F_i \right)^2,$$

where

$$\Phi^{-1}(\alpha_i|\mu, \sigma) = \mu + \frac{\sqrt{3}\sigma}{\pi} \ln \frac{\alpha_i}{1 - \alpha_i}.$$  

The solution is

$$\hat{\mu} = \hat{F} - \sqrt{3}\hat{\sigma} \frac{\sum_{i=1}^{p} \ln \frac{\alpha_i}{1 - \alpha_i}}{p\pi}, \quad \hat{\sigma} = \frac{\sqrt{3}\pi p}{3} \frac{\hat{F} \sum_{i=1}^{p} \ln \frac{\alpha_i}{1 - \alpha_i} - p \sum_{i=1}^{p} \ln \frac{\alpha_i}{1 - \alpha_i}}{\left(\sum_{i=1}^{p} \ln \frac{\alpha_i}{1 - \alpha_i}\right)^2 - p \sum_{i=1}^{p} \ln \frac{\alpha_i}{1 - \alpha_i}^2},$$

where $\hat{F} = \frac{1}{p} \sum_{i=1}^{p} F_i$.

### 3. Mean-choke portfolio selection model

In reality, there usually exist transaction costs when buying and selling securities. Therefore, we take them into account in this paper. Let $t_i$ denote the uncertain return rates of the $i$th securities which are defined as $r_i = (p_i + 0.0)/p_{i0}, i = 1, 2, \ldots, n$, respectively, where $p_{i0}$ are the estimated closing prices of the securities $i$ in the future time, $p_i$ the closing prices of the securities $i$ at present. Let $t_i$ be the transaction cost rate of buying and $t_i$ the transaction cost rate of selling. Then the investors’ actual return rate from security $i$ is

$$\xi_i = \frac{p_i(1 - t_i) - p_{i0}(1 + t_i)}{p_{i0}(1 + t_i)} = \frac{1 - t_i}{1 + t_i} - \frac{t_i + t_i}{1 + t_i}.$$  

Let $F$ be the uncertain economic environment factor given by experts, and $f_i$ the $i$th company's uncertain uniqueness factor given by experts. By using the method introduced in Section 2, the return rates of security $i$ can be obtained by $r_i = a_i + b_i F + c_i f_i, i = 1, 2, \ldots, n$, respectively. Then, according to Eq. (3), the investors’ actual uncertain return rate is

$$\xi_i = \frac{1 - t_i}{1 + t_i} (a_i + b_i F + c_i f_i) - \frac{t_i + t_i}{1 + t_i}.$$  

Since investors are usually required by law or by their own preference to make a portfolio investment with required diversification level and they usually have some preference and requirement on investment limitations on certain securities, we consider the requirement and preference in the model. Let $x_i$ denote the investment proportions in securities $i$, and $y_i$ be binary variables defined by

$$y_i = \begin{cases} 1, & \text{if security } i \text{ is selected}, \\ 0, & \text{otherwise}, \end{cases} \quad i = 1, 2, \ldots, n.$$
Suppose that the investors pre-give an investment threshold return level $s$ and the tolerable chance level $\gamma$ of the portfolio return failing to reach the threshold. Furthermore, they require that the total number of the selected securities in the portfolio should not be less than a preset number $d$, and the minimum investment proportions on securities belonging to set $L_p$ should not be higher than $l_i$ where $l_i > 0$ and $\sum_{i \in L_p} b_i < 1$, and the maximum investment proportions on securities belonging to set $H_p$ should not be less than $h_i$ where $h_i < 1$. Then in order to meet their requirements and to pursue the maximum expected portfolio return under risk control, the investors should select the portfolio according to the following model:

$$
\max E \left[ \sum_{i=1}^{n} x_i \frac{1 - t_i}{1 + t_b} (a_i + b_i F + c_i f_i) - \sum_{i=1}^{m} x_i \frac{t_b + t_b}{1 + t_b} \right]
$$
subject to:

$$
M \left( \sum_{i=1}^{n} x_i \frac{1 - t_i}{1 + t_b} (a_i + b_i F + c_i f_i) - \sum_{i=1}^{m} x_i \frac{t_b + t_b}{1 + t_b} \leq s \right) \leq \gamma \ (I)
$$

$$
\sum_{i=1}^{n} y_i \geq d, \quad x_i \geq l_i, \quad x_i \leq h_i, \quad \forall i \in L_p, H_p
$$

$$
\sum_{i=1}^{n} x_i = 1, \quad y_i \in [0, 1], \quad i = 1, 2, \ldots, n
$$

(5)

where $M$ is the uncertain measure and $E$ the expected value operator of the uncertain variables. Please note that the constraint $\sum_{i=1}^{n} y_i \geq d$ is diversification constraint. Though superficially, it seems that the constraint has nothing to do with the decision variables $x_i$, we can see from the definition formula (4) that the constraint does have connection with the decision variables $x_i$. In fact, the model (5) is equivalent to the following model:

$$
\max E \left[ \sum_{i=1}^{n} x_i \frac{1 - t_i}{1 + t_b} (a_i + b_i F + c_i f_i) - \sum_{i=1}^{m} x_i \frac{t_b + t_b}{1 + t_b} \right]
$$
subject to:

$$
M \left( \sum_{i=1}^{n} x_i \frac{1 - t_i}{1 + t_b} (a_i + b_i F + c_i f_i) - \sum_{i=1}^{m} x_i \frac{t_b + t_b}{1 + t_b} \leq s \right) \leq \gamma \ (I)
$$

$$
y_i = 1, \quad \text{if } x_i > 0
$$

$$
y_i = 0, \quad \text{if } x_i = 0
$$

$$
\sum_{i=1}^{n} y_i \geq d
$$

$$
x_i \geq l_i, \quad \forall i \in L_p, H_p
$$

$$
x_i \leq h_i, \quad \forall i \in H_p
$$

$$
\sum_{i=1}^{n} x_i = 1
$$

$$
y_i \in [0, 1], \quad \forall i = 1, 2, \ldots, n
$$

(6)

4. Equivalents of the mean-chance model

**Theorem 1.** Let $\Psi$ denote the continuous uncertainty distribution function of the uncertain factor $F$ whose inverse function $\Psi^{-1}(\alpha)$ exists and is unique for each $\alpha \in (0, 1)$, and $\Phi_i$ the continuous uncertainty distribution function of the factor $f_i$ whose inverse function $\Phi_i^{-1}(\alpha)$ exists and is unique for each $\alpha \in (0, 1)$, $i = 1, 2, \ldots, n$, respectively. Let $\nu$ be the expected value of $F$ and $e_i$ the expected value of $f_i$, $i = 1, 2, \ldots, n$, respectively. Suppose $b_i \geq 0$ for $i = 1, 2, \ldots, m$ and $c_i \geq 0$ for $i = 1, 2, \ldots, k$, and $b_i < 0$ for $i = m + 1, m + 2, \ldots, n$ and $c_i < 0$ for $i = k + 1, k + 2, \ldots, n$. Then the mean-chance model (5) can be transformed into the following form:

$$
\max \sum_{i=1}^{n} x_i \left[ \frac{1 - t_i}{1 + t_b} (a_i + b_i e + c_i e_i) - \frac{t_b + t_b}{1 + t_b} \right]
$$
subject to:

$$
\sum_{i=1}^{n} (1 - t_i) a_i - t_b - t_b x_i + \sum_{i=1}^{m} \frac{1 - t_i}{1 + t_b} b_i \Psi^{-1}(\gamma) x_i
$$

$$
+ \sum_{i=m+1}^{n} \frac{1 - t_i}{1 + t_b} b_i \Psi^{-1}(1 - \gamma) x_i + \sum_{i=1}^{k} \frac{1 - t_i}{1 + t_b} c_i \Phi_i^{-1}(\gamma) x_i
$$

$$
+ \sum_{i=k+1}^{n} \frac{1 - t_i}{1 + t_b} c_i \Phi_i^{-1}(1 - \gamma) x_i \geq s
$$

(7)

**Proof.** The objective function of the model (7) can be obtained directly from Theorem 4 in the Appendix.

According to Eq. (12) in Theorem 3 in the Appendix and the monotonicity property of uncertain variable, the risk constraint (I) in model (5) can be transformed into the following form:

$$
\sum_{i=1}^{n} (1 - t_i) a_i - t_b - t_b x_i + \sum_{i=1}^{m} \frac{1 - t_i}{1 + t_b} b_i \Psi^{-1}(\gamma) x_i
$$

$$
+ \sum_{i=m+1}^{n} \frac{1 - t_i}{1 + t_b} b_i \Psi^{-1}(1 - \gamma) x_i + \sum_{i=1}^{k} \frac{1 - t_i}{1 + t_b} c_i \Phi_i^{-1}(\gamma) x_i
$$

$$
+ \sum_{i=k+1}^{n} \frac{1 - t_i}{1 + t_b} c_i \Phi_i^{-1}(1 - \gamma) x_i \geq s
$$

Thus, the theorem is proved.

According to Theorem 1, we can easily get the following theorem.

**Theorem 2.** When the factor $F$ is normal uncertain variable $F \sim \mathcal{N}(e, \sigma)$, and the uniqueness factor $f_i$ normal uncertain variable $f_i \sim \mathcal{N}(e_i, \sigma_i)$, $i = 1, 2, \ldots, n$, respectively, the mean-chance model (5) can be converted into the following equivalent:

$$
\max \sum_{i=1}^{n} x_i \left[ \frac{1 - t_i}{1 + t_b} (a_i + b_i e + c_i e_i) - \frac{t_b + t_b}{1 + t_b} \right]
$$
subject to:

$$
\sum_{i=1}^{n} (1 - t_i) a_i - t_b - t_b x_i + \sum_{i=1}^{m} \frac{1 - t_i}{1 + t_b} b_i \nu \ln \frac{1 - \gamma}{1 - \gamma} \sum_{i=1}^{n} \frac{1 - t_i}{1 + t_b} c_i \sigma_i x_i \geq s
$$

(8)

$$
\sum_{i=1}^{n} y_i \geq d
$$

$$
x_i \geq l_i, \quad \forall i \in L_p
$$

$$
x_i \leq h_i, \quad \forall i \in H_p
$$

$$
\sum_{i=1}^{n} x_i = 1
$$

$$
y_i \in [0, 1], \quad \forall i = 1, 2, \ldots, n.$$
5. Genetic algorithm

The equivalent models (7) and (8) are complex nonlinear programming models and are difficult to use traditional methods to solve them. GA is good at solving complex optimization problems. In this section, we will propose a GA for solving the proposed mixture integer portfolio selection problem.

5.1. Representation structure

We use a chromosome \( C = (g_1, g_2, \ldots, g_n, g_{n+1}, g_{n+2}, \ldots, g_{2n}) \) to represent the solution where \( g_i, i = 1, 2, \ldots, n \) represent \( x_i \), \( i = 1, 2, \ldots, n \), and \( g_{n+i}, i = 1, 2, \ldots, n \) represent \( y_i \), \( i = 1, 2, \ldots, n \), respectively. The solution between \( g_i \) and \( x_i \) for \( i = 1, 2, \ldots, n \) are

\[
x_i = \frac{g_i}{g_1 + g_2 + \cdots + g_n}, \quad i = 1, 2, \ldots, n,
\]

which ensures that \( x_1 + x_2 + \cdots + x_n = 1 \) always holds, and the solution between \( g_{n+i} \) and \( y_i \) for \( i = 1, 2, \ldots, n \) are

\[
y_i = g_{n+i}, \quad i = 1, 2, \ldots, n.
\]

5.2. Initialization process

Initialization of \( pop_size \) feasible chromosomes are completed through the following process:

First, generate a random point \( (g_1, g_2, \ldots, g_n) \) from the hyper-cube \([0, 1]^n\).

Second, randomly generate an integer vector \( (g_{n+1}, g_{n+2}, \ldots, g_{2n}) \) from integer set \([0, 1]\). If \( g_{n+i} = 1 \), let \( g_i = g_i \), otherwise, if \( g_{n+i} = 0 \), let \( g_i = g_i \), where \( i = 1, 2, \ldots, n \), respectively. Then, if \( g_i > 0 \), let \( g_{n+i} = 1 \), if \( g_i = 0 \), then \( g_{n+i} > 0 \), if \( g_i = 0 \). And in the meantime if \( g_{n+i} = 1 \), then \( g_i > 0 \), if \( g_{n+i} = 0 \), then \( g_i = 0 \). In other words, we have made it sure that when the ith security is selected, \( x_i = 1 \) and \( x_i > 0 \), and when the ith security is not selected, \( y_i = 0 \) and \( y_i = 0 \).

Third, if \( \sum_{i=1}^{n} g_i = 0 \), turn back to the first step. Otherwise, let \( x_i = g_i / \sum_{i=1}^{n} g_i \), \( i = 1, 2, \ldots, n \), to ensure that \( x_1 + x_2 + \cdots + x_n = 1 \) always holds.

Fourth, check if the chromosome at this stage is feasible. If the chromosome is feasible, then a feasible initial chromosome is obtained. Otherwise, repeat the above steps until a feasible initial chromosome is obtained. The feasibility of the chromosome \( (g_1, g_2, \ldots, g_n, g_{n+1}, g_{n+2}, \ldots, g_{2n}) \) is checked as follows:

\[
\text{If } \sum_{i=1}^{n} y_i < d, \quad \text{return 0};
\]

\[
\text{If } \sum_{i=1}^{n} y_i + \sum_{i=1}^{n} x_i \leq \sum_{i=1}^{n} 1 - t_1 b_i \psi^{-1}(1) x_i + \sum_{i=1}^{n} 1 - t_2 c_i \Phi_{-1}(1) x_i + \sum_{i=1}^{n} 1 - t_3 c_i \Phi_{-1}(1 - y) x_i < s, \quad \text{return 0};
\]

\[
\text{If } x_i < l_i, \quad i \in L_p \quad \text{return 0};
\]

\[
\text{If } x_i > h_i, \quad i \in H_p \quad \text{return 0};
\]

in which 1 means feasible, and 0 non-feasible. In case when \( F \) and \( f_i \), \( i = 1, 2, \ldots, n \) are all normal uncertain variables, the feasibility of the chromosome \( (g_1, g_2, \ldots, g_n, g_{n+1}, g_{n+2}, \ldots, g_{2n}) \) is checked as follows:

\[
\text{If } \sum_{i=1}^{n} y_i < d, \quad \text{return 0};
\]

\[
\text{If } \sum_{i=1}^{n} x_i \left[ \frac{1 - t_1}{1 + t_3} (a_i + bh e + c_i x_i) - t_2 + t_3 \right] + \sqrt{3(1 - t_3)} \ln \frac{\gamma}{1 - \gamma} \sum_{i=1}^{n} x_i | h \sigma_i + | c_i x_i | | < s, \quad \text{return 0};
\]

\[
\text{If } x_i < l_i, \quad i \in L_p \quad \text{return 0};
\]

\[
\text{If } x_i > h_i, \quad i \in H_p \quad \text{return 0};
\]

5.3. Selection process

The selection of chromosomes is done by spinning roulette wheel, which is a fitness-proportional selection. The selection process of the chromosomes is done as follows:

First, Compute the objective values of all the chromosomes and re-arrange them such that the chromosome with higher objective value is given smaller ordinal number. That is, after arrangement, \( C_1, C_2, \ldots, C_{pop_size} \) are the chromosomes from good to bad.

Second, compute the rank-based evaluation function, denoted by \( eval(C_i) \), for each chromosome. The rank-based evaluation function is defined as follows,

\[
eval(C_i) = \nu(1 - \nu)^{j-1}, \quad j = 1, 2, \ldots, pop_size,
\]

where \( \nu \in (0, 1) \) is a parameter, and \( j = 1 \) means the best individual, \( j = pop_size \) the worst individual.

Third, compute the cumulative probability \( P_i \) for each chromosome \( C_i \),

\[
P_0 = 0, \quad P_i = \sum_{k=1}^{i} \text{Eval}(C_k), \quad k = 1, \ldots, pop_size.
\]

Then divide all \( P_i \)'s, \( k = 1, 2, \ldots, pop_size \), by \( P_{pop_size} \) such that \( P_{pop_size} = 1 \).

Fourth, randomly generate a real number \( m \) from the interval \( (0, P_{pop_size}) \). Select the jth chromosome \( C_j \) \( (1 \leq j \leq pop_size) \) if \( P_{j-1} < m \leq P_j \). It can be seen that the probability of the number \( m \) falling in \( (P_{j-1}, P_j) \) is the probability that jth chromosome will be selected. The probability is proportional to the fitness of the chromosome.

Fifth, repeat the fourth step \( pop_size \) times, then \( pop_size \) copies of chromosomes are selected.

5.4. Crossover operation

Determine a parameter \( P_c \) as the probability of crossover first. The chromosomes \( C_i \) are selected as parents when the randomly generated real number \( h \) from \([0, 1]\) is less than \( P_c \) at the jth selection, where \( j = 1, 2, \ldots, pop_size \). Let \( C_1, C_2, C_3, \ldots \) denote the selected parents. Crossover operation on each pair is illustrated on \( C_1 = (F_1', R_1') \) and \( C_2 = (F_2', R_2') \) as follows, where \( F_1' = (g_1^{(1)}, g_2^{(1)}, \ldots, g_n^{(1)}) \) and \( F_2' = (g_1^{(2)}, g_2^{(2)}, \ldots, g_n^{(2)}) \) denote the front \( n \) numbers of genes of the chromosomes \( C_1' \) and \( C_2' \), respectively, and \( R_1' = (g_{n+1}^{(1)}, g_{n+2}^{(1)}, \ldots, g_{2n}^{(1)}) \) and \( R_2' = (g_{n+1}^{(2)}, g_{n+2}^{(2)}, \ldots, g_{2n}^{(2)}) \) denote the rear \( n \) numbers of genes of the chromosomes \( C_1' \) and \( C_2' \), respectively.

First, cross front \( n \) numbers of genes of the pair chromosomes \( C_1' \) and \( C_2' \) Generate a random number \( e \) from the open interval \((0, 1)\). Then the front \( n \) numbers of genes are crossed by \( F_1'' = e \cdot F_1' + (1 - e) \cdot F_2', F_2'' = (1 - e) \cdot F_1' + e \cdot F_2' \).
Second, cross rear $n$ numbers of genes of $C_1'$ and $C_2'$, First, randomly generalize one integer $k$ such that $n+1 < k \leq 2n$. Then, exchange the genes of $R'_1$ and $R'_2$ between $k$ and $2n$ and produce two children’s rear part as follows.

$R'_1 = (g_{n+1}^{(1)}, g_{n+2}^{(1)}, \ldots, g_{n+k-1}^{(1)}, g_{n+k}^{(2)}, g_{n+k+1}^{(2)}, \ldots, g_{2n}^{(2)})$.

$R'_2 = (g_{n+1}^{(2)}, g_{n+2}^{(2)}, \ldots, g_{n+k-1}^{(2)}, g_{n+k}^{(1)}, g_{n+k+1}^{(1)}, \ldots, g_{2n}^{(1)})$.

Third, for each $C_1'' = (F''_1, R'_1)$ and $C_2'' = (F''_2, R'_2)$, let $g_{n+i} = 1$ if $g_i > 0$; Otherwise, let $g_{n+i} = 0$ if $g_i = 0$, where $i = 1, 2, \ldots, n$, respectively. If $\sum_{i=1}^{n} g_i = 0$, go back to the first step of the crossover. Otherwise, let $x_i = g_i / \sum_{i=1}^{n} g_i$ and $y_i = g_{n+i}, i = 1, 2, \ldots, n$, respectively. Then check if the new pair of offsprings $C_1''$ and $C_2''$ are feasible. If they are checked to be feasible, we take them as children and replace their parents with them; otherwise, we keep the feasible one if it exists, and then redo the crossover operation until two feasible children are obtained or a given number of cycles is finished. In this case, we only replace the parents with the feasible children.

5.5. Mutation operation

Determine a parameter $P_m$ as the probability of mutation first. The chromosomes $C_1$ are selected as parents when the randomly generated real number $h$ from $[0, 1]$ is less than $P_m$ at the jth selection, where $j = 1, 2, \ldots, pop\_size$. Mutation operation is done on each selected parent $C' = (F', R')$ as follows, where $F' = (g_1, g_2, \ldots, g_n)$ denote the front $n$ numbers of genes of the chromosome $C'$ and $R' = (g_{n+1}, g_{n+2}, \ldots, g_{2n})$ the rear $n$ numbers of genes of the chromosome $C'$.

First, mutate the front $n$ numbers of genes of the chromosome. Randomly choose a mutation direction $D$ in $\mathbb{R}^n$. Let $M$ be an appropriately large positive number. Mutate $F'$ by $F' + M \cdot D$ and a new front part of the chromosome $F''$ is formed.

Second, mutate the rear $n$ numbers of genes of the chromosome. First, randomly choose a mutation position $k$ such that $n+1 < k \leq 2n$. Then, initialize $g_{n+k}, g_{n+k+1}, \ldots, g_{2n}$ from integer set $\{0, 1\}$, and a new rear part of the chromosome $R''$ is formed.

Third, for $C'' = (F'', R'')$, let $g_{n+i} = 1$ if $g_i > 0$; Otherwise, let $g_{n+i} = 0$ if $g_i = 0$, where $i = 1, 2, \ldots, n$, respectively. If $\sum_{i=1}^{n} g_i = 0$, go back to the first step of the mutation. Otherwise, let $x_i = g_i / \sum_{i=1}^{n} g_i$ and $y_i = g_{n+i}, i = 1, 2, \ldots, n$, respectively. Then check if the offspring $C''$ is feasible. If it is feasible, then replace the parent $C'$ with it. Otherwise, repeat the above mutation process until the feasible child is obtained.

### Genetic algorithm
After selection, crossover and mutation, the new population is ready for its next evaluation. The GA will continue until a given number of cyclic repetitions of the above steps is met. We summarize the algorithm as follows.

**Step 1.** Input the parameters of GA: pop\_size, $P_c$, $P_m$, $\nu$.

**Step 2.** Initialize feasible pop\_size chromosomes.

**Step 3.** Select the chromosomes by spinning the roulette wheel.

**Step 4.** Update the chromosomes by crossover and mutation operations.

**Step 5.** Repeat the third and the fourth step for a given number of cycles.

**Step 6.** Take the best chromosome as the solution of portfolio selection.

6. An example

To illustrate the modeling idea and to test the effectiveness of the designed GA, we present an example here. The example is performed on a personal computer with the following parameters in the GA: the population size is 30, the probability of crossover $P_c = 0.6$, the probability of mutation $P_m = 0.6$, the parameter $\nu$ in the rank-based evaluation function is 0.05.

**Example 1.** Suppose that there are 24 stocks. The transaction cost rate of buying is 0.3% and the transaction cost rate of selling is 0.4%. The experts give their individual judgments about the relationship between monthly stock return rates and the economic environment factor and the company’s uniqueness factor independently. Then according to Eq. (1) in Section 2, the coefficients $a_i, b_i, c_i$ for stocks $i, i = 1, 2, \ldots, 24$, are obtained and given in Table 1. Furthermore, the experts believe that the values of the economic environment factor and uniqueness factor of each stock’s company are normal uncertain variables. By using the method provided in Section 2, the distribution of $F$ is $F \sim \mathcal{N}(0.022, 0.02)$ and the distributions of $f_i, i = 1, 2, \ldots, 24$, are obtained and given in Table 2.

Suppose that the investors set the return threshold at 2.0%, and give their tolerance toward the occurrence chance of the return failing to reach the threshold at 5%. They require that the selected portfolio should include at least 8 stocks. Though they do not have special preferences on different stocks and do not impose lower and upper bound limitations on stock investment, to avoid trivial investment they require that the minimum proportions of the selected stocks should not be less than 2.0%. Then in order to pursue the maximum expected return under risk control and diversification and investment limitation requirements, according
to model (8) in Section 4, the uncertain mean-chance model is built as follows:

\[
\begin{align*}
\max & \quad \sum_{i=1}^{24} 0.9930 (x_i a_i + 0.022b_i + c_i e_i) - 0.0070 \\
\text{subject to:} & \quad \sum_{i=1}^{24} 0.9930 (x_i a_i + 0.022b_i + c_i e_i) \\
& \quad - 1.6121 \sum_{i=1}^{24} x_i (|0.02b_i| + |c_i|\sigma_i) \geq 0.0270 \\
& \quad \sum_{i=1}^{24} y_i \geq 8 \\
& \quad \sum_{i=1}^{24} x_i = 1 \\
& \quad x_i \geq 0.02, \quad i = 1, 2, \ldots, 24 \\
& \quad y_i \in [0, 1], \quad i = 1, 2, \ldots, 24.
\end{align*}
\]

A run of the GA with 1000,000 generations shows that among 24 securities, in order to gain the maximum expected return under risk control and diversification and investment limitation requirements, the investors should assign their capital according to Table 3. The corresponding maximum expected return rate is 9.34%.

To test the effectiveness of the designed GA, we run the GA with different parameter settings. The parameters and the corresponding objective values are shown in Table 4. We calculate the relative error, i.e., (optimal expected value − actual expected value)/optimal expected value × 100%, where the optimal expected value is the maximum one of all the expected values obtained. It can be seen from Table 4 that the relative errors do not exceed 1%, indicating that the designed GA is robust to the set parameters and effective for solving the proposed problem.

### 7. Conclusions

In this paper, we discuss a portfolio selection problem with security returns given by experts’ evaluations instead of historical data. A factor method for evaluation of security returns based on experts’ evaluations is proposed and a mean-chance model for optimal portfolio selection is developed taking transaction costs and investors’ diversification and investment limitation requirements into account. The factor method of evaluation can make good use of experts’ knowledge on the effects of economic environment and the companies’ uniqueness on security returns and incorporate the contemporary relationship of security returns in the portfolio. The use of chance of portfolio return failing to reach the threshold can help investors easily tell their tolerance toward risk and thus facilitate a decision making. To solve the proposed nonlinear programming problem, a GA is presented. The results of the numerical example show that the proposed GA is effective in solving the problem.

### Acknowledgments

This work was supported by National Natural Science Foundation of China Grant No. 71171018 and Specialized Research Fund for the Doctoral Program of Higher Education Grant No. 20130006110001.

### Appendix. Uncertain variable

To better describe the uncertainty which is subject to human beings’ imprecise estimations, Liu (2007) proposed an uncertain measure and further developed uncertainty theory based on normality, duality, subadditivity and product axioms.

**Definition 1.** Let \( \Gamma \) be a nonempty set, and \( \mathcal{L} \) a \( \sigma \)-algebra over \( \Gamma \). Each element \( A \in \mathcal{L} \) is called an event. A set function \( \mathcal{M}(\{A\}) \) is called an uncertain measure if it satisfies the following three axioms (Liu, 2007):

(i) (Normality Axiom) \( \mathcal{M}(\Gamma) = 1 \).

(ii) (Duality Axiom) \( \mathcal{M}(\{A\}) + \mathcal{M}(\{\bar{A}\}) = 1 \).

(iii) (Subadditivity Axiom) For every countable sequence of events \( \{A_i\} \), we have

\[ \mathcal{M}\left( \bigcup_{i=1}^{\infty} A_i \right) \leq \sum_{i=1}^{\infty} \mathcal{M}(A_i). \]

The triplet \( (\Gamma, \mathcal{L}, \mathcal{M}) \) is called an uncertainty space.

It can be proven that any uncertain measure \( \mathcal{M} \) is increasing. That is, for any events \( A_1 \subseteq A_2 \), we have

\[ \mathcal{M}(A_1) \leq \mathcal{M}(A_2). \]

In order to define product uncertain measure, Liu (2009) proposed the fourth axiom as follows:

(iv) (Product Axiom) Let \( (\mathcal{L}_k, \mathcal{M}_k) \) be uncertainty spaces for \( k = 1, 2, \ldots \). The product uncertain measure \( \mathcal{M} \) is an uncertain measure satisfying

\[ \mathcal{M}\left( \prod_{k=1}^{\infty} A_k \right) = \min_{t \leq x < \infty} \mathcal{M}_k[A_k]. \]

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

\[ \{\xi \in B \} = \{y \in \Gamma | \xi(y) \in B\} \]

is an event.

An uncertainty distribution function is used to characterize an uncertain variable and is defined as follows.

**Definition 3 (Liu, 2007).** The uncertainty distribution \( \Phi : \mathbb{R} \rightarrow [0, 1] \) of an uncertain variable \( \xi \) is defined by

\[ \Phi(t) = \mathcal{M}[\xi \leq t]. \]

For example, by a normal uncertain variable, we mean the variable that has the following normal uncertainty distribution

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

where \( e \) and \( \sigma \) are real numbers and \( \sigma > 0 \). For convenience, it is denoted in the paper by \( \xi \sim \mathcal{N}(e, \sigma) \).
Definition 4 (Liu, 2009). The uncertain variables $\xi_1, \xi_2, \ldots, \xi_n$ are said to be independent if

$$\mathcal{M}\left(\bigcap_{i=1}^{n} (\xi_i \in B_i) \right) = \min_{1 \leq i \leq n} \mathcal{M}(\xi_i \in B_i)$$

for any Borel sets $B_1, B_2, \ldots, B_n$ of real numbers.

When the uncertain variables are represented by uncertainty distributions, the operational law is given by Liu (2010) as follows:

Theorem 3 (Liu, 2010). Let $\xi_1, \xi_2, \ldots, \xi_n, \xi_{n+1}, \ldots, \xi_{n+m}$ be independent uncertain variables with uncertainty distributions $\Phi_1, \Phi_2, \ldots, \Phi_n, \Phi_{n+1}, \ldots, \Phi_{n+m}$, respectively. Let $f(t_1, t_2, \ldots, t_n)$ be strictly increasing with respect to $t_1, t_2, \ldots, t_n$, and $g(t_1, t_2, \ldots, t_n, t_{n+1}, \ldots, t_{n+m})$ be strictly increasing with respect to $t_1, t_2, \ldots, t_n$ and strictly decreasing with respect to $t_{n+1}, t_{n+2}, \ldots, t_{n+m}$. Then

$$\xi = f(\xi_1, \xi_2, \ldots, \xi_n)$$

is an uncertain variable with inverse uncertainty distribution function

$$\psi^{-1}(\alpha) = f(\Phi_1^{-1}(\alpha), \Phi_2^{-1}(\alpha), \ldots, \Phi_n^{-1}(\alpha)), \quad 0 < \alpha < 1.$$  \hspace{1cm} (11)

and

$$\eta = g(\xi_1, \xi_2, \ldots, \xi_n, \xi_{n+1}, \ldots, \xi_{n+m})$$

is an uncertain variable with inverse uncertainty distribution function

$$\psi^{-1}(\alpha) = g(\Phi_1^{-1}(\alpha), \ldots, \Phi_n^{-1}(\alpha), \Phi_{n+1}^{-1}(1-\alpha), \ldots, \Phi_{n+m}^{-1}(1-\alpha)), \quad 0 < \alpha < 1.$$  \hspace{1cm} (12)

To tell the size of an uncertain variable, Liu defined the expected value of uncertain variables.

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

$$\mathbb{E}[\xi] = \int_{-\infty}^{\infty} \mathcal{M}(\xi \geq r)dr - \int_{-\infty}^{0} \mathcal{M}(\xi \leq r)dr$$

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

When the uncertainty distribution function $\Phi$ or the inverse uncertainty distribution function $\Phi^{-1}$ of the uncertain variable $\xi$ is known, it is easy to verify that the expected value of $\xi$ can also be obtained via (Liu, 2007)

$$\mathbb{E}[\xi] = \int_{0}^{\infty} (1 - \Phi(r))dr - \int_{-\infty}^{0} \Phi(r)dr,$$  \hspace{1cm} (13)

$$\mathbb{E}[\xi] = \int_{0}^{1} \Phi^{-1}(\alpha)d\alpha.$$

Theorem 4 (Liu, 2010). Let $\xi_1$ and $\xi_2$ be independent uncertain variables with finite expected values. Then for any real numbers $a_1$ and $a_2$, we have

$$\mathbb{E}[a_1 \xi_1 + a_2 \xi_2] = a_1 \mathbb{E}[\xi_1] + a_2 \mathbb{E}[\xi_2].$$  \hspace{1cm} (14)

References


