ISSN: 2688-8645

An Empirical Study on the Influence of Investor Sentiment on Fund Excess Return from the Perspective of Network Information

Jingcai Wang[1], Zejiong Zhou[2, *]

[1]School of Finance, Anhui University of Finance and Economics, Bengbu, Anhui, China

[2]School of Economics, Anhui University of Finance and Economics, Bengbu, Anhui, China

*aczzj123456@163.com

Abstract

The stock selection ability and timing ability of fund managers determine the excess return of the fund, and the influence of investor sentiment on investor behavior is bound to affect the excess return of the fund through the fluctuation of asset price and market transaction. Firstly, based on the stock bar comment data, the deep learning model is trained to classify the comments. The results show that the overall sentiment of online investors presents negative characteristics. Secondly, the investor sentiment index is constructed according to the classification results to quantify investor sentiment. The results show that the trend of Shanghai stock index is negatively correlated with the investor sentiment index extracted from the 10 day trend item. Finally, based on panel data, this paper empirically studies the impact of investor sentiment on the excess return of China's open-end stock funds. The results show that there is a negative correlation between investor sentiment based on online text information and the excess return of open-end stock funds when considering the cross-sectional fixed effect of panel data. After classifying the funds by investment type, there is a negative correlation between investor sentiment based on network text information and the excess return of positive growth, steady growth, index and optimized index funds. The negative correlation between investor sentiment and the excess return of index funds is the most obvious, followed by optimized index and positive growth funds, and the least obvious is stable growth funds.

Keywords

Investor sentiment; Fund; Excess return; Network information.

1. Introduction

In 2020, the sudden COVID-19 made the world economy suffer heavy losses. In the background of the market generally not optimistic about the future, investor sentiment continued to be depressed. China's Shanghai stock index was 7.72% lower than the first trading day after the Spring Festival. The S & P 500 index continued to melt in March 9th, March 12th and March 16th. It is evident that investor sentiment has affected the price of financial assets. As Baker and Wurgler said, the focus of the problem is no longer whether investor sentiment will affect stock prices, but how to better measure investor sentiment and quantify its impact [1].

China's institutional investors represented by funds have significant irrational behavior and are vulnerable to investor sentiment, forming market anomalies such as herding. At the same time, as a professional institutional investor in China's capital market, the investment decision of the fund will have a far-reaching impact on the investment target, the financial market and even the national economy. Studying the impact of investor sentiment on fund investment return and investment behavior will help us better identify the professional investment ability of the fund and recognize the impact of investor sentiment on different investors [2]. Therefore, the research on the relationship between investor sentiment and fund excess return is of great significance.

Starting with the open-end stock fund, this paper empirically studies the relationship between investor sentiment and fund excess return. In order to quantify investor sentiment more truly and accurately, we choose to use network information as the data source of investor sentiment measurement, apply deep learning technology to text sentiment analysis of network information, construct investor sentiment index according to the results of text sentiment analysis, and empirically study the impact of investor sentiment on the excess return of China's securities investment funds, In order to identify and quantify the investor sentiment contained in the data more efficiently in the face of network information big data, and provide new ideas for the research on the impact of investor behavior and investor sentiment on asset pricing.

2. Literature Review

To study the impact of investor sentiment on fund excess return, we need to accurately measure investor sentiment to explore its impact on securities investment funds.

2.1. Investor Sentiment Measurement

The academic research on the quantification of investor sentiment has been very mature. Scholars' research on investor sentiment index has experienced the evolution process from "subjective or objective single index measurement" to "composite index measurement considering subjective and objective factors comprehensively" [3].

Subjective or direct indicators are mainly investor intelligence index, CCTV market index, investor confidence index, consumer confidence index, etc. Representative research results include: Fisher and Statman (2000) found that investor intelligence index can be used as a measure of investor sentiment [4]. Marianoh and encina (2021) empirically analyzed the relationship between Bloomberg investor sentiment index and abnormal returns and volume shocks of major European financial companies, and found that the impact interpretation ability of Bloomberg investor sentiment index is low [5]. Xue Fei (2005) found that consumer confidence index can well measure investor sentiment. When it is good, small market stocks perform better [6].

The objective or indirect indicators use the public statistical data of market transactions as an alternative indicator of investor sentiment, mainly including closed-end fund discount, IPO issuance, first day income, trading volume, etc. Representative research results include: De Long et al. (1990) believe that the discount rate of closed-end funds can reflect the change of noise traders' mood, and investor sentiment is the systematic risk factor of asset pricing [7]. Jones (2002) found that high trading volume is often accompanied by low return, which can be used as a reverse predictor of stock return [8]. Long et al. (2020) studied whether the volatility index can represent investor sentiment, and the results show that the volatility index can reflect investor sentiment [9].

Composite indicators are composed of one or more direct and indirect indicators. Many scholars believe that the discount of closed-end funds, IPO issuance and first day income, net redemption of mutual funds, trading volume, zero share trading ratio and fund position ratio can better reflect the emotional changes of investors [4]. Representative research results include: Baker and Wurgler (2006) selected six indirect indicators of closed-end fund discount, trading volume, IPO number and first day income, dividend income and stock / securities issuance ratio to construct a composite index to measure investor sentiment. It is found that this index can more comprehensively reflect the psychological changes of investors than a single index [10]. Hu et al. (2021) selected the number of new accounts opened for a shares, the number of IPOs, the return on the first day of IPO, the discount of closed-end funds and the

ISSN: 2688-8645

turnover rate as emotional proxy variables, constructed a comprehensive index of investor sentiment through principal component analysis, and studied the dynamic relationship between investor sentiment, stock market return and volatility. It was found that the fluctuation of investor sentiment had a significant impact on stock market returns [11]. Xie Shiqing and Tang Sixun (2021) chose the number of newly established funds, turnover rate, P/E ratio, margin trading balance, trading volume and investor confidence index as proxy variables of investor sentiment, and explored investor sentiment and macroeconomic fluctuations through structured vector autoregressive model. The results show that, Positive investor sentiment has a positive impact on the stock market return in the short term. The explanation contribution rate of investor sentiment to the variance of stock market return is much higher than that of macroeconomic fluctuation [12].

In addition to the above three types of indicators, the advent of the era of big data makes it a new trend for academia to measure investor sentiment by using computer technology and massive online text data.

Representative research results include: Gao et al. (2020) constructed the emotion index of 38 countries from 2004 to 2014 according to the Google search behavior of families, and found that the emotion index is a reverse predictor of market return [13]. Wang Haoqing (2019) established a mixed model of investor emotion recognition based on long and short memory model and naive Bayesian model to classify the emotional tendency of stock bar comments and generate an emotion index. Then he verified that the emotion index has a certain prediction ability for the Shanghai Composite Index [14]. Chen Xiaojuan (2020) constructed the investor sentiment tracking index and investor sentiment consistency index with the interactive trading platform of Shenzhen Stock Exchange and the Oriental Fortune stock bar platform as data sources, and found that the investor sentiment tracking index and other investor sentiment indexes have a positive impact on the stock turnover rate [15].

2.2. Investor Sentiment and Securities Investment Funds

Studies have shown that the change of stock returns will further affect institutional investors, especially securities investment funds. Beaumont et al. (2008) found that investors have a significant impact on the volatility of fund returns through the study of fund market data in the United States from 1998 to 2004. The higher investor sentiment is, the greater the volatility of returns is [16]. Hudson, Yan and Zhang (2020) examined the role of investor sentiment in the herding effect of institutional investors and found that there is herding effect in the investment behavior of UK fund managers, and herding effect has a one-way investor sentiment effect [17]. Yi Li and Li Shimei (2020) used the error detection rate method to test whether fund managers can choose the investment time according to the investor's emotion from four aspects: influencing factors, investment style, fund characteristics and sustainability. The results show that a few fund managers adopt the negative emotion timing strategy, and a certain proportion of fund managers adopt the positive emotion timing strategy, most fund managers do not adjust portfolio risk exposure according to changes in investor sentiment [18]. Hou Hui et al. (2021) used the flow data of open-end funds to establish an emotional contagion index to investigate the impact of emotional contagion on the fluctuation of heavy stocks and the synchronization of stock prices in the process of fund investment. The study found that emotional contagion can promote the fluctuation of heavy stocks [19].

To sum up, most of the academic circles still use composite indicators to measure investor sentiment, and use this to study the impact of investor sentiment on the excess return of financial assets such as stocks and securities investment funds. Although the investor sentiment measured by these indicators is getting better and better in objectivity, accuracy and authenticity, it still does not use the real statements made by investors themselves as the data source, and there is still room for improvement in the objectivity, accuracy and authenticity of measurement. In recent years, there has been a trend in academia to use big data technology, deep learning technology and natural language processing technology to process investor speech information to measure investor sentiment. However, most of the research results on the impact of investor sentiment on fund excess return still remain at the level of using subjective, objective or composite indicators to measure investor sentiment. There are still few academic achievements based on investors' speech information to study the impact of investor sentiment on fund excess return, and the only literature uses data with a large time span such as week and quarter as the data processing object. This paper aims to connect big data, computers and finance, use big data technology to collect massive text comments from investors on the Internet, and use deep learning technology to identify the emotions contained in the massive text information published by investors, so as to construct the investor sentiment index, Using the daily excess return data from 647 open-end stock funds and the investor sentiment index to construct panel data regression, in order to study its ability to explain the excess return of open-end stock funds, and provide new ideas for the research on the relationship between investor sentiment and excess return of securities investment funds.

3. Theoretical Basis

This paper empirically studies the impact of investor sentiment based on network information on fund excess return, and accurately measuring investor sentiment and fund excess return is the premise of our empirical research. In order to accurately measure the excess return of the fund, the first half of this section first introduces how to use CAPM model to calculate the excess return of the fund; In order to accurately measure investor sentiment based on Web text information, it is necessary to use deep learning model to assist text sentiment analysis.

3.1. CAPM Model and Fund Excess Return

There are many calculation methods of fund excess return, which are mainly composed of different basic models and benchmark combinations. The basic models mainly include: standard market model, T-M model, H-M model and asymmetric response model; Benchmark combinations include: single factor combination, three factor combination and four factor combination [20]. The CAPM model used in this paper is the standard market single factor model. The formula of using CAPM model to measure fund excess return is as follows:

$$r_i - r_f = \alpha_i + \beta_i (r_m - r_f) \tag{1}$$

Among them, α_i is the measurement index of the fund's excess return, r_f is the risk-free interest rate, r_m is the market portfolio yield, and β_i is the beta value of the fund portfolio. If $\alpha_i > 0$, the fund has achieved excess return. If $\alpha_i < 0$, the fund not only did not achieve excess returns, but performed worse than the market portfolio. If $\alpha_i = 0$, the fund did not achieve excess returns.

3.2. Text Emotion Analysis, Deep Learning and Investor Sentiment Measurement

The text information published by investors in the stock bar contains rich emotions. In order to extract emotional features from the text information and accurately measure investors' emotions, this part introduces the theoretical basis and practical steps of using deep learning model to assist emotion classification for text emotion analysis.

Text emotion analysis refers to the process of analyzing, processing, summarizing and reasoning the text containing subjective emotion by using natural language processing and computational linguistics [21]. At present, text emotion analysis technology is divided into rule-based methods and machine learning methods.

Because this paper uses the text information from stock bar and has a huge amount of data, it is suitable to use the deep learning model in machine learning to assist text emotion analysis. The practical steps are as follows: the first step is to use a crawler to crawl the text data from the stock bar forum. The second step is to clean the text corpus, segment words and remove stop words. The third step is to train the neural network model and solve the neural network parameters by using some text corpus marked with emotional tendency as input. In the fourth step, the trained neural network is used to analyze the emotional tendency of the remaining text corpus and output it. The fifth step is to measure investor sentiment by using the results of emotional tendency analysis.

4. Empirical Analysis

The empirical analysis steps of this paper are carried out in the following order: the first step is to crawl the text information of stock evaluation and collect the data of fund return and control variables. The second step is to preprocess the stock evaluation text information into computer recognizable information. The third step is to mark part of the stock evaluation text as the training corpus to train the deep learning model, evaluate the deep learning model and debug the parameters. The fourth step is to use the trained in-depth learning model to analyze the text emotion of all stock evaluation text information, and construct the investor emotion index according to the emotion classification results to quantify the investor emotion. The fifth step is to calculate the excess return of the fund according to the CAPM model and the fund return data. The sixth step is to construct the panel data regression between investor sentiment index and fund excess return, and empirically study the relationship between investor sentiment and fund excess return.

4.1. Data Acquisition

4.1.1. Text data acquisition

In order to measure investor sentiment as comprehensively and objectively as possible, this paper refers to the ranking of Alexa ranking website (www.alexa.com, may 2021), and finally selects Dongfang fortune.com stock forum as the text data source after screening. After selecting the source of text data, we compiled a special crawler program to crawl the title text data of stock reviews in the stock bar forum from January 1, 2020 to December 31, 2020 from the Shanghai index bar under the Oriental Wealth bar.

The Shanghai Stock Index (00000 1) directly reflects the overall trend of stocks listed on the Shanghai Stock Exchange, and is also the epitome of the overall trend of China's stock market. It has long become the reference basis for Chinese investors and securities practitioners to study and judge the trend of stock prices. And in 2020, China and the world economy experienced COVID-19's attack. The Shanghai stock index trend was not only a sharp decline, but also a sharp rise and shock market, which made investors' emotional characteristics very rich and suitable for analysis.

After crawling the text data of stock evaluation, we first counted the time distribution of post volume in a day. According to the statistical analysis results, investors prefer to publish stock comments in the two time periods of 9-11 and 13-15, which is also the market trading time period during the trading day. In contrast, there are relatively few posts in the non trading time period.

Then, we counted the time distribution of posts in a week. According to the statistical analysis results, users' posts in a week are mainly concentrated on Monday to Friday. They are concentrated on trading days, and the number of posts on weekends is significantly reduced.

4.1.2. Fund yield data acquisition

As mentioned above, the trend characteristics of Shanghai stock index in 2020 are very rich, so each trading day from January 1, 2020 to December 31, 2020 is selected as the sample research range. Select open-end stock funds as the research object, excluding QDII, QFII funds and index funds. After deleting some samples with missing data, this paper finally obtains effective samples from 647 funds, including the fund's daily return, risk-free interest rate, market premium factor $(r_m - r_f)$ in CAPM model, β_i value to measure the fund's risk, etc. the data are all from the RESSET database.

4.1.3. Control variable selection

Learn from Wang Jue and Chen Yongshuai's (2019) research on fund excess return [2], and select the following control variables: fund return volatility (RISK), using the standard deviation of quarterly return to take the natural logarithm; The age of the Fund (AGE), taking the natural logarithm of the month from the establishment of the fund to December 2020; Fund size(SIZE): adopt the natural logarithm of the net asset value of the fund; Fund flow(FLOW). Among them, the fund capital flow is based on the definition of capital flow by Xiao Jun and Shi Jin (2011) [22]. Assuming that the new capital flows at the end of the period, it is obtained from the following formula (2):

$$FLOW_{it} = [TNA_{it} - TNA_{it-1} \times (1 + R_{it})] / TNA_{it-1} \tag{2}$$

Where, $FLOW_{it}$ represents the fund flow, TNA_{it} represents the net asset value of Fund i at the end of t, and R_{it} represents the return rate of fund i at the end of t. $FLOW_{it} > 0$ represents new capital inflow, $FLOW_{it} < 0$ represents capital outflow.

4.2. Emotional Classification of Stock Evaluation

The original text data can not be directly used for the training and debugging of the deep learning model. It can be used for the training of the deep learning model only after the text is preprocessed and transformed into the information that can be recognized by the computer.

4.2.1. Text preprocessing

After obtaining the stock bar comment data, the text data is preprocessed for subsequent research and analysis. The processing process will affect the accuracy of emotion analysis. It mainly includes data cleaning, word segmentation and de stop words. The processed text data can be used for in-depth learning.

(1) Data cleaning. Regular expression and manual filtering are used to clean the initial text data to remove the content that reduces the accuracy of emotion analysis.

(2) Participle. Use the open source THULAC word segmentation tool to segment the stock evaluation text, and cut the Chinese sentence into separate idioms for input into the deep learning model.

(3) Go to stop words. The Stoplist is downloaded from the Internet, the words after word segmentation are compared and screened, and the functional words without actual meaning are eliminated. After completing the text preprocessing, a total of 808719 stock evaluation text corpora for in-depth learning are obtained.

4.2.2. Deep learning model training

The deep learning model regards the text as a sequence direct input model, and uses the internal neural network to realize the automatic learning and extraction of text features, which alleviates the dependence of the traditional machine learning model on feature engineering construction. With the development of deep learning, the neural network model using deep learning has been widely used in emotion classification tasks [23].

(1) Word embedding. This paper uses Word2Vec word embedding technology, uses Python's gensim toolkit, trains Word2Vec model with Chinese corpus after text preprocessing, and obtains word vector based on Wiki Chinese corpus, so as to achieve the purpose of transfer learning. In the subsequent deep learning model training, Word2Vec word vector is used as the first layer input of neural network.

(2) Classifier training. After getting the word embedding vector, we can build a deep learning model, train and debug the model. The first step is to segment the wiki Chinese corpus, and then serialize the Chinese corpus using Keras's Tokenizer class under Tensorflow toolkit. The second step is to construct a matrix containing the word vectors corresponding to all words in Wiki Chinese corpus. The third step is to judge 4067 sentences with positive emotional tendency and 4067 sentences with negative emotional tendency from the stock evaluation text data that has completed text preprocessing, and extract them. The fourth step is to segment the extracted sentence, and then convert the word after word segmentation into the word vector corresponding to the word to obtain the initial weight of Keras embedding layer. In the fifth step, after getting the Embedding layer, connect the LSTM as the neural network constructed in the deep learning model. Step 6: model training.

(3) Model evaluation and debugging. The deep learning model is developed using Keras toolkit. Tensorflow is used as the back end during training, and GPU acceleration is enabled. In each iteration, 70% of the labeled corpus is randomly selected for model training, and the remaining 30% is used for verification. The training results are shown in Figure 1 and figure 2 below.

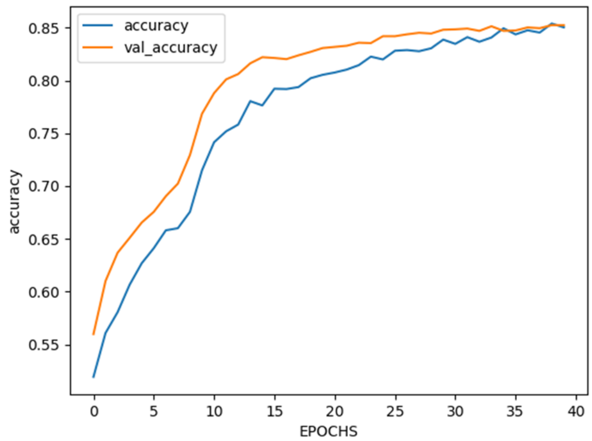

Figure 1. Model training accuracy

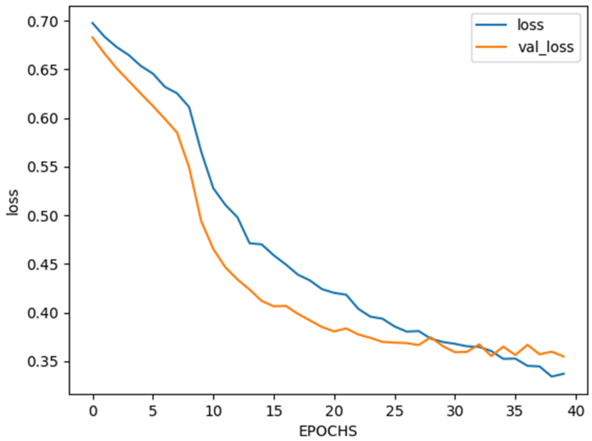

Figure 2. Model training loss

In Figures 1 and 2, EPOCHS represents the number of iterations of the model, accuracy represents the accuracy and val_accuracy represents the accuracy of the model on the validation set, loss represents the loss of the model in the training set, and loss represents the loss of the model on the validation set. It can be seen that after the first 15 iterative training, the accuracy of model prediction has been greatly improved in both training set and verification set, and the loss has been greatly reduced. After 40 iterations of training, the accuracy of the model has reached 85.51% on the training set, and the loss has been reduced to 0.3323. The accuracy of the test set is 85.24%, the loss is reduced to 0.3549, and the model tends to be stable. It can be seen that the deep learning model performs well in emotion classification and can be used for emotion classification of stock evaluation text.

4.2.3. Emotion classification

After completing the stock evaluation data preprocessing and deep learning model training, you can use the trained deep learning model to classify the 808719 stock evaluation texts. The classification results are optimistic or pessimistic, which can also be understood as bullish and bearish.

In the above, we have completed the training of deep learning model, so in this part, we only need to convert the preprocessed stock evaluation text into the corresponding word vector sequence according to the previously constructed dictionary and word vector matrix, and then use the trained model to predict, so as to classify the stock evaluation text. It can be seen from the classification results, that the accuracy of emotion classification results of stock evaluation texts is high. After the classification of all stock evaluation texts is completed, we will save the classification results in Excel files. These classification results are the data source for constructing investor sentiment indicators.

4.3. Construction of Investor Sentiment Index

After completing the sentiment classification of stock evaluation text, we counted the classification results, constructed the bullish index according to the method of Antweiler and Frank [24], and took the bullish index as an alternative to the investor sentiment index. According to the research of Antweiler and Frank, the results obtained by the current method of constructing bullish indicators are very similar. They investigated different aggregation functions with homogeneity. The specific construction methods are as follows:

Assuming that X^{up} and X^{down} represent the number of "bullish" and "bearish" stock reviews in a day, the bullish index BI is:

$$BI = \frac{X^{up} - X^{down}}{X^{up} + X^{down}} \tag{3}$$

According to the construction formula of bullish index, the larger the BI value, the more bullish posts on the day, and the higher the bullish mood. The smaller the BI value is, the more bearish posts is and the higher the bearish mood is. However, this form of bullish index does not take into account the impact of the number of posts on the bullish index. Generally speaking, the more posts you post, the stronger your mood. Therefore, according to the method of Zhuang Shaobo [24], the second form of bullish index is introduced

$$BI^* = \ln\left(\frac{1 + X^{up}}{1 + X^{down}}\right) \tag{4}$$

The bullish index constructed according to formula (4) considers the impact of the number of posts on the index on the basis of formula (3), and we take BI* as an alternative to the investor sentiment index.

After completing the construction of investor sentiment index, in order to have an intuitive understanding of investor sentiment index, we have made the sequence diagram of investor sentiment index. As can be seen from the sequence diagram, the investor sentiment index on all trading days is less than 0, and the maximum value is only -0.0793, which shows that the investor sentiment from Dongfang fortune Shanghai stock index bar is generally bearish on the Shanghai stock index and generally pessimistic. Then, we extracted the trend item of investor sentiment index by using the 10 day moving average and compared it with the trend of Shanghai stock index in the same period.

We calculated the correlation between the investor sentiment index and Shanghai Stock Exchange Index, and the result shows that the correlation between the two is -0.7215, indicating that there is a negative correlation between the rise and fall of the index and the change of investor sentiment.

4.4. Investor Sentiment Index and Fund Excess Return

Before studying the relationship between investor sentiment index and fund excess return, it is necessary to calculate the excess return of each fund. According to formula (1), first, we subtract the daily risk-free rate of return r_f from the rate of return r_i of 647 fund samples on each trading day. Secondly, we multiply the β_i of each fund relative to the CSI 300 index by the market premium factor $(r_m - r_f)$. Finally, we subtract the two and get the excess return α_i of the 647 funds on each trading day in the sample observation period.

After obtaining the excess return of each sample fund, we first consider the overall explanatory ability of the investor sentiment index to the excess return of the fund, and take the investor sentiment index BI* obtained above as the explanatory variable to determine the excess return of each fund αit as the explained variable, it introduces the standard deviation of return (RISK), fund age (AGE), fund size (SIZE) and fund flow (FLOW) as the control variables. Based on the panel data, OLS method is used to construct the mixed estimation model regression, as shown in the following formula (5).

$$\alpha_{it} = \beta_0 + \beta_1 \cdot BI^* + \beta_2 \cdot RISK_i + \beta_3 \cdot AGE_i + \beta_4 \cdot SIZE_i + \beta_5 \cdot FLOW_i + \varepsilon_{it} \tag{5}$$

Where β_0 is a constant and β_1 to β_5 are explanatory variable coefficients, where $i = 1, 2 \cdots N$ represents different funds and $t = 1, 2 \cdots T$ represents different periods. The regression results based on the panel data of 647 funds are shown in Table 2 below.

Table 2. Results of regression analysis of mixed estimation model

Variables	Results of regression analysis
BI*	-0.0701*** (-9.0660)
RISK	-0.1078*** (-8.0420)
AGE	0.1110*** (16.2027)
SIZE	0.0385*** (15.4975)
FLOW	0.0005*** (3.4953)
β0	-1.4127*** (-26.8400)

It can be seen from table 2 that without considering the possible fixed and random effects of panel data, the regression coefficient of investor sentiment index BI* is -0.0701, which is significant at the level of 1%, indicating that the higher the level of investor sentiment, the lower the excess return of the fund; In addition to the negative regression coefficient of the standard deviation of return, the regression coefficients of fund age, fund size and fund flow are positive, indicating that the higher the fluctuation level of fund return, the longer the duration of the fund, the stronger the fund assets and the more fund capital inflows, the higher the excess return of the fund.

Considering the possible fixed effect and random effect of panel data, we use Hausman test to test whether the panel data used in this paper has fixed effect or random effect. The test results are shown in Table 3 below.

Table 3. Hausman test results

Effect test	Chi-square statistic	P value
Cross section random effect	142.4964	<0.0001

It can be seen from table 3 that the statistics of cross-sectional random effect are significant at the level of 1%, which rejects the original assumption that there is cross-sectional random effect in panel data. In order to verify the fixed effect of panel data, we use F test for panel data, and the test results are shown in Table 4 below.

Table 4. F test results

Effect test	statistic	P value
Cross section F	6.4859	<0.0001
Cross section Chi-square	4152.0120	<0.0001

It can be seen from table 4 that the statistics of F test are significant at the level of 1%, rejecting the original assumption that there is no cross-sectional fixed effect in the panel data, indicating that there is cross-sectional fixed effect in the panel data, and there are heterogeneous differences between different funds. This difference can be attributed to the differences in fund management ability, which is certain for each fund.

We know that the panel data has cross-sectional fixed effect, so we use the panel least square method with dummy variables to construct the cross-sectional fixed effect regression model, as shown in the following equations (6) and (7).

$$\begin{aligned}\alpha_{it} = \beta_0 + k_1 D_1 \beta_0 + k_2 D_2 \beta_0 + \cdots + k_n D_n \beta_0 + \cdots + k_i D_i \beta_0 + \beta_1 \cdot BI^* \\ + \beta_2 \cdot RISK_i + \beta_3 \cdot AGE_i + \beta_4 \cdot SIZE_i + \beta_5 \cdot FLOW_i + \varepsilon_{it}\end{aligned} \tag{6}$$

$$D_n = \begin{cases} 1, n = i \\ 0, n \neq i \end{cases} \tag{7}$$

Among them, Dn is the dummy variable, kn is the coefficient before the dummy variable, and the other variables are the same as those in equation (5). The regression results are shown in Table 5 below

Table 5. Regression analysis results of cross section fixed effect model

Variables	Regression analysis results
BI*	-0.0867*** (-10.9712)
RISK	-0.2308*** (-13.7649)
AGE	0.0531* (-1.7701)
SIZE	-0.0434*** (-4.8815)
FLOW	0.0002 (1.2224)
β0	0.9619*** (5.1168)

It can be seen from table 5 that considering the cross-sectional fixed effect of panel data, the regression coefficient of investor sentiment index BI* is -0.0867, which is significant at the level of 1%, indicating that on the whole, there is a negative correlation between investor sentiment based on online text information and the excess return of open-end stock funds. Because the trend of investor sentiment based on network text information is negatively correlated with the trend of Shanghai stock index, while the trend of enterprise stock which is the subject of open-end stock fund investment is positively correlated with the trend of Shanghai stock index, the trend of investor sentiment is negatively correlated with the trend of enterprise stock invested by the fund. At the same time, the stock trend of these enterprises will directly determine the excess return of fund investment. Therefore, there is a negative correlation between the trend of investor sentiment and the excess return of the fund.

Different from table 4, the regression coefficient of the standard deviation of the rate of return risk is doubled, the regression coefficient of the fund age is doubled, the regression coefficient of the fund size becomes negative, and the regression coefficient of the fund flow is no longer significant at the significance level of 10%, indicating that considering the fixed effect of the cross-section of the panel data. The negative correlation between the fluctuation level of fund return and fund excess return is more prominent; The positive correlation between the duration of the fund and the excess return of the fund is weakened; The stronger the fund assets, the lower the excess return of the fund; The amount of fund inflow no longer has explanatory power to the excess return of the fund. The fluctuation of fund returns reflects the level of fund management. High-level managed funds can seek reasonable excess returns while keeping the difference between return fluctuations and market fluctuations small, while low-level managed funds will lose some excess returns due to the sharp fluctuation of fund returns; The duration of the fund also reflects the level of fund management. A good fund can survive for a long time because the ability of the fund to seek excess returns has been recognized by investors, while a fund with short duration cannot obtain high excess returns due to poor management level and blind pursuit of bright performance; With the expansion of fund scale, in order to ensure reputation, performance and business income and attract more investors to buy, the trading operation of the fund is more stable, which makes large-scale funds unable to obtain high excess returns, while small-scale funds can obtain more excess returns due to radical trading operation.

Considering the heterogeneity between the cross-sections of panel data, and in order to further explore the impact of investor sentiment on the excess returns of different types of open-end stock funds, we divide all 647 funds into four types according to the investment type: 138 positive growth funds, 224 stable growth funds, 151 index funds and 134 optimized index funds. After classifying the funds, we use the panel least square method with dummy variables to construct the cross-sectional fixed effect model regression for the panel data from different types of funds according to the fund type. The regression results are shown in Table 6 below.

Table 6. Results of regression analysis of positive growth funds

variable	Active growth fund	Steady growth fund	Index Fund	Optimizing index fund
BI*	-0.0773***	-0.0570***	-0.1250***	-0.1073***
	(-4.2270)	(-4.3495)	(-7.8726)	(-6.1446)
RISK	-0.2779***	-0.1900***	-0.2202***	-0.2925***
	(-6.6895)	(-6.4259)	(-6.8716)	(-8.4430)
AGE	-0.1846**	-0.1382**	0.0653	-0.1047*
	(-2.3244)	(-2.2873)	(1.1995)	(-1.8643)
SIZE	-0.0292	-0.0334**	-0.0004	-0.0471**
	(-1.5418)	(-2.2781)	(-0.0172)	(-2.1500)
FLOW	0.0243***	0.0001	0.0131	0.0099**
	(4.0873)	(1.1605)	(1.3671)	(2.1545)
β0	1.3513***	1.2104***	-0.4854	1.1237***
	(3.1616)	(3.6010)	(-1.1180)	(2.7550)

It can be seen from table 6 that investor sentiment has an impact on the excess returns of positive growth funds, steady growth funds, index funds and optimized index funds α The regression coefficients of it are negative and significant at the level of 1%, indicating that there is a negative correlation between investor sentiment based on online text information and excess returns of positive growth, steady growth, index and optimized index funds. Further observation shows that the negative correlation between investor sentiment index and excess return of index funds is the most obvious, with a coefficient of -0.1250. Optimized index funds and positive growth funds are the second, with coefficients of -0.1073 and -0.0773 respectively. The negative correlation between the investor sentiment index and the excess return of steady growth funds is the least obvious, with a coefficient of -0.0570, indicating that different types of funds have different sensitivity to investor sentiment: index funds need to ensure the accuracy and timeliness of tracking the corresponding index, the fund return is more closely related to the index trend, and the index trend is affected by investor sentiment, Therefore, the excess return of the fund is more affected by investor sentiment. Growth funds focus on value investment and do not track the constraints of the corresponding index. They invest more in high-quality companies that can bring rich returns in the future. The relationship between fund income and index trend is not as close as index funds. Therefore, fund income is relatively less affected by investor sentiment.

At the same time, we find that the regression coefficients of fund age, fund size and fund flow of index funds are not significant at the level of 10%; Except for index funds, the return standard deviation risk and the regression coefficient of fund age of active growth funds, steady growth funds and optimized index funds are negative. The regression coefficient of size is significantly negative at the level of 10%; The regression coefficient of fund flow is significantly positive at the level of 10%.

5. Conclusion

This paper uses the in-depth learning model to classify the stock evaluation texts from the eastern fortune stock bar and the Shanghai stock index bar, constructs the investor sentiment index according to the emotional classification results, and empirically studies the impact of investor sentiment based on network information on the excess return of the fund. The conclusions are as follows: (1) overall, the investor sentiment based on online text information presents negative characteristics. (2) The trend of Shanghai stock index is negatively correlated with the investor sentiment index extracted from the 10 day trend item. (3) Considering the cross-sectional fixed effect of panel data, there is a negative correlation between investor sentiment based on online text information and the excess return of open-end stock funds. (4) After classifying the funds by investment type, investor sentiment based on network text information has a negative correlation with the excess returns of positive growth, steady growth, index and optimized index funds. The negative correlation between investor sentiment and the excess returns of index funds is the most obvious, followed by optimized index and active growth funds, Steady growth funds are the least obvious.

Acknowledgments

This work is supported by Anhui University of Finance and Economics National College Students' Innovation and Entrepreneurship training Program, “An empirical study on the impact of investor sentiment on the excess return of the fund -- from the perspective of network information” (No. 202110378009).

References

[1] Baker M, Wurgler J. Investor Sentiment in the Stock Market [J].The Journal of Economic Perspectives, 2007, 21(2): 129-151.

[2] Wang Jue, Chen Yongshuai. Can catering to investor sentiment bring excess returns to securities investment funds? [J]. Management review, 2018, 30 (7): 3-15.

[3] Li shuaiqi. Research on the construction and application of investor sentiment index incorporated into China wave index [D]. Harbin Engineering University, 2018.

[4] Fisher K L, Statman M． Investor Sentiment and Stock Returns[J]． Financial Analysts Journal, 2000, 56(2):16-23.

[5] Mariano G S, Encina M de V M. Influence of Bloomberg’s Investor Sentiment Index: Evidence from European Union Financial Sector[J]. Mathematics, 2021, 9(4):20-27.

[6] Xue Fei. Research on investor behavior based on emotion [D]. Fudan University, 2005.

[7] De Long J B, Shleifer A, Summers L H, et al. Noise Trader Risk in Financial Markets [J]. Journal of Political Economy, 1990, 98(4):120-129.

[8] Jones C M. A Century of Stock Market Liquidity and Trading Costs [J]. SSRN Electronic Journal, 2002.

[9] Long W, Zhao M Y, Tang Y R. Can the Chinese volatility index reflect investor sentiment?[J]. International Review of Financial Analysis, 2020, 73(07):211-220.

[10] Baker M, Wurgler J. Investor Sentiment and the Cross-Section of Stock Returns [J]. Journal of finance, 2006, 61(4):1645-1680.

[11] Hu J S, Sui Y Y, Ma F. The Measurement Method of Investor Sentiment and Its Relationship with Stock Market [J]. Computational Intelligence and Neuroscience, 2021(08):80-89.

[12] Xie Shiqing, Tang Sixun. The impact of investor sentiment and macroeconomic fluctuations on stock market returns [J]. Macroeconomic research, 2021 (02): 99-107.

[13] Gao Z Y, Ren H H, Zhang B H. Googling Investor Sentiment around the World [J]. Journal of Financial and Quantitative Analysis, 2020, 55(2):549-580.

[14] Wang Haoqing. Sentiment identification of retail investors based on comments of stock bar forum [D]. Shanghai Normal University, 2019.

[15] Chen Xiaojuan. Research on the influence of investor sentiment on stock price [D]. Nanjing University of Posts and telecommunications, 2020.

[16] Beaumont R, Daele M V, Frijns B, et al. Investor sentiment, mutual fund flows and its impact on returns and volatility [J]. Managerial Finance, 2008, 34(11):772-785.

[17] rational investor sentiment data[J]. Review of Behavioral Finance, 2019, 12(2):97-118.

[18] Hudson Y W, Yan M L, Zhang D L. Herd behaviour & investor sentiment: Evidence from UK mutual funds [J]. International Review of Financial Analysis, 2020(02): 71-89.

[19] Yi Li, Li Shimei. An empirical study on the emotional timing ability of Chinese open-end fund managers [J]. Journal of Guangxi University of Finance and economics, 2020, 33 (01): 99-116.

[20] Hou Hui, Naren mandura, Liu Jian, et al .Emotional contagion, individual stock volatility and stock price synchronization -- Empirical Evidence from fund heavy positions [J]. Journal of Northeast University (NATURAL SCIENCE EDITION), 2021,42 (05): 748-754.

[21] Zhou zejiong, Shi Benshan. An empirical analysis on the performance persistence of open-end funds in China [J] Exploration of economic issues, 2004 (09): 58-62.

[22] Zhao Yanyan, Qin Bing, Liu Ting. Text emotion analysis [J] Journal of software, 2010,21 (08): 1834-1848.

[23] Xiao Jun, Shi Jin. Fund performance and capital flow: is there a "redemption anomaly" in China's fund market? [J]. Economic research, 2011, 46 (01): 112-125.

[24] Ain Q T, Ali M, Riaz A, et al. Sentiment Analysis Using Deep Learning Techniques: A Review [J]. International Journal of Advanced Computer Science and Applications, 2017, 8(6):30-38.

[25] Antweiler W, Frank M Z. Is All That Talk Just Noise? The Information Content of Internet Stock Message Boards [J]. The Journal of Finance, 2004, 59(3):78-89.