A Fuzzy Behavioral Portfolio Model Based on Text Sentiment Analysis

Qiansheng Zhang¹, Tiancheng Hou², Hongyi Jiang³, Suifu Gan⁴, Lingyue Zhang⁵

¹,²,³ School of Finance, Guangdong University of Foreign Studies, Guangzhou 510006
⁴ School of Informatics, Guangdong University of Foreign Studies, Guangzhou 510006

Abstract—A Fuzzy Behavioral Portfolio model (FBPM) is proposed for security investment with insufficient market information and uncertain emotion influence on investment return and risk. Based on the general behavioral portfolio theory, trapezoidal fuzzy number is employed to characterize investment return and risk. Text emotion analysis based on emotional lexicons is introduced to obtain the market investment emotion and then adjust the investment value function to obtain the optimal portfolio solution of security. In the end, the back testing on Shanghai Composite Index, Shenzhen Component Index and Growth Enterprise Market Index implies the advantage of FBPM on both return and risk.

Keywords—Text Emotion Analysis; Behavior Investment of Security; Fuzzy Portfolio Decision; Fuzzy Return; Fuzzy Risk

I. INTRODUCTION

The key of a portfolio theory is the characterization of investor behavior, which includes choosing the method of measuring investment returns and risks as well as establishing the procedure and objectives of decisions. Previously, relevant studies mostly characterize returns and risks with random variables and their statistics. One of the renowned theories, the Modern Portfolio Theory of Markowitz (1952) [1], uses mean and variance values to respectively characterize investment returns and risks. With the restriction of given risk levels, it maximizes the expectations of returns to obtain an optimal securities portfolio. However, there are several limitations in these theories.

The first one is that it is improper to model security market with only random variables because the psychological activities of investors cannot be sufficiently characterized by randomness. Furthermore, compared with quantitative statistics, qualitative fuzzy languages are preferred in describing the price states of the security market. In addition, it’s difficult for investors to determine the precise distributions of the returns of securities. Generally, it is impractical to obtain the precise probability distribution, but relatively easy to estimate the degree of membership function and fuzzy possibility distribution. So fuzzy return rate will be used in our model.

Secondly, “rational investors” and “maximized expectation of utility” theories have been challenged intensively by extreme situations of market, psychological experiments and other phenomena contradicting to the Expected utility hypothesis. An improved version of the
Markowitz theory is Behavioral Portfolio Theory (BPT) (Shefrin & Statman, 2000) [2], which is based on Prospect Theory (Kahneman & Tversky, 1979)[3]. By considering irrational behavior bias, multiple psychological accounts and loss aversion, it characterizes the behavior of irrational investors. We will propose an improved behavioral portfolio model based on it, which takes the mentioned irrational behavior into account.

At last, in traditional theories, decisions among investors are isolated, which does not hold in real world where investors interact with each other frequently. The Flock Effect in the security market indicates that investors are likely to follow the tendencies of market (Banerjee, 1992) [4]. In this article, we will analyze the sentiment orientations of texts of relevant comments toward investment on Sina Weibo to obtain proper estimations of the sentimental states of the market, and eventually build up a portfolio model considering sentimental influences. For sentimental analysis, we will use the sentimental lexicon method developed by Jiang, Huang, Cai and Wang (2015) [5], which utilizes the ICTCLAS (Institute of Computing Technology, Chinese Lexical Analysis System) (Zhang, Yu, Xiong & Liu, 2003) [6] for segmentation, the National Taiwan University Semantic Dictionary (NTUSD) (Ku & Chen, 2007) [7] and HowNet (Dong Z., Dong Q. & Hao, 2006) [8] for sentimental classification.

In all, based on the traditional BPT, this article constructs a new behavioral portfolio model by using trapezoidal fuzzy numbers to characterize the features of returns and risks. We use test sentiment analysis to obtain the sentimental parameters of the investment market and consequently, adjust the prospect value function of investors. As a result, an improved model considers the cognitive bias, vagueness of cognition and market sentiment is constructed to promote the efficiency of portfolio decisions. Moreover, by properly solving the parameters, it settles the problems that the traditional BPT cannot be validated with real-world data.

II. ANALYSIS ON THE CLASSICAL SINGLE PSYCHOLOGY ACCOUNT MODEL OF BPT

Given that there are \( m \) kinds of securities available for investment. Price state space of the security market is \( S = \{ s_1, s_2, ..., s_t \} \), \( p_j \) is the probability of the presence of \( s_j \).

Let \( r_{ij} \) be return rate of the \( i^{th} \) security under the state \( s_j \), \( w = (\omega_1, \omega_2, ..., \omega_m)\) be asset allocation vector, \( W_0 \) be initial wealth owned by investors and \( W_j \) be eventual wealth under the price state \( s_j \). We have

\[
W_j = W_0 \cdot \sum_{i=1}^{m} (1 + r_{ij})\omega_i \tag{1}
\]
Under the cumulative representation (Tversky & Kahneman, 1992) [9], the prospect value functions of investors toward the \(i\_{th}\) security under the price state \(s_j\) is

\[
v_{ij} = \begin{cases}  
(r_{ij} - R_j)\gamma, & r_{ij} > R_j \\
-\theta(R_j - r_{ij})\delta, & r_{ij} < R_j 
\end{cases}, \quad (i = 1, \ldots, m; \ j = 1, \ldots, t) \tag{2}
\]

where \(R_j\) represents reference return rate under the state \(s_j\), in other words, the return rate satisfying the investors under this state. \(\theta > 1\) illustrates the degree of loss aversion. \(\gamma, \delta\) illustrate respectively the convexities of profit and loss value function. As a result, comprehensive prospect values under all potential states is obtained

\[
V_i = \sum_{j=1}^{t} v_{ij} \pi_j(p_j) \tag{3}
\]

where \(\pi_j\) is the subjective probability of investors toward the state \(s_j\), which is a function of the objective probability \(p_j\).

According to the definitions, the classical single psychological account model of BPT is obtained

\[
\max V = \sum_{i=1}^{m} V_i \omega_i \tag{4}
\]

\[
\text{s.t.} \quad \left\{ \begin{array}{l}
\text{Prob}\{W \geq A\} \geq \alpha \\
\sum_{i=1}^{m} \omega_i \leq 1
\end{array} \right.
\]

where \(W\) is the random variable whose state space is \(\Omega = \{W_1, W_2, \ldots, W_t\}\), \(A\) is the expected wealth of investors and \(\alpha\) is the lowest allowed probability that \(W\) is greater than \(A\). The second condition is for budget constraint.

There are some points remained to be discussed:

(1) It is improper to use specific numbers to characterize return rates because the feature of return rates is of fuzziness. Especially when we introduce multiple market states in our model, it is difficult to directly assign specific values as the expected returns conditioning on various states.
(2) To determine the objective probabilities of future market states is difficult. To assign value directly to the probabilities seems to be inconvincible and arbitrary. When we introduce cognitive bias including Flock Effect in the model, it doesn’t hold that the subjective probabilities are based purely on objective probabilities, so we cannot regard the subjective probability as a function of the objective probability only.

(3) Because objective probability distributions of under various states are unknown, it is not feasible to introduce the probability restriction on the eventual wealth. This leads to the fact that there are no clear advantages compared with traditional risk restriction.

To solve these problems, considering the fuzziness of expectations of investors toward market, this article introduces trapezoidal fuzzy numbers to behavioral portfolio model. We use fuzzy expectations and fuzzy variances both proposed by Carlsson and Fullér (2001) [10] to characterize fuzzy returns and fuzzy risks of securities. We change the subjective probabilities and objective probabilities in the traditional model into the subjective possibilities distributions of market, which will be obtained in the chapter of text sentiment analysis. Our new model is constructed as following.

### III. CONSTRUCTION OF FUZZY BEHAVIORAL PORTFOLIO MODEL (F-BPM)

Because of incompleteness of market information, deficit of historical statistics of return rates and instability of sentiments, investors cannot approach accurate values of returns and risks of securities. So, we can introduce fuzziness and depict returns by trapezoidal fuzzy numbers defined as below:

Definition 1. A trapezoidal fuzzy number $\tilde{r}_{ij} = (a_{ij}^{ij}, b_{ij}^{ij}, \alpha_{ij}^{ij}, \beta_{ij}^{ij})$ represents the fuzzy return of the $i_{th}$ security under state $s_j$. Its membership function $\tilde{r}_{ij}(r)$ is defined as

$$\tilde{r}_{ij}(r) = \begin{cases} 
1 - \frac{a_{ij}^{ij} - r}{\alpha_{ij}^{ij}} & a_{ij}^{ij} - \alpha_{ij}^{ij} \leq r \leq a_{ij}^{ij} \\
0 & \text{Else}
\end{cases}$$

where $[a_{ij}^{ij}, b_{ij}^{ij}]$ is the tolerance interval of return rate $r$. The left width is $\alpha_{ij}^{ij}$ and the right is $\beta_{ij}^{ij}$. The $\lambda$ cut set is

$$[\tilde{r}_{ij}]^\lambda = [a_{ij}^{ij} - (1 - \lambda)\alpha_{ij}^{ij}, b_{ij}^{ij} - (1 - \lambda)\beta_{ij}^{ij}] = [a_{1j}^{ij}(\lambda), a_{2j}^{ij}(\lambda)]$$
Definition 2. The fuzzy co-variance of $\hat{r}_{ij}$ and $\hat{r}_{kj}$ of the $i_{th}$ and $k_{th}$ securities under state $s_j$ is

$$\tilde{\sigma}^2(\hat{r}_{ij}, \hat{r}_{kj}) = \frac{1}{2} \int_0^1 \lambda (a_{ij}^2(\lambda) - a_{ij}^1(\lambda))(a_{kj}^2(\lambda) - a_{kj}^1(\lambda))d\lambda$$

where the $\lambda$ cut sets of the $i_{th}$ and $k_{th}$ securities under state $s_j$ is

$$|\hat{r}_{ij}|^\lambda = [a_{ij}^1(\lambda), a_{ij}^2(\lambda)],$$

$$|\hat{r}_{kj}|^\lambda = [a_{kj}^1(\lambda), a_{kj}^2(\lambda)].$$

The difference between real return rate and reference return rate is $\Delta \hat{r}_{ij} = |\hat{r}_{ij} - R_j|$. Consequently, the fuzzy prospect value function of the $i_{th}$ security under state $s_j$ can be written as

$${\tilde{v}}_{ij} (\hat{r}_{ij}) = \begin{cases} (\Delta \hat{r}_{ij})^\gamma, & \hat{r}_{ij} > R_j \\ -\theta(\Delta \hat{r}_{ij})^\delta, & \hat{r}_{ij} < R_j \end{cases} #(5)$$

Let $p_j$ be the estimation of the possibility that the market is under state $s_j$. In other words, $p_j$ is the weight of market sentiment toward state $s_j$. $V_i$ is the comprehensive prospect value of the $i_{th}$ security and $V$ is the comprehensive prospect value of all securities.

$$V_i = \sum_{k=1}^{m} p_j E({\tilde{v}}_{ij})$$

$$= \sum_{j=1}^{t} p_j \int_0^1 \lambda [ {\tilde{v}}_{ij} (a_{ij}^1(\lambda)) + {\tilde{v}}_{ij} (a_{ij}^2(\lambda))]d\lambda$$

$$V = \sum_{i=1}^{m} V_i \omega_i #(6)$$

And the comprehensive risk of portfolio is
\[
\sigma^2 = \sum_{k=1}^{m} \sum_{i=1}^{m} \sum_{j=1}^{t} p_j \omega_i \omega_k \tilde{\sigma}^2 \left( \tilde{r}_{ij}, \tilde{r}_{kj} \right)
\]

\[
= \sum_{k=1}^{m} \sum_{i=1}^{m} \sum_{j=1}^{t} p_j \omega_i \omega_k \frac{1}{2}
\]

\[
\int_{0}^{1} \lambda \left[ a_{ij}^2 (\lambda) - a_{ii}^j (\lambda) \right] \left[ a_{kj}^i (\lambda) - a_{kj}^i (\lambda) \right] d\lambda \quad (7)
\]

As a result, we can construct a Fuzzy Behavioral Portfolio Model (F-BPM), which is

\[
\max_w V \quad (8)
\]

s. t. \[
\begin{align*}
\sigma^2 &= M \\
\mathbf{w}^T \mathbf{F} &= 1
\end{align*}
\]

where M is the maximum of the tolerable risk, and F is a unit vector.

F-BPM is a quadratic programming problem, and its dual program can be solved with convex quadratic programming algorithms. Moreover, as there is short-sale constraint in Chinese market, we add it into F-BPM to keep consistency with the market. Thus, the final model is

\[
\min \frac{1}{2} \sigma^2 \quad (8')
\]

s. t. \[
\begin{align*}
V &= \mu \\
\mathbf{w}^T \mathbf{F} &= 1 \\
\mathbf{w} &\geq 0
\end{align*}
\]

where \( \sigma^2 \) is the comprehensive risk, V being the comprehensive prospect value of all securities, \( \mu \) being the required minimum V and \( \mathbf{w} \) is the allocation vector and \( \mathbf{F} \) is a unit column vector.

Due to the existence of short-sale constraint, the efficient portfolio frontier is of a piecewise function. The analytic solution of this programming only can be achieved by reducing the original problem into a groups of new quadratic programming problems recursively. For simplicity, and since finding the solution numerically is stable and efficient, in this article, we will apply interior-point-convex algorithm of MATLAB to solve this problem.
IV. OBTAINING PARAMETERS

Some papers of portfolio theory ended after building a model or just giving a numerical example of their model by setting some random parameters. The calibration of each specific parameter may not be explained clearly, making those models difficult to be applied in the real world. To avoid that, this section gives a complete explanation on the way to obtain the parameters in our model by using historical market data.

A. Market state probability

In the model (8), the probability $p_j$ is a key parameter, which indicates the subjective estimation of the likelihood that the market will be in state $s_j$ in the next day. It represents the overall expectation of investors regarding the market return in the future.

To estimate this parameter, we use the method of sentiment analysis. We can extract the general attitude of investors toward market, by analyzing texts gathered from Chinese microblogging website Sina Weibo, where investors can make comments on the market condition.

Our sentiment analysis process mainly follows that from Jiang et al. (2015) [5], which consists of six steps:

1. Fetching text

   We select 50 stock market commentators who have the greatest number of followers in Sina Weibo. Then we use an automated program to fetch all their post during a specific period, including all comments of their post made by followers.

2. Data cleaning

   We calculate the proportion of finance-related words on each text gathered, and filter out all texts that do not contain these words or in which the proportions of these words are lower than 50% among content words.
(3) Tokenizing

The acquired texts are in Chinese, meaning that there is no space between each two words, making it hard to apply the technique of emotional lexicons. Therefore, we use the ICTCLAS to segment text into words. The system is reliable and renowned for its segmentation precision rate (97.58%, result from official evaluation in China national 973 project).

(4) Extracting emotion vector

Based on the emotional lexicons, we can extract an emotion vector from a text, which consists of tuples \((W_{ik}, C_{ik}, w_{ik})\) of the emotion word \(W_{ik}\), its surrounding context \(C_{ik}\) and the attitude of the word represented by a real number \(w_{ik}\) (for word \(i\) in text \(k\)). The number is positive if the emotion is positive and vice versa. The absolute value indicates emotional strength of the word.

(5) Handle negation and adverb

If negation words occur in the context \(C_{ik}\) for odd number of time, then negate \(w_{ik}\) as \(-w_{ik}\) in the tuple.

For each of the adverb, find the strength \(s\) of the expression from the emotional lexicons, then reassign \(w_{ik} = s * w_{ik}\).

As a result, the emotion \(Emo_k\) of each text can be expressed as a sum

\[
Emo_k = \sum_{i} w_{ik}
\]

for each word \(W_{ik}\).

(6) Calculate the market state probability \(p_j\)

For convenience, we only consider that the market has three states, i.e. positive (\(j=1\)), neutral (\(j=2\)), and negative (\(j=3\)), though extending the number of states is trivial.
By setting a threshold $\epsilon > 0$, we can classify any text into positive if $Emo_k > \epsilon$, neutral if $|Emo_k| < \epsilon$ or negative if $Emo_k < -\epsilon$. By first manually classify some text, the threshold can be chosen to approximate manual classification.

After that, count the number of texts belonging to each class. Then the market probabilities of future states may be estimated by the proportion of texts in each class. For example, the positive state probability $p_1 = \frac{\text{number of positive text}}{\text{number of text in total}}$.

B. Parameters for fuzzy rate of return $r_{ij} = (a_{ij}, b_{ij}, \alpha_{ij}, \beta_{ij})$

After knowing how to estimate $p_j(t)$ for each $t$, we can obtain the above four parameters from historical data.

Define the conditional frequency on market state $s_j$ for some specific return $r$, of the asset $i$, as:

$$f_{ij}(r) = \sum_{t=0, r_i = r}^N p_j(t)$$

Then the conditional cumulative distribution function of return on market state $s_j$ for asset $i$ can be defined as:

$$P_{ij}(r) = \int_{-\infty}^{r} f_{ij}(s)ds$$

From which we can calculate the percentile.

We set $a_{ij} - \alpha_{ij}$ be the 20th percentile, $a_{ij}$ as the 40th percentile, $b_{ij}$ as the 60th percentile and $b_{ij} + \beta_{ij}$ as the 80th percentile of the above distribution.

C. Benchmark returns

The benchmark returns $R_j$ mainly defines how investors normally expecte to earn under the market state $s_j$. Such expectation originated from what investors got in the past. Therefore, it can be expressed as the average return of the past under some market state:
$$R_j = r_{0j} = \frac{\sum_{t=0}^{N} p_j(t) r_0(t)}{\sum_{t=0}^{N} p_j(t)}$$

Namely, the average of a market index returns weighted by the probability of time $t$ to be under the market state $s_j$.

D. Determining Parameters $\theta$, $\gamma$, $\delta$ in value function.

These three parameters take the same value across many literature, originally given by Kahneman [9]. We also set them to the classic value: $\gamma = \delta = 0.88$, $\theta = 2.25$.

E. Expected return (Subjective)

Both our model and the traditional MPT model need users to set how much they want to earn, namely, the parameter $\mu$. This parameter should be set according to one’s risk preference. During the back-testing in the next section, we will set it be the value that minimizes the risk of the resulting portfolio, to reflect the profitability of the model under a conservative notion.

These parameters are all our model needed to produce an optimal portfolio. As an application, we will do a back-testing to verify the theory.

V. BACK-TESTING THE FUZZY-BPM MODEL

From the mutual fund separation theorem, all investment to the market may be separated into two parts: investment to riskless asset and investment to market portfolio. Hence, we will use several representative indexes to simulate all investment targets (stocks) available to investor. Since our sentimental data is from Chinese microblog website, Weibo, so the back-testing will also conduct on Chinese stock market and its indexes. We use three
indexes in our experiment: Shanghai Composite Index, Shenzhen Component Index, and Growth Enterprise Index. These three indexes cover most of the stocks available in Chinese stock market.

The simulated trades are made daily, based on historical daily closing price data. For each day, the historical data will be input into the parameter estimation procedure to update the parameters. Then the parameters are used to calculate the optimal portfolio by our F-BPM model and, for comparison, the traditional MPT model. Afterwards, the difference between the theoretical optimum and the current portfolio will be traded on current closing price, assuming there are no trading fees.

The test is conducted on price data collected from TDX terminal, dated from May 5th, 2016 to January 24th, 2017. The texts fetched for sentiment analysis from Sina Weibo is of the same period.

![Figure 1: Accumulative wealth and utility in the back-testing](image)

In the figure, line “F-BPM”, “Random”, “MPT” and “Index” denote the results from using our model, trading randomly, using Modern Portfolio theory and holding the index respectively. Index holding means to buy each index weighted according to its market value.
TABLE 1: STATISTICS ON RATE OF RETURN (ANNUALIZED)

<table>
<thead>
<tr>
<th></th>
<th>F-BPM</th>
<th>Random</th>
<th>MPT</th>
<th>Index</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rate of Return</td>
<td>10.47%</td>
<td>0.29%</td>
<td>7.36%</td>
<td>7.93%</td>
</tr>
<tr>
<td>Std</td>
<td>12.18%</td>
<td>13.54%</td>
<td>13.02%</td>
<td>12.33%</td>
</tr>
<tr>
<td>Sharpe Ratio</td>
<td>0.8596</td>
<td>0.0214</td>
<td>0.5653</td>
<td>0.6431</td>
</tr>
</tbody>
</table>

For each trading strategy, the mean, standard deviation and Sharpe ratio of daily return are shown in the table.

From the figure and table, we can know clearly that after a period, our model surpasses others in many ways. Specifically, the return rate of our model is substantially higher than trading randomly or just following the market by buying indexes, showing the its effectiveness. Comparing to the traditional MPT model, F-BPM earn 3.11% more annually, which is a big improvement by 42.26%. Meanwhile, the risk of earning such profit than MPT decreased 0.84%. Thus, the Sharp ration of our model is significantly higher than that of MPT.

Apart from what these statistics tell, since F-BPM takes loss aversion into account, by predicting the market state by sentiment analysis, losses are reduced when the market turn into downward path in the later period. Especially in the figure of utility, comparing our model and MPT, the performance difference within downward path is clear.

These results show that the Fuzzy Behavioral Portfolio Model (F-BPM) can yield a better optimal portfolio under Chinese market environment.

VI. CONCLUSION

The Fuzzy Behavioral Portfolio Model is supposed to solve some problems of traditional portfolio theory, arising from ignoring the psychological and socioeconomic factors. Also, our model is supposed to address the difficulty of reflecting subjective expectation and calculating probability constraints. By setting the fuzzy return and fuzzy variance as the
objective function and constraint of optimization, we have the classical Behavioral Portfolio Theory (BPT) fuzzified. By using sentimental analysis, we made possible the incorporation of market emotion into a portfolio optimization model. Finally, by back-testing on the historical Chinese market data, we verified the effectiveness and advantages of the F-BPM model.

Of course, there are still some problems that need to be solved in future study.

As the number of active user on the Weibo platform was gradually decreasing in the past few years, it is doubtful that the attitude of commentators on Weibo can represent the so-called “market attitude”. Also, the method of sentimental lexicons is efficient in analyzing large amount of texts, but not accurate enough when encountered texts with multi-dimensional attitude. These problems may reduce the accuracy in estimating market state probability, which may in turn damage the whole optimization process. Acknowledging this, we are building a more powerful text emotion classifier from a bigger data, in hope to address these drawbacks in the future research.

ACKNOWLEDGMENT

This paper is supported by Innovative School Project in Higher Education of Guangdong, China under grant GWTP-GC-2014-03, and the Student Innovation and Entrepreneurship Training Program of Guangdong under grant 201611846035, 110-GK161017.

REFERENCES


