Research on the Construction and Validity of Stock Market Investor Sentiment Index based on Stock Bar Comments

Yue Wu

Hunan University, Changsha 410082, China

Abstract. In the present day, the internet has become the primary place for many retail investors in China to invest, and stock forums are often the primary choice for investors to exchange and discuss. According to behavioural finance, irrational investor decisions are often the source of instability in the stock market. Therefore, the study of investor sentiment has significant implications for the optimisation of investor decisions, the design of market surveillance mechanisms and the study of behavioural finance theory. This paper combines empirical research as well as theoretical analysis, a comprehensive analysis of the raw data extracted based on Eastern Fortune stock bar comments to establish the corresponding sentiment index, and will also give an analysis and conclusion on the validity verification of the sentiment index. The research includes the following two parts: (1) Using text mining and web crawlers, text-based data and numerical data are extracted, and data mining is completed after comprehensive analysis and screening, and text sentiment analysis is used to complete the establishment of the sentiment index for investors. (2) The full-day, trading day and overnight investor sentiment indices were compared with the SSE Composite Stock Index as the research object through correlation analysis with the mainstream six sentiment proxy variables to prove their significant correlation, and then the KNN regression based on simulated annealing and the LightGBM regression model based on genetic algorithm were used to verify the prediction of investor sentiment on the stock price movement of the SSE Composite Index respectively The empirical validity of the regression models is verified.In the end, this article will propose tailored policy optimization suggestions for both investors and market regulatory agencies, in accordance with the results of the analysis.

Keywords: Stock market; Investor sentiment; Data mining; Text sentiment analysis; Validity analysis

1. Introduction

The formal system of the Chinese capital market is not perfect and the information environment still needs to be improved.[1] The advent of the Internet era has led to the flourishing of social media such as stock bars, and the exchange of information and analysis among stockholders has significantly changed the information environment for A-shares in China.[2] Investor sentiment contained in the comments of investors in stock internet forums can reflect some of the investment tendencies of Chinese retail investors, who are the largest number of retail investors in the A-share stock market and whose investment behaviour is not as professional and systematic as that of institutional investors, which can lead to irrational investment decisions.[3]As an important research guide for behavioural finance theory, investor sentiment can be of great help in understanding and analysing the investment behaviour of retail investors. Therefore, the study of investor sentiment embedded in stock bar commentaries has practical implications for the optimisation of investor decisions, the design of market monitoring mechanisms and the study of behavioural finance theory.

2. Research Methodology and Content

In this paper, we take the Eastern Fortune stock bar as the research site, and use text mining and web crawlers to combine text-based data and numerical data to extract multimodal data such as user relationships, published opinions and investment history from the Eastern Fortune stock bar, and use text sentiment analysis to construct an investor sentiment index. And the all-day, trading day and overnight investor sentiment indices and the SSE Composite Stock Index were used as the research objects to verify the validity of investor sentiment through correlation analysis, KNN regression based on simulated annealing and LightGBM regression model based on genetic algorithm, respectively. The specific research steps are as follows:

1. In this paper, Python is used to crawl the information related to the postings of users within the stock bar, and data processing is carried out on the acquired information, and text sentiment analysis methods are applied to quantify the content of the postings. The opening and closing time periods respectively. In this paper, we also use python to process the data to remove all relevant legal holidays when the stock market is closed.

2. This paper analyses the validity of the calculated sentiment indices from a financial econometric perspective. The validity analysis is divided into two steps: correlation analysis and regression analysis, in order to investigate the degree of correlation and consistency between the sentiment indices obtained in this paper and traditional investor sentiment proxies from a financial econometric perspective. Regression analysis is also used to predict the relationship between the sentiment index and the price of the SSE Composite Index, and the model is used to predict whether there is a necessary relationship between the two changes in behaviour.

3. Data Mining of Investor Sentiment Texts in Stock Bars

3.1 Emotional text data sample selection

Only with the selection of unique and representative internet sentiment data as the basic condition can we further improve the validity and significance of quantification in the process of accurately measuring and quantifying investors' sentiment using big data on the internet platform.[4]

The domestic stock market is dominated by individual investors, who make up the majority of investors, and the effective share of this group is 99.7%, which is very close to 100%. The differences between individual investors and the asymmetry of information will obviously lead to a significant shortcoming in the way individuals are informed of market changes and how they deal with them, as well as their investment sentiment, compared to institutionalised and specialised institutions and professional investors.[5]Further, due to the limitations of their professional skills such as qualifications and experience, individual investors often use the internet to obtain instant and free information, while market information is fetched, so finding the right stock exchange platform to judge investors' decisions is very effective.[6]

Some of the more active and recognised stock exchange platforms on the internet in China are: Financial Sector Forum, Sina Finance Stock Bar, Tao Stock Bar, Oriental Fortune and many other well-known media and websites. If we examine the website based on Alexa's evaluation, the stock bar of Oriental Fortune dominates, occupying the first place among all similar websites, with a daily average of more than 20 million visitors and a daily page view of more than 130 million times, making it simply a giant among forums. Its internal stock bar alone accounts for nearly two-fifths of all visits to this type of board, with nearly 28 page views per capita.[7] In addition to being far more popular than other stock bars, the Eastern Fortune stock bar also has special posting writing specifications and requirements to ensure that posting text information is complete and easy to save and trace, and has technical advantages such as easy data crawling.

Therefore, this paper will choose the posting board of the stock forum of Oriental Fortune as the data source to further collect relevant raw analysis data.[8]

The posting data in the board mainly has the following types of data: number of views, comments, posting topics, posting authors, posting time and last interaction time, etc. As the main posts of the bar have the function of accurately reflecting the main expectations of individual investors for market changes, this paper will set the key data for quantitative sentiment measurement based mainly on the main post topic data. The relevant data types crawled using the web crawler are: number of post views, post subject text and posting time.[9]

3.2 Emotional text data crawling

The information on the web pages in the Eastern Fortune Stock Forum website (http://guba.eastmoney.com/) is defined through the Hypertext Markup Language (HTML), which means that the web pages contain a large number of tags, and the document format and content is mainly organized by harmonious tags that unify the scattered resource information into a logical whole presented to the viewer This means that the website contains a large number of tags and the format and content of the documents are mainly organised by the Harmony tags, which bring together scattered resources into a logical whole.[10] Therefore, if the user needs to access the information in the Oriental Wealth webpage, it is necessary to understand the overall structure of the webpage, locate the required information through the code, and achieve precise access and storage.

Table 1 Excerpts from the data acquisition section function code

```
#   Set request headers
headers = { 'User-Agent':
UserAgent(verify_ssl=False).random.}
# Set the time range
start_date = '2022-01-01'
end_date = '2022-12-31'
# Store a list of post messages
post_data = []
# Crawl post titles
    info['title'] = div.xpath('. /span[3]/a/@title')[0]
    if info['title'] is None.
        info['title'] = 'null'
except IndexError.
    info['title'] = 'null'
# Crawl for posting times
    info['time'] = div.xpath('. /span[5]/text()')[0]
    if info['time'] is None.
        info['time'] = 'null'
except IndexError.
    info['time'] = 'null'
if start_date <= info['time'] <= end_date: # Determine if the
posting time is within the time range
    post_data.append(info)
```

In this paper, based on the characteristics of the target webpage, the Jupyter Notebook environment in Anaconda is used to write scripting code using the Python language, calling requests and lxml packages to extract information from the website, following the established code to achieve accurate search for the information required in the stock bar forum webpage, and automatically downloading and storing it in the local database. The crawl crawled data on the number of posts read, titles and posting times of stock bar users. The sample data spans the period 1 January 2022 to 31 November 2022, with a total of 4,382,514 posts crawled.

The following are excerpts of some of the functional code from Table 1 that was used to implement the crawl and store the web information.By following the above steps, the exact series of text information required can be extracted from the Eastern Fortune Share Forum website and stored in a local csv table for later data analysis and processing.

3.3 Emotional text data cleaning

In accordance with the code shown in Table 2, this paper processes the data obtained, eliminating the information of investors whose ip addresses belong to foreign countries, and at the same time needs to delete duplicate posts, posts containing advertisements, blank posts and invalid characters. Duplicate posts are simply retained, while posts containing advertisements, blank posts, etc. are not useful for the psychological study of real investors' emotions, and may even bring results contrary to investors' emotions due to profit factors such as selling lessons, so this paper chooses to delete such a series of posts, and obtains 3,277,747 valid text data belonging to some Chinese stock bar investors (including the mainland part and Hong Kong, Macao and Taiwan).

Table 2 Extracts of valid text data

Posting time	*Posting title*
11/25/2022 09:55	*There is only hope if we break the 120-day line, otherwise we will jump back and forth [show hand].*
10/20/2022 12:37	*The money lost on other stocks is back on Ping An Bank.*
9/27/2022 15:11	*3000 points held! What about tomorrow? The actual operation today is attached!*
6/10/2022 08:40	*stable [earn big]*
5/19/2022 09:32	*Another countdown list one [bared teeth]*
4/7/2022 10:06	*day good, good noodle house*
1/19/2022 22:59	*Hurry up and pull the Chinese character head ah! There is just so much money in stock, how much money has been pulling large cap stocks?*

4. Stock Bar Investor Sentiment Index Construction

4.1 Sentiment quantification of valid posting data

4.1.1 Building a complete corpus of emotion dictionaries

In this paper, after extracting valid data related to investors' postings, the sentiment dictionary method is chosen to extract and analyse the intrinsic sentiment features contained in investors' text data, using the idea of "constructing an initial sentiment dictionary - setting up a threshold to classify sentiment - constructing a sentiment index - comparing the sentiment of the day with that of the overnight". Combined with the sentiment dictionaries used in existing literature on text sentiment analysis, several Chinese sentiment dictionaries, including SnowNLP, BosonNLP, NTUSD, Hownet and Li Jun Chinese positive and negative dictionaries from Tsinghua University, were de-integrated to build a more complete corpus of initial sentiment dictionaries. At the same time, this paper also integrates several commonly used deactivation lexicons, such as SnowNLP, HHU and Baidu, in order to construct the initial deactivation lexicon.

4.1.2 Set thresholds to classify posting sentiment

In this paper, the initial deactivation lexicon is used to remove the postings containing deactivated words from the valid text crawled, and then the SnowNLP library that comes with python is used to retrieve the sentiment module to train the positive and negative words in the initial sentiment lexicon lexicon, and then the trained lexicon is used to extract the sentiment for the above data, extracting the "positive" and "negative" words in each user's posting respectively. "The probability of each post being positive is then obtained after the proportion of positive and negative words is calculated. In this paper, we set corresponding thresholds for the sentiment probability, setting the sentiment probability of [0,0.5) as negative sentiment and the sentiment probability of (0.5,1) as positive sentiment, thus making a reasonable division for the sentiment orientation of each posting, and the amount of sentiment analysis of the posting reached 97.337%, so that the valid posting data of the crawled stock bar text could be fully analyzed and utilized. The final statistics are as follows: 2,240,842 negative sentiment data and 1,036,905 positive sentiment data were obtained.

This paper conducts a sentiment analysis of user comments on stock bars in China for 24/7 i.e. (0:00-24:00), dividing all positive and negative stock bar user comment data from January-November 2022 into 11 groups by month.It is clear from Figure 1 that negative sentiment messages from stock bar users dominate, with the number of each month significantly higher than positive sentiment messages.

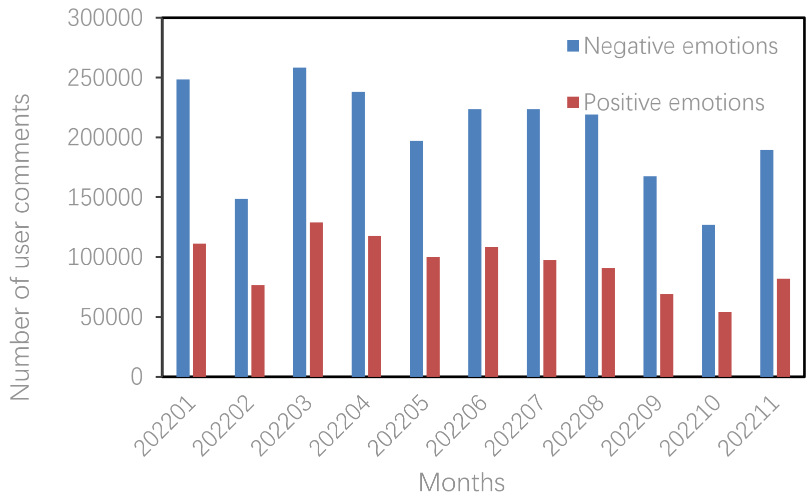

Figure 1 Monthly chart of positive and negative comments from stock bar users

This is in line with investors' expectations of using a forum such as a stock bar, i.e. a greater tendency to be willing to express their negative emotions such as loss and regret on the forum.

4.2 Construction of sentiment index of stock bar users

4.2.1 Mood index throughout the day

In this section, an all-day sentiment index is chosen to reflect the daily stock bar investor sentiment from 00:00-24:00. The posting data are excluded from the content posted during holiday periods when the A-share stock market is closed, such as the Chinese New Year and National Day holidays.

On the basis of the sentiment quantification of postings within the time interval of the stock bar already, this paper first chose to construct a bullish sentiment index for the first day, referring to the construction method of investor sentiment given by Duan Jiangjiao et al[11], taking into account the existence of neutral sentiment, formula (1) is as follows:

$$I_i = \frac{2 \times PT_i^{pos} - PT_i^{neg}}{PT_i} \tag{1}$$

Where, is the number of positive comment posts on the first day, is the number of negative comment posts on the first day, is the total number of posts on the day, the range of values between, and when the higher, the higher the bullish sentiment index of stock bar investors, at the time, the day stock bar posts are bullish sentiment. Conversely, the lower the bullish sentiment of investors.

Combined with the share bar China user posting volume and bullish sentiment index, this paper follows formula (2) to construct the investor sentiment index:

$$SI_i = I_i \times \ln(1 + PT_i) \tag{2}$$

Where, when less than 0, it can be considered that stock bar investors have a pessimistic sentiment tendency for the future trend of the general market, and this sentiment will arise when investors' expectations deviate greatly from the actual situation of the market. Pessimism is manifested by fear of market declines, an increased willingness to be risk averse, leading to a sell-off of assets, which further drives the market lower. Figure 2 illustrates the investor sentiment index for all valid trading days in the sample period from 1 January 2022 to 30 November 2022.

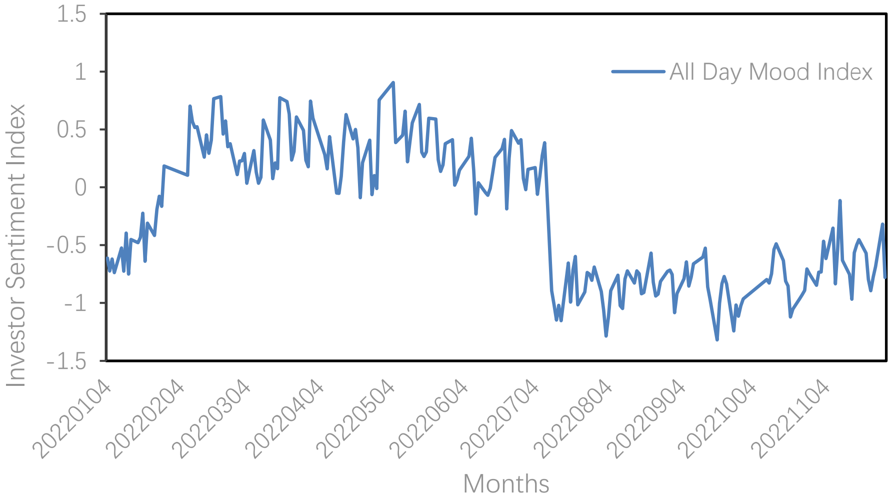

Figure 2 Statistical chart of the Effective Investor Sentiment Index

4.2.2 Trading day and overnight sentiment indices

In this section, the trading day sentiment index and the overnight sentiment index are chosen to more closely reflect the sentiment of stock bar investors during two different time periods: trading day and overnight. Trading day means 9:30am to 15:30pm Monday to Friday, while overnight means 15:30am to 9:30pm the next day after the market closes, and both exclude holiday periods when the A-share stock market is closed, such as the Chinese New Year and National Day holidays from the data.

As can be seen from the overnight and trading day sentiment indices fold chart 3, trading day sentiment fluctuates broadly in line with the overnight sentiment index, and comparison of the full day sentiment indices shows a broadly comparable trend, proving the reliability of the findings to be strong. In Figure 3, the sentiment of retail investors in the stock bar is more positive during the trading day period until July 2022, which may be related to the more favourable sentiment of investors during the opening period when stocks are up.

After July 2022, investors' sentiment was more positive during the overnight period than during the trading day, which, combined with the rise and fall of the A-share stock market, may be due to the fact that investors are more likely to have negative sentiment due to losses during the market as a result of the large fall in the stock market, and that there is a "negative proximate cause effect" when investors review the market after the market.

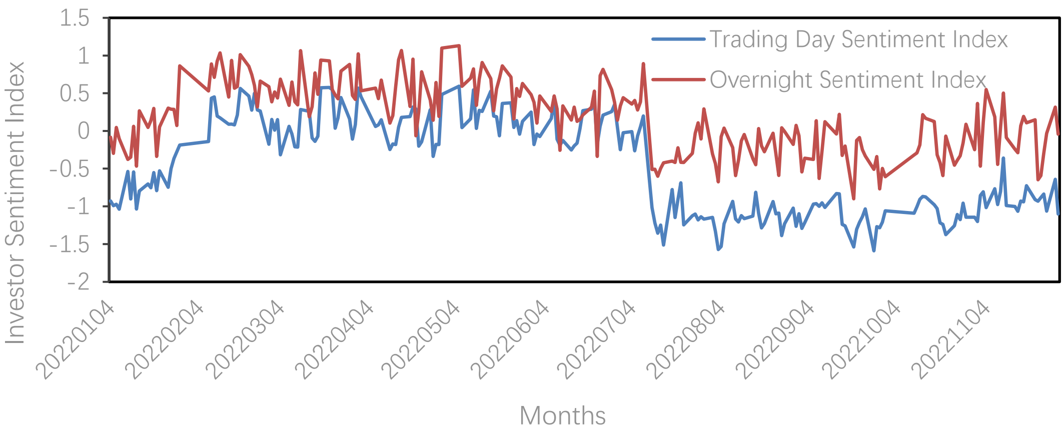

Figure 3 Comparison of investor sentiment indices

This is due to the fact that investors are more likely to have negative emotions due to losses during the market, and when the market is reviewed after the market, there is a "negative proximate effect" where investors believe that there is a good chance of a bottoming out after a continuous decline in the stock market, and therefore have expectations for the next trading day. As a result, investors may choose to add to their investment in what they believe to be the right range after the resumption of the market, thus "getting their money back".

5. Analysis of the validity of the impact of sentiment indices on the stock market

5.1 Correlation analysis with mainstream emotional proxy variables

As the investor sentiment index constructed in this paper is a high-frequency indicator constructed using sentiment text data mining and text analysis, this paper combines each sentiment proxy variable with the monthly frequency investor sentiment index separately to complete a correlation analysis of six different sets of variable comparisons to verify its validity as a sentiment measure. In this paper, the sample time span of the model is set from January 2022 to November 2022, with a total of 11 monthly data, and Table 3 shows the data table of sentiment proxy variables.

Table 3 Data table of emotional proxy variables

Date	*China:Consumer Confidence Index*	*China:Investor Confidence Index*	*Turnover:SSE Composite Index*	*Turnover Rate:SSE Composite Index*	*Turnover:SSE Composite Index*	*Number of new A-share accounts opened on SSE*
Month	*—*	*—*	*billion shares*	*%*	*Billion*	*households*
202201	*121.50*	*54.60*	*6625.72*	*15.81*	*82548.10*	*2344200.00*
202202	*120.50*	*51.70*	*5377.25*	*12.83*	*62559.47*	*2428700.00*
202203	*113.20*	*52.20*	*8702.01*	*20.96*	*100541.96*	*3964600.00*
202204	*86.70*	*50.80*	*7477.91*	*17.98*	*76450.60*	*2147700.00*
202205	*86.80*	*62.10*	*6747.27*	*16.19*	*71584.48*	*2100000.00*
202206	*88.90*	*65.40*	*8739.39*	*20.97*	*105106.90*	*2352400.00*
202207	*87.90*	*61.50*	*6880.87*	*16.46*	*87708.70*	*1967500.00*
202208	*87.00*	*56.00*	*7050.01*	*16.64*	*94875.14*	*2141000.00*
202209	*87.20*	*54.50*	*5490.16*	*12.93*	*65573.37*	*1805400.00*
202210	*86.80*	*57.20*	*4138.63*	*9.63*	*54453.87*	*1433900.00*
202211	*85.50*	*60.50*	*6869.08*	*15.96*	*85844.37*	*1773300.00*

Note: Data from Wind.

The sentiment proxy variables are consumer confidence index, investor confidence index, SSE Composite Index turnover, SSE Composite Index trading volume, SSE Composite Index turnover rate and the number of new investors opening accounts in A-shares in the previous month, which are all obtained from Wind.

5.2 Normality test

In this paper, for the quantitative data, the normality test is first carried out, firstly, the Shapiro-Wilk test is carried out on the data, that is, the normality test carried out on small samples less than 5000, when the obtained P>0.05, it means that it does not show significance and conforms to the normal distribution, and vice versa, it means that it does not conform to the normal distribution; if the absolute value of kurtosis of a sample is less than If the absolute value of kurtosis of a sample is less than 10 and the absolute value of skewness is less than 3, and it is found that the sample basically has the characteristics of normal distribution after visual analysis through normal distribution histogram, etc., then it can be said that the sample conforms to normal distribution. The results obtained are shown in Table 4:

Table 4 Normality test results

Variable Name	*Median*	*Mean*	*Skewness*	*Kurtosis*	*S-W test*	*K-S test*
China Consumer Confidence Index	*87.2*	*95.636*	*1.246*	*-0.442*	*0.65(0.000***)*	*0.403(0.040**)*
China Investor Confidence Index	*56*	*56.955*	*0.398*	*-1.041*	*0.941(0.536)*	*0.142(0.957)*
Turnover	*6869.078*	*6736.209*	*-0.281*	*0.158*	*0.941(0.536)*	*0.195(0.728)*
Change of hands	*16.192*	*16.034*	*-0.246*	*0.178*	*0.938(0.502)*	*0.201(0.693)*
Turnover	*82548.103*	*80658.816*	*-0.067*	*-0.967*	*0.976(0.941)*	*0.097(1.000)*
Number of new SSE A-share accounts opened	*2141000*	*2223518.182*	*2.073*	*5.874*	*0.785(0.006***)*	*0.285(0.278)*
All-day sentiment index	*-8.291*	*-4.925*	*-0.015*	*-1.958*	*0.85(0.042**)*	*0.223(0.573)*

*Note: ***, ** and * represent 1%, 5% and 10% levels of significance respectively.*

The S-W test was chosen where the sample size N of the China Consumer Confidence Index was less than 5000. The test shows that the China Consumer Confidence Index sample does not satisfy a normal distribution with a significance level of 0.000***, indicating that the original hypothesis is rejected. According to the results of the data analysis, the kurtosis of this data set is -0.442, with an absolute value of less than 10, and the skewness is 1.246, with an absolute value of less than 3. Also, as observed through the histogram of normal distribution Figure 4, the graph shows a bell-shaped curve, indicating that the data set basically meets the normal distribution. Although it is not an absolute normal distribution, it is still acceptable.

5.3 Constructing a predictive model of investor sentiment - SSE Composite Index

5.3.1 KNN regression prediction model construction based on simulated annealing

In this paper, we use simulated annealing algorithm to build the model and set the initial parameters as follows: initial temperature: 100 degrees; termination temperature: 2 degrees; number of iterations per degree: 100; cooling factor: 0.98. In order to verify the effectiveness of the constructed investor sentiment index in predicting the closing price sequence of the stock index, a control group without the investor sentiment index is set up and compared with the experimental group. . The prediction effectiveness of the investor sentiment index was assessed by comparing the model prediction results of the experimental and control groups.

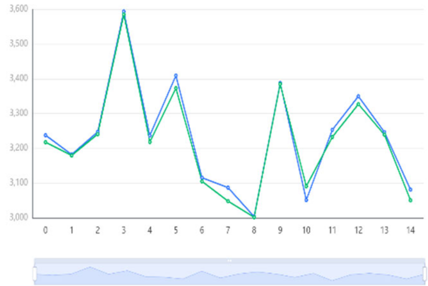

Figure 4 Test Data: SSE Index Forecast Results Chart (KNN)

Table 5 Results of the evaluation of the KNN model for the all-day sentiment index

Number set	MSE	RMSE	MAE	MAPE	R^2
Training set	644.284	25.383	19.06	0.594	0.976
Cross-validation set	1419.545	36.209	27.004	0.84	0.946
Test set	655.887	25.61	21.63	0.684	0.971

Figure 4 plots the results of the evaluation of the all-day sentiment index-stock index model for the test data and Table 6 shows the results of the model evaluation. It can be seen that the R2 is higher for both the test and training sets, with over 97%, indicating that the stock index model using the all-day sentiment index is more accurate than the model using only the mean when comparing the predicted values to the case of using only the mean. The RMSE (root mean squared error) was below 26 for both the training and test sets and below 37 for the cross-validation set, taking smaller values, while the MAPE (mean absolute percentage error) was below 1 for each data set, indicating that the model is highly accurate.

Table 6 Trading Day sentiment index KNN model evaluation results

Number set	MSE	RMSE	MAE	MAPE	R^2
Training set	682.409	26.123	18.912	0.588	0.973
Cross-validation set	1346.334	35.702	26.179	0.812	0.943
Test set	1548.164	39.347	26.498	0.842	0.958

Table 7 Overnight sentiment index KNN model evaluation results

Number set	MSE	RMSE	MAE	MAPE	R^2
Training set	670.69	25.898	18.977	0.593	0.975
Cross-validation set	1413.756	36.916	27.872	0.868	0.937
Test set	757.796	27.528	20.345	0.625	0.973

Tables 5 and 6 show the evaluation results of the trading day sentiment index forecasting model and the overnight sentiment index forecasting model respectively, from the results it can be seen that both the test machine and the training set can explain 95% of the results, indicating a high degree of prediction, because this paper has chosen a data shuffling function for K nearest neighbours (KNN), so there may be inconsistencies in the models trained each time, the training data set are all the SSE Composite Index for the interval January 2022 to November 2022.

6. Conclusion

From the above validity analysis, it is evident that an investor sentiment index based on stock bar comments can have a significant impact on the stock market by:

(1) The investor sentiment index constructed in this paper is significantly correlated with the mainstream sentiment proxy variables, and there is an inevitable link between it and the price

behaviour of the SSE Composite Stock Index, providing reliable data support for the development of the regression predictions in this paper.

(2) The investor sentiment index constructed in this paper has a relatively good predictive effect on the SSE Composite Stock Index series, indicating that the presence of investor sentiment can cause investors to irrationally purchase or bid for a stock, thus causing a change in the price of that stock. Realistically explained, as the sentiment of group investors has an aggregating and amplifying effect, while individual investors in the stock market usually attribute the high investment returns gained to one's wise investment decisions and the losses to the poor external environment. Therefore when the market underperforms, investor sentiment performs generally and therefore has relatively little impact on the stock market.

(3) The combination of the full-day, day-ahead and overnight sentiment indices constructed in this paper provides a better prediction of changes in the composite stock index, and the degree of correlation with the true value suggests that the model is reasonably effective in achieving the index prediction problem. Realistically interpreted, investors want to gain and avoid losses, so the share price fluctuations during the opening trading period will have a greater impact on investors' psychological volatility, which is caused by investors' expectations as well as loss concerns, and these emotions will cause investors to care about factors other than intrinsic value and thus make irrational decisions.

References

[1] Yang L. S., Yang J.. A study on monetary policy, investor sentiment and stock market liquidity - an empirical analysis based on TVP-SV-VAR model[J]. Price Theory and Practice, 2021(09):142-145+203.

[2] Xu Qiqi, Bo Zhongpu, Jiang Cuiman. A study on the relationship between online sentiment and stock market returns based on quantile Granger causality [J]. Management Science, 2017, 30(3):147-160.

[3] Yu, Q. J., Zhang, B.. Investors' limited attention and stock returns - an empirical study using Baidu index as attention [J]. Financial Research, 2012(08):152-165.

[4] Zhao Longkai, Lu Ziyu, Wang Zhiyuan. An empirical investigation of the relationship between stock returns and Baidu search volume[J]. Financial Research, 2013(04):183-195.

[5] Zhang Yihao, Li Yuan, Su Zhongfeng, Zhang Zelin. Can internet search predict the stock market? [J]. Financial Research, 2014(02):193-206.

[6] Yang X. L., Shen H. B., Zhu Y. Local preferences, investor sentiment and stock returns:empirical evidence from online forums[J]. Financial Research, 2016(12):143-158.

[7] Zhou Xiaohua, Huang Miao. Limited investor attention, composite sentiment indicators and stock returns[J]. Finance and Accounting Monthly, 2017(03):99-105.

[8] Wang Meijin, Sun Jianjun. Chinese stock market returns, return volatility and investor sentiment [J]. Economic Research, 2004(10):75-83.

[9] Wang Xingfen, Liu Changhao. Iron ore futures price volatility and investors' market sentiment: an analysis of the sentiment index collected from "Oriental Fortune" Internet news data[J]. Price Theory and Practice, 2020(11):81-84+183.

[10] Meng Y, Chang J. The impact of investor sentiment on the scale effect [J]. Statistics and Information Forum, 2019, 34(04):98-104.

[11] Yin H, Wu XY. The predictive role of investors' high-frequency sentiment on intraday stock returns [J]. China Industrial Economics, 2019(08):80-98.