



## Research article

## The weekly cycle of investor sentiment and the holiday effect-- An empirical study of Chinese stock market based on natural language processing

Qing Liu<sup>a,b</sup>, Xinyuan Wang<sup>b,c</sup>, Yamin Du<sup>a,b,\*</sup><sup>a</sup> College of Economics and Management, Huainan Normal University, China<sup>b</sup> Graduate School of Management of Technology, Pukyong National University, South Korea<sup>c</sup> School of Economics and Management, Hulunbuir University, China

## ARTICLE INFO

## Keywords:

Investor sentiment  
Weekly cycle  
Holiday effect  
Stock market  
Behavioral finance

## ABSTRACT

Investor sentiment is an important factor that affects stock prices, stock market returns, and asset pricing. However, the fluctuation patterns and factors influencing investor sentiment have received less attention from scholars. This study uses text messages from stock investors' social networks and natural language processing techniques to reveal sentiment fluctuation laws of stock market investors. An investor confidence index (ICI) is constructed by quantifying sentiment in investor messages on social networks. By taking this index as a proxy for sentiment, we measure the candidate fluctuation periods of investor sentiment using a Fourier transform. The significance test then determines the significant cycle of investor sentiment within seven days. Based on this, cluster analysis further reveals that investor sentiment in the 7-day cycle has a 5 + 2 cycle of variability. That is, from Monday to Friday, investor sentiment is disturbed by stock market sentiment showing profit-seeking and risk-averse preferences, while during the weekend holiday, stock market disturbance to investor sentiment becomes lower, investor sentiment is substantially higher, and volatility is narrowed, showing a typical holiday effect. The analysis also shows that the recurring cycle of 5-day trading days and 2-day holidays is a direct exogenous factor contributing to the 7-day cycle of investor sentiment. This study provides a new perspective for studying "investor sentiment," "day of the week effect," and "behavioral finance."

## 1. Introduction

Investor sentiment is closely related to the "day-of-the-week effect" (Berument and Kiyamaz, 2001), stock market returns" (Fisher and Statman, 2000), "asset pricing" (Brown and Cliff, 2005; Ljungqvist et al., 2006), "the closed-end fund puzzle" (Lee et al., 1991), "stock market liquidity" (Liu, 2015) and even "stock market crises" (Zouaoui et al., 2011). Insight into the fundamental patterns of fluctuations in stock market investor sentiment is of great significance in the study of financial policy, financial risk, and investment portfolios. This study uses convolutional neural networks to classify investor messages from social networks according to the strength of investor confidence, as indicated in the text. Further, an investor confidence index is constructed as a sentiment proxy following the methodology of Antweiler and Frank (2004). Combined with stock market data, we investigate two basic issues related to investor sentiment in the stock market in depth: the fluctuation pattern of investor sentiment and the basic causes affecting the fluctuation of investor sentiment. This study offers a fresh point of reference for

research on investor sentiment and the analysis of related economic events.

From De Long et al. (1990), who first hypothesized that "investors are influenced by emotions," to Baker and Wurgler (2007), who explicitly emphasized that "the question is no longer whether investor sentiment affects stock prices as it did decades ago, but how to measure investor sentiment and quantify its impact," it took 21 years, during which a branch of finance closely related to investor sentiment, behavioral finance, was also developed. In previous studies, scholars fully discussed how investor sentiment affects stock market returns (Brown and Cliff, 2004; Fisher and Statman, 2000); how investor sentiment affects asset prices (Lemmon and Portniaguina, 2006; Ljungqvist et al., 2006); and how investor sentiment can be used to predict stock returns (Neal and Wheatley, 1998). In recent years, in addition to traditional theoretical validation, researchers have begun to apply investor sentiment to other markets, such as renewable energy (Song et al., 2019), cryptocurrencies (López-Cabarcos et al., 2021), and crude oil markets (Qadan and Nama, 2018) among others. The relationship between the stock market and

\* Corresponding author.

E-mail address: [duyamin@pukyong.ac.kr](mailto:duyamin@pukyong.ac.kr) (Y. Du).

investor sentiment has also been discussed using new approaches. [Schadner \(2021\)](#) used multifractal detrended fluctuation analysis to study the multifractal property and temporal persistence of U.S. and European stock market sentiment. This gave him more information about how investors acted. [Wang et al. \(2022\)](#) studied investor sentiment and the stock market through multilayer network analysis. It was concluded that fluctuations in either direction in the investor sentiment layer affect the stock return layer. However, many economic phenomena related to stock market investor sentiment are still controversial decades after they were proposed, such as the “day of the week effect” in the stock market.

In 1965, Fama first pointed out the anomaly of Monday stock returns ([Fama, 1965](#)), initiating the study of the “day of the week effect” in the stock market. The distribution of stock returns can vary depending on the day of the week ([Gibbons and Hess, 1981](#)). The most typical feature is the anomalous return on Mondays and Fridays, suggested by [Lakonishok and Levi, 1982](#)). Recently, there has been increasing awareness that changes in investor sentiment may be a potential explanation for the weekday effect ([Chiah and Zhong, 2019, 2021](#)), and [Kim and Ryu \(2022\)](#) used changes in investor sentiment to explain post-weekend stock price declines in a very recent study. As discussed earlier, researchers have used investor sentiment to explain the “day of the week effect” and have reached various possible conclusions, but how does investor sentiment fluctuate? How does investor sentiment relate to holidays? However, there is still a lack of systematic research and extensive market validation.

Investor sentiment is complex and challenging to measure, and traditional sentiment proxies fall into three main categories based on market indicators, survey indices, and special events. Proxies based on market indicators indirectly measure investor sentiment using market indicators such as trading volume, closed-end fund discounts, first-day returns on initial public offerings (IPOs), and the number of IPOs. The method of sentiment proxies is based on survey indices that quantify investor sentiment by collecting investors' optimistic or pessimistic expectations about the stock market through surveys such as the Consumer Confidence Index ([Brown and Cliff, 2005](#)), UBS/GALLUP Investor Optimism Index ([Lemmon and Portniaguina, 2006](#)), and Investment Newsletter ([Qiu and Welch, 2004](#)). An approach based on special events, often using special social events as sentiment proxies, COVID-19 is perhaps the most representative current example. [Naseem et al. \(2021\)](#) analyzed the impact of COVID-19 on investor psychology and the impact of investor psychology on the stock market. Market indicator-based sentiment proxies are often impure and mixed with many extraneous bubbles, making it difficult to determine whether this evidence is a coincidence of multiple effects. Proxy methods based on survey indices are usually limited by the sample size, and the questionnaire respondents' answers are subjective ([Da et al., 2015](#); [Singer, 2002](#)). An approach based on special events is typically used to study particular social phenomena and is not universal.

In the Internet era, it has become an effective complement to the above-mentioned traditional sentiment proxies for measuring investor sentiment based on search engines. This approach has the advantages of high-frequency accessibility and objective data ([Da et al., 2015](#)). [Da et al. \(2011\)](#) argued that searches reflect investor attention. Thus, Google search frequency is a new, direct measure of investor attention. However, investor sentiment obtained using either of these proxies is only a statistical description of the data indicators. [López-Cabarcos et al. \(2017\)](#) show that investor profiles help understand the influence of social networks on investor sentiment and the influence of social networks on stock markets. Improvements in natural language processing provide us with new ways to measure how people feel about things in text and on social media. [Sprenger et al. \(2014\)](#) used natural language techniques to analyze the existence of associations between tweet sentiment and stock returns, news and trading volume, and disagreement and volatility, as well as the mechanisms that lead to the effective aggregation of information in microblogging forums. This method of mining social media sentiment using natural language processing techniques has the

advantage of easy data availability and high credibility as a direct and effective investor sentiment indicator ([Arbieu et al., 2021](#); [Hirschberg and Manning, 2015](#)).

In this study, based on credible natural language processing techniques and large sample size, we study the fluctuation patterns and factors influencing stock market investors' sentiment fluctuations. We used crawler technology to obtain more than 2.66 million comments from investors on China's SSE from 2019 to 2021 and used a convolutional neural network model to classify the comments into three categories: “positive,” “neutral,” and “negative” according to the strength of investors' confidence in the stock market. Based on this, we constructed an investor confidence index (ICI). Taking ICI as a proxy for sentiment and SSE stock data as a proxy for stock prices, we combine investor sentiment panel data and stock market price panel data to conduct an in-depth study of the patterns and underlying causes of investor sentiment fluctuations in the stock market. We also consider the special characteristics of investor sentiment in atypical trading cycles (weekly holidays > two days).

The main contribution of this paper is that we have revealed the cyclical fluctuation pattern of investor sentiment for the first time in the Chinese stock market, giving a comprehensive explanation for investor sentiment, the “day of the week effect,” and the holiday effect, and our main findings are as follows: Investor sentiment shows a 7-day cycle with a 5 + 2 pattern; During the five trading days, investor sentiment showed the “rational person” characteristic of seeking profits and avoiding risks; During the 2-day holiday, investors briefly forgot to “work,” and investor sentiment increased significantly, and the increase was not affected by the stock market, showing the “social person” characteristic.

The subsequent sections of this paper are organized as shown in [Figure 1](#): Section 2 is the hypothesis, data, and methodology, which describes the data sources and quantification of investor sentiment, Section 3 discusses the investor sentiment cycle and the holiday effect, and Section 4 addresses investor sentiment fluctuations in atypical trading cycles and Section 5 concludes.

## 2. Hypothesis, data, and methodology

### 2.1. Hypothesis

Broadly speaking, investor sentiment is a belief about future cash flows and investment risk ([Baker and Wurgler, 2007](#)), and investors' assessment of future cash flows and investment risk in the stock market is derived from fundamental judgments about past and current stock markets. Thus, stock market fundamentals affect investor sentiments. Research shows that investor sentiment affects stock market returns, prices, and cash flows ([Baker and Wurgler, 2007](#); [McGurk et al., 2020](#); [Sayim and Rahman, 2015](#)). Based on these facts, the first hypothesis of this study is proposed.

**H1.** Investor sentiment is significantly related to stock prices.

[Hong and Wang \(2000\)](#) used social media data to study individual-level emotional changes on social media. They found that people were happier on the weekends. Therefore, investor sentiment on social media is better on weekends. This leads to the second hypothesis.

**H2.** Investor sentiment is higher during holidays, and there is a holiday effect.

Scholars have shown that human emotions deteriorate over the course of the day ([Hong and Wang 2000](#)). Research on the effect of the day of the week suggests that human emotions vary depending on the day of the week it is ([Chiah and Zhong, 2019, 2021](#)). Based on the above analysis and **H2** hypothesis, the third hypothesis of this study is proposed.

**H3.** There is a significant cycle of volatility in investor sentiment, which is associated with the cyclical opening and closing of the stock market.

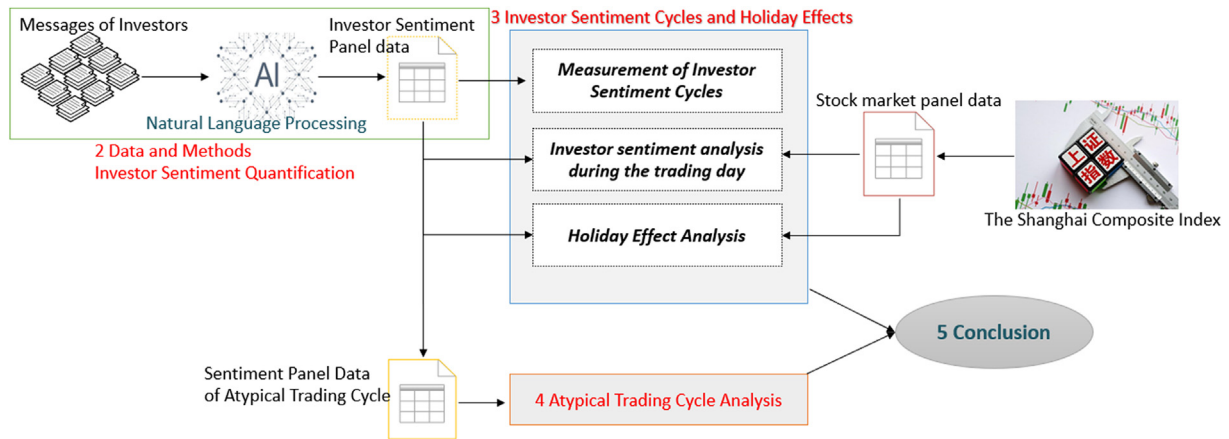


Figure 1. Research framework.

2.2. Data sources

Orient Fortune is China's leading integrated Internet wealth management operator that provides Internet-based financial information, data, and trading services to more than 100 million users. The company's "stock bar is the most popular Chinese stock- and fund-exchange community. The investor messages used in this study were from the "Stock Bar" of Oriental Fortune. As shown in Figure 2, each message has a title, author name, date, reading number, comment number, text, and so on, similar to the text messages used by Antweiler and Frank (2004), where investor messages are generally short and mainly express personal opinions without too much analysis. The number of readings conveys the message's ability to spread. These comments may represent the approval or disapproval of other investors. At the same time, we are not concerned with who the authors of the messages are, so we collect the message information, mainly including the title, text, date, number of readings, and number of comments.

In this study, we use the last three years of investor message records to obtain conclusions in line with the current situation. We used crawler technology to obtain all investor messages on the "SSE Index Bar" from 2019-01-01 to 2021-12-31. Figure 3 shows the distribution of investor messages over 36 months, from which we can see that the monthly investor messages are not uniform. The total number of investor messages exceeded 2.66 million, with a total reading of 1.892 billion, a total comment count of 5,943,800, a maximum daily message volume of 15, 2 and 61, an average daily message volume of 2,432, and a minimum daily message volume of 13. Detailed statistics are presented in Table 1.

Since the investor comment message discussion topic is the SSE Composite Index, this study uses the SSE Composite Index as a proxy for the composite stock price. We collected the opening price, closing price, highest stock price, lowest stock price, volume, and other index data of the SSE index to construct the stock market data panel and used the statistical interval consistent with the investor's message data for a total of 730 trading days.

---

TITLE Insufficient downward momentum. The May Stock Index is dominated by the horizontal accumulation of downward momentum

NAME Cat catching mouse

DATE 2021-05-10 13:46:28

READERS 221

COMMENTS 0

TEXT: Because of the sharp decline some time ago, many chips were missed. Stop losing money Coupled with the recent emergence of several slogan-style stabilizations of stock measures, stock indices are expected to continue to fall in mid-late June.

---

TITLE Possible plunge in the afternoon, watch out for risks.

NAME Re Qulong

DATE 2021-04-10 12:22:33

READERS 200

COMMENTS 5

TEXT Possible plunge in the afternoon, watch out for risks.

---

Figure 2. Samples of investor messages.

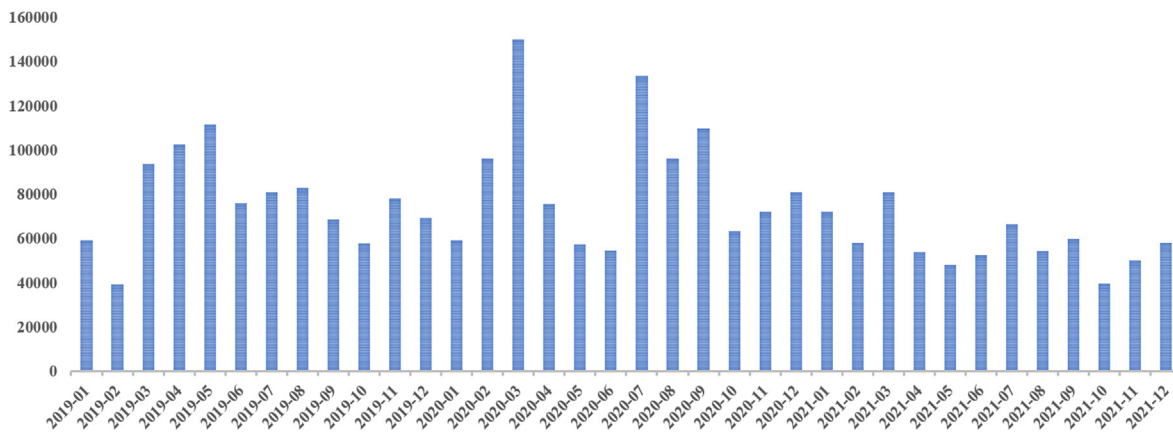


Figure 3. Monthly distribution of investor messages.

Table 1. Statistical description of investor messages.

	Daily posts volume	Daily reading volume	Daily comment volume
max	15261	44924349	43562
mean	2433	1729255	5433
min	13	24426	22

Investors' messages contain the expression of investors' sentiments about stock market quotes, so stock market quotes and investors' sentiments have a natural causal relationship. In addition, based on existing research, investors' sentiments counteract stock market fundamentals (Baker and Wurgler, 2007; Neal and Wheatley, 1998; Zouaoui et al., 2011). Thus, investor sentiment and stock market sentiment have a causal link, and the correlation between investor sentiment and fundamental stock market sentiment can be understood as a causal link between the two.

2.3. Sentiment classification of messages

As the number of investor messages is as high as 2.66 million, it is difficult to classify the sentiment of all messages manually, and we use supervised learning to solve this problem. As shown in Figure 4, 1.1% (about 30,000) of all the messages were sampled randomly for manual classification, and the messages were classified into three categories, "positive," "neutral," and "negative," according to the strength of confidence in the market expressed in the investors' messages. Several convolutional neural networks were trained using 90% of the manually labeled data as the training dataset. The remaining 10% of the labeled data were used to evaluate the classification performance of the neural

networks to obtain a useable deep learning model, which was finally used to classify unclassified text for sentiment.

2.3.1. Text encoding

Before using the deep learning model for text classification, it was necessary to encode the text data. The encoded text generates a set of feature vectors that represent the words in the dataset. Then, the feature vectors are input into the network model to output the sentiment classification of the message.

The text data of the investors' messages consist of a variety of contents, except for the text, and the rest of the contents will not contribute to the sentiment classification of the text, which is called noise in text processing. In text preprocessing, we remove numbers, punctuation marks, extra spaces, and special characters that are difficult to understand, such as (, [, {, &, etc., and replace the web links in the text with the string "URL" and the username with the string "USERNAME," and then start text encoding.

Word2vec is a simple two-layer neural network for processing text and representing words as vectors. The model can transform the input text data into a set of vectors as the output (Goldberg and Levy, 2014). Owing to the peculiarities of Chinese, the text needs to be split into words before encoding. As shown in Figure 5, we split the sentences into words and then use Word2vec to encode and transform the text into word vectors.

2.3.2. Network training and text classification

A word vector without sentiment classification can be represented as a (vector, ?) whereas a sentiment-classified word vector whose sentiment is deterministic can be denoted as (vector, sentiment). Using a large amount of classified data, we trained a neural network to learn the classification rules of the labeled datasets. Then, we attempted to classify

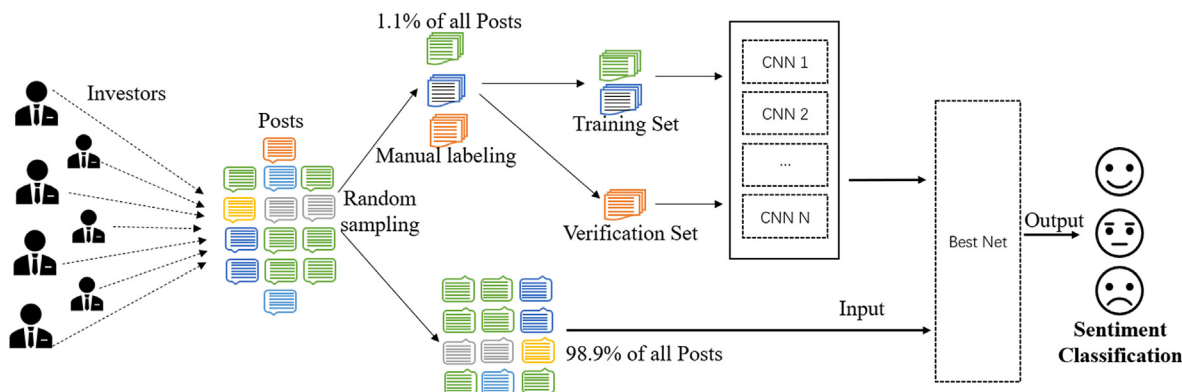


Figure 4. Flowchart of text sentiment classification.

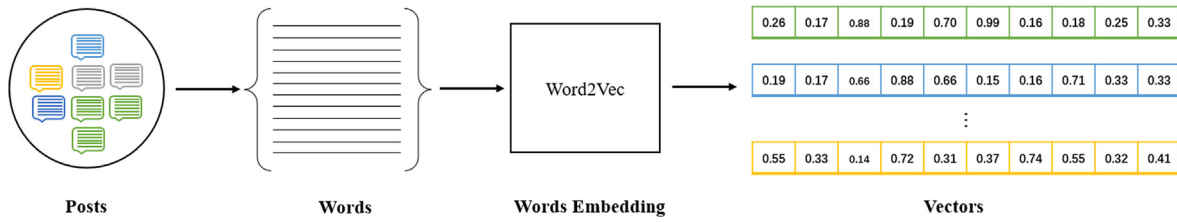


Figure 5. Flowchart of text encoding.

the unclassified messages to test their classification performance. If its classification accuracy can reach the expected result, we can use this network model to help classify the sentiment; its classification process is shown in Figure 6.

We trained several convolutional neural networks using 27,000 manually labeled data and then calibrated the performance of the networks using 3020 sample data to obtain the network model with the best classification performance for classifying unclassified messages by sentiment tendency. The classification performance of the network model used in this study is shown in Table 2, which shows that the classification accuracy of the samples with “positive” sentiment is 90.99%, and the classification accuracy of the samples with “negative” sentiment is 94.62%, and the classification accuracy of the samples with “neutral” sentiment is 81.83%. The combined accuracy of the network model is 89.14%, which is higher than the 88.1% of Antweiler and Frank (2004) and 85.4% of Xiong et al. (2017). In particular, the classification accuracy of the positive and negative samples, which are the most critical for constructing the investor confidence index, are 90.99% and 94.62%, respectively, which are significantly higher than the accuracy rates of similar studies. Therefore, the classification model can be considered to have a high degree of confidence and no systematic error.

2.4. Investor confidence index and Agreement Index

After obtaining the sentiment classification of all messages, the investor confidence index (ICI) was defined to study the aggregate sentiment of investors expressed by the daily average of 2,433 posts. Referring to Antweiler’s method of defining the Bullish Index, taking  $N_t^{positive}$  to represent the number of messages with confidence in the stock market on day t and  $N_t^{negative}$  to represent the number of messages without confidence, the investor confidence index is constructed by the formula

$$ICI = \ln \left[ \frac{1 + N_t^{positive}}{1 + N_t^{negative}} \right] \tag{1}$$

The higher the proportion of messages expressing confidence in the market in the stock bar, the higher the ICI, indicating that investors are

Table 2. Confusion matrix of prediction performance.

True classification	Prediction classification			Accuracy
	Positive	Negative	Neutral	
Positive	1010	79	21	90.99%
Negative	21	880	29	94.62%
Neutral	24	154	802	81.83%
Total				89.14%

Note: This table is constructed with reference to the confusion matrix commonly used in field of machine learning.

more confident in the market, and the lower the proportion of messages expressing confidence in the market in the stock bar, the lower the ICI, indicating that investors are lacking confidence in the market, in this paper, the ICI is used as a proxy for investor sentiment, so a high ICI represents high sentiment and a low ICI represents low sentiment.

Investors’ sentiment tendencies disagree, and with reference to Antweiler’s suggestion, we constructed the Agreement of Investors’ Opinions Index (Agree Index or Agree) as follows:

$$Agree = 1 - \sqrt{1 - \left( \frac{N_t^{positive} - N_t^{negative}}{N_t^{positive} + N_t^{negative}} \right)^2} \tag{2}$$

If all messages were positive or negative, the Agree Index was one. A larger Agree Index indicates that investors disagree less. As suggested by Liu et al. (2022), formula (2) was not used for correlation analysis in this study.

2.5. Variables definition

Using the above work, we constructed a panel database sorted by date. In addition to the Investor Confidence Index and Agree Index, we also collected the daily message volume, message reading volume, and comment volume. For the convenience of data processing, in this study,

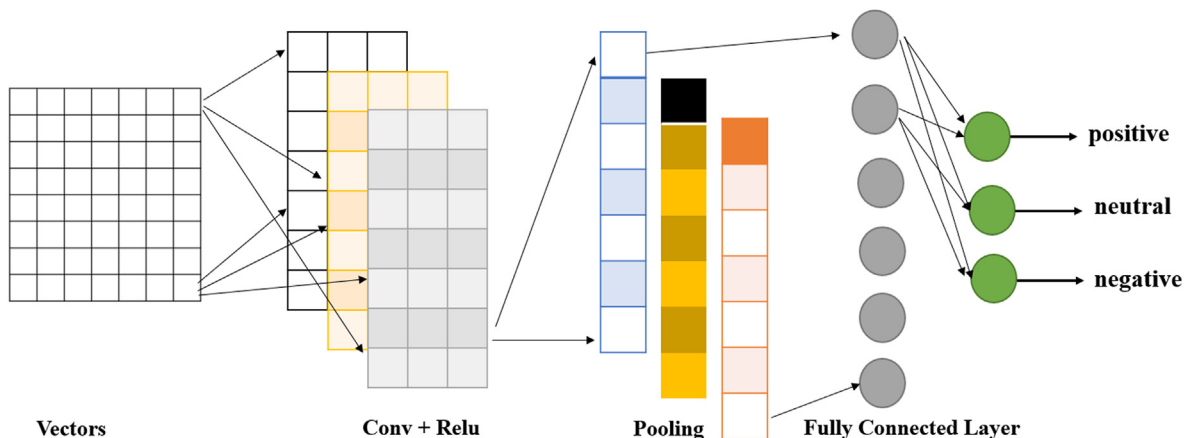


Figure 6. Sentiment classification diagram of convolutional neural network.

we used numbers from 1 to 7 to represent Monday to Sunday. To study the impact of stock prices on investor sentiment, we introduce data related to the SSE index as a proxy for stock market quotes. Table 3 presents definitions of the variables used in this study.

The statistical data used in this study are from 2019-01-01 to 2021-12-31. However, there is a data gap on 2019-01-01; therefore, the actual data start date is 2019-01-02, and there are 1094 consecutive observations. Table 4 shows a statistical description of the main data used in the study. Panel A shows the basic structure of the database, and Panel B is a statistical description of the panel database. Panel C shows the statistical description of the stock price panel; Panel D shows the statistical description of the stock price panel on Friday; and Panel E is a description of statistical indicators for the complete trading week. The size of the statistical sample for Panel C is 730 because there are only 730 trading days in the sample of 1094 observations.

**Table 3.** Definition of variables.

Variables of stock markets		Variables of sentiment	
P <sub>O</sub>	Opening Price	WI	Day of the week number. The range of values is from 1 to 7, representing Monday to Sunday, respectively.
P <sub>C</sub>	Closing Price	PV	Volume of messages
P <sub>PC</sub>	Previous day's closing price	RV	Reading Volume
P <sub>H</sub>	The highest stock price of the day	CV	Comments volume
P <sub>L</sub>	The lowest stock price of the day	ICI	The investor confidence index, used as a proxy for investor sentiment
P <sub>H-L</sub>	Range of stock price fluctuation on the day	Agree	Agreement index which describes the disagreement of investors' opinions
P <sub>M</sub>	Up/Down Amount of the day	ICI <sub>WI</sub>	ICI for a day of the week
V	The volume of the day	ICI <sub>6,5</sub>	Incremental ICI on Saturday, obtained by ICI <sub>6</sub> - ICI <sub>5</sub>
MV	Amount of the day's trading	ICI <sub>7,6</sub>	Incremental ICI on Sunday, obtained by ICI <sub>7</sub> - ICI <sub>6</sub>
P <sub>O5</sub>	Friday's opening price	ICI <sub>MAX</sub>	The maximum ICI value for the day of the week in the trading cycle
P <sub>C5</sub>	Friday's closing price	ICI <sub>MEAN</sub>	The average ICI value for the day of the week in the trading cycle
P <sub>H5</sub>	Friday's highest stock price	ICI <sub>MIM</sub>	The lowest ICI value for the day of the week in the trading cycle
P <sub>L5</sub>	Friday's lowest stock price	ICI <sub>MAX-MIM</sub>	The difference between the high and low ICI of the day of the week in the trading cycle
P <sub>M5</sub>	Friday's up/down amount		
P <sub>H5-L5</sub>	Friday's range of fluctuations		
P <sub>HW</sub>	The highest stock price of the week		
P <sub>LW</sub>	The lowest stock price of the week		
P <sub>HW-LW</sub>	The difference between P <sub>HW</sub> and P <sub>LW</sub> .		
P <sub>DELTA-HW-LW</sub>	The cumulative stock price movement of the week, obtained by adding up the daily P <sub>H-L</sub>		
P <sub>DELTA-M</sub>	The cumulative up/down amount of the week, obtained by adding up P <sub>M</sub>		
P <sub>DELTA- M </sub>	The cumulative absolute up/down amount of the week, obtained by adding up the absolute value of P <sub>M</sub>		

### 3. The investor sentiment cycle and the holiday effect

#### 3.1. Measurement of the investor sentiment cycle

Table 5 shows the results of the correlation analysis between the day of the week number (WI) and investor sentiment (ICI), where WI uses numbers 1–7 to represent Monday to Sunday. The ICI and WI showed a strong and highly significant correlation ( $r = 0.53141$ ,  $p < 0.00001$ ). This finding suggests that investor sentiment can differ significantly depending on the day of the week. As such, just like the day of the week has an effect on stock returns, the day of the week has a big effect on investor sentiment. Investor sentiments may have a volatile cycle.

If an investor's sentiment series is a periodic time series, it may itself be close to a sine wave. If the sentiment cycle contains a significant sine wave, a sine wave can be found using the Fourier transform (Jackson and Lacey, 2020). By using the Fourier transform to expand the time series data into linear combinations of trigonometric functions, the coefficients of each expansion term are obtained; the larger the Fourier coefficient, the more likely it is that the period of the sine wave corresponding to it is the fluctuation period of the series (Boehme and Bracewell, 1966; Goldblum et al., 1988).

Table 6 presents the most likely cycles of investor sentiment fluctuations and their amplitude values calculated by the Fourier transform, which can then be used to determine the significance of the fluctuation cycles using autoregressive coefficients (Cowan et al., 1992). Table 7 lists the autocorrelation coefficients for the three possible cycles. The analysis shows that investor sentiment shows a significant autocorrelation ( $AC = 0.627567$ ) when the period is seven days, which indicates that investor sentiment fluctuates in a cycle of 7 days. Figure 7 shows the autoregressive line plot of investor sentiment for some dates; this 7-day cycle can be clearly observed from the phase differences.

#### 3.2. The 5 + 2 cycle of investor sentiment

We use cluster analysis to examine whether there is a significant difference in investor sentiment between trading days and holiday weekends. Clustering is a type of unsupervised learning, and the k-means algorithm is a popular clustering algorithm based on error minimization (Likas et al., 2003) in which N samples are assigned to k centroids by minimizing the mean square distance from the data points to the nearest centroid (Kanungo et al., 2002). Because the k-means algorithm is based on the distance measure of spatial geometric distance, if the data gap between variables is large (e.g., PV, RV, CV, etc. in this study), the data must be normalized (Mohamad and Usman, 2013).

To explore the underlying mechanisms of investor sentiment fluctuations, a comprehensive sentiment panel consisting of five indicators—PV, RV, CV, ICI, and Agree—is clustered using the k-means clustering algorithm. The number of clusters  $k = 2$  is set to differentiate investor sentiment on trading days from that on holidays, and the cluster centroids are listed in Table 8. It can be seen that classification 0 is characterized by lower message volume, reading volume, and comment volume, with higher investor sentiment and agreement index, while classification 1 is characterized by higher message volume, reading volume, and comment volume, with lower investor sentiment and agreement index.

Table 9 provides statistics on the classification of investor sentiment on trading days and holidays for the 5-variable clustering. Among the 1094 days of the study sample, 80.08% of the trading day samples with classification 1 are characterized by a high number of messages, readings, and comments, and low investor sentiment and the Agree Index, and 93.57% of the holidays with classification 0 are characterized by a low number of messages, readings, and comments, and a high investor sentiment and agreement index. The results of the 5-variable cluster analysis show that trading days and holidays have significant differences in terms of investor sentiment. Figure 8 shows a scatter plot of investor sentiment for all data samples using the date as the x-axis and from

**Table 4.** Panel data of investor sentiment.

Panel A. The basic structure of the sentiment panel									
date	WI	PV	RV	CV	ICI	Agree			
2019-01-02	3	5,695	5,278,103	8,740	-0.304	0.011			
2019-01-03	4	4,796	4,677,579	10,445	-0.563	0.038			
2019-01-04	5	4,595	3,574,075	6,097	-0.271	0.009			
2019-01-05	6	442	1,534,157	3,492	0.053	0.000			
Panel B. Statistical description of the sentiment panel									
Indicators	MV	RV	CV	ICI	Agree				
count	1,094	1,094	1,094	1,094	1,094	1,094			
mean	2,432.53	1,729,255.00	5,433.12	0.0571	0.0136				
std	1,963.42	2,033,129.00	3,723.27	0.3279	0.0168				
min	13.00	24,426.00	22.00	-0.6825	0.0000				
25%	611.25	813,450.80	3,151.25	-0.1943	0.0018				
50%	2,245.00	1,341,684.00	4,787.50	0.0115	0.0072				
75%	3,352.75	2,085,342.00	6,625.25	0.3001	0.0192				
max	15,261.00	44,924,350.00	43,562.00	0.9967	0.1140				
Panel C. Statistical description of the stock price panel									
Indicators	P <sub>O</sub>	P <sub>C</sub>	P <sub>PC</sub>	P <sub>H</sub>	P <sub>L</sub>	P <sub>H-L</sub>	P <sub>M</sub>	V (million)	MV (million)
count	730.00	730.00	730.00	730.00	730.00	730.00	730.00	730.00	730.00
mean	3,192.43	3,195.49	3,193.92	3,213.33	3,172.42	40.90	1.57	28,474.85	34,462,820.00
std	308.15	307.52	308.18	308.67	305.66	21.23	35.25	10,542.81	14,340,630.00
min	2,446.02	2,464.36	2,464.36	2,488.48	2,440.91	11.18	-229.92	10,993.20	9,759,257.00
25%	2,922.75	2,924.13	2,923.64	2,936.88	2,905.93	25.90	-16.37	20,564.50	22,625,000.00
50%	3,233.08	3,224.45	3,223.77	3,254.13	3,209.11	35.99	1.72	27,739.89	32,850,000.00
75%	3,483.27	3,485.06	3,484.06	3,501.87	3,456.58	49.85	20.57	34,990.36	44,975,000.00
max	3,721.09	3,715.37	3,715.37	3,731.69	3,692.82	163.80	180.07	66,727.63	84,300,000.00
Panel D. Statistical description of the stock price panel on Friday									
Indicators	P <sub>O5</sub>	P <sub>C5</sub>	P <sub>L5</sub>	P <sub>H5</sub>	P <sub>M5</sub>	P <sub>H5-L5</sub>			
count	129.00	129.00	129.00	129.00	129.00	129.00	129.00	129.00	129.00
mean	3,197.99	3,200.21	3,218.76	3,177.02	-1.06	41.74			
std	311.75	309.59	310.04	309.04	35.88	19.91			
min	2,446.02	2,514.87	2,515.32	2,440.91	-136.56	15.15			
25%	2,919.26	2,927.37	2,939.34	2,902.69	-20.75	28.10			
50%	3,250.15	3,224.53	3,274.33	3,219.42	0.48	37.27			
75%	3,484.93	3,488.47	3,520.97	3,459.75	19.61	49.57			
max	3,691.19	3,703.11	3,722.87	3,681.64	95.81	134.16			
Panel E. Description of statistical indicators for the complete trading week									
Indicators	P <sub>HW</sub>	P <sub>LW</sub>	P <sub>HW-LW</sub>	P <sub>DELTA-HW-LW</sub>	P <sub>DELTA-M</sub>	P <sub>DELTA-MI</sub>			
count	129.00	129.00	129.00	129.00	129.00	129.00			
mean	3,243.67	2,926.22	317.45	206.61	4.88	126.79			
std	308.53	113.50	223.57	76.47	76.13	65.73			
min	2,574.41	2,515.51	40.55	91.00	-187.09	39.87			
25%	2,958.24	2,873.99	84.50	149.11	-33.98	81.77			
50%	3,326.46	3,000.00	326.46	193.63	5.99	107.31			
75%	3,538.01	3,000.00	538.01	234.88	49.07	141.83			
max	3,723.85	3,000.00	723.85	496.90	230.51	365.05			

Note: Only the first four rows of data are shown in Panel A. The 25%, 50%, and 75% quartiles in Panel B indicate quartiles. The unit of measurement for share prices is the RMB.

Monday to Sunday as the y-axis. Each point in the plot represents a day's sentiment, and different sentiment classifications are rendered using different colors, from which this finding can be clearly observed.

To further validate the above findings, clustering analysis was performed again using the k-means algorithm with two direct sentiment indicators, ICI and Agree; the centroids of the clustering results are shown in Table 10, where low ICI and low Agree are used to characterize classification 0, while classification 1 is characterized by high ICI and high Agree.

Table 11 shows the classification of investor sentiment on trading days and holidays for 2-variable clustering, where 79.69% of trading days are classified as 0, characterized by low investor sentiment and an agreement index, and 92.28% of holidays are classified as 1, characterized by high investor sentiment and an agreement index. In line with the results of the 5-variable cluster analysis, trading days and holidays show significant differences in investor sentiment. Figure 9 shows a scatter plot of investor sentiment data, using the date as the x-axis and Mon-Sun as the y-axis. Each point in the plot represents a day's sentiment, and

**Table 5.** Correlation analysis between investor sentiment and the day of the week.

	WI
ICI	0.53141****

Note: \*, \*\*, \*\*\*, and \*\*\*\* represent significance at the 0.1, 0.05, 0.01, and 0.001 levels, respectively.

**Table 6.** The top three fluctuation periods and their amplitude values.

periods	547 3 7
Power	73.893 77.569 111.504

**Table 7.** Autoregressive coefficients of sentiment fluctuation.

periods	AC
547	0.061686
3	-0.038586
7	0.627567

different sentiments are rendered using different colors, from which this finding can be clearly observed.

The k-means algorithm is a widely adopted unsupervised learning algorithm (Sinaga and Yang, 2020). Through cluster analysis, we conclude that investor sentiment differs significantly between holiday weekends and trading days. To avoid possible erroneous conclusions due to single-algorithm validation, we supplemented the above findings with multiple supervised learning algorithms. The 1094 statistical samples included 783 trading days and 311 holiday weekends. We use the label “0” for trading days and the label “1” for holiday weekends and randomly split the 1094 samples into two parts: a training set and a test set. We used the training set to train the machine learning model and allowed the model to learn the classification rules. The machine learning model was then used to predict the test set data. If investor sentiment changes significantly between trading days and holidays, standard machine-learning models should be able to accurately predict the labels for each group.

We used “Logistic Regression,” “Decision Tree,” “Random Forest,” “SVM” and “Boosting Strategy” to classify the test set data. We do not intend to discuss the performance advantages and disadvantages of these models. Our point is that if common machine learning models can accurately separate holiday and trading day investor sentiment, then there is a typical difference between holiday and trading day investor sentiment. This difference is a difference in the sentiment itself, rather than relying on a specific machine learning model. The classification results are listed in Table 12. The average accuracy of the 5-variable sentiment classification for all models is 92.30%, with 89.42% for the holiday sample and 93.47% for the trading-day sample. The average accuracy of the 2-variable sentiment classification was 82.57%, including 77.57% for the holiday sample and 88.02% for the trading-day sample. It

can be seen that there is a significant difference between investor sentiment on holidays and trading days.

In summary, investor sentiment on holidays and trading days differs significantly. The repeated cycle of trading days and holidays makes the 7-day cycle of investor sentiment to show a 5 + 2 fluctuation pattern. As shown in Figures 8 and 9, although this 5 + 2 fluctuation pattern is occasionally broken, it always returns to track again.

### 3.3. Analysis of investor sentiment on the trading days

The cluster analysis results show that investor sentiment is significantly different on trading days versus holidays, and stock market sentiment is more likely to influence investor sentiment on trading days than on holidays. Table 13 presents the correlation analysis between investor sentiment and stock market indices from Monday to Friday, which shows that investor sentiment has a weak but statistically significant positive correlation with the opening, closing, highest, and lowest stock prices of the stock market ( $r > 0.32$ ,  $p < 0.0002$ ), where higher stock prices lead to higher investor sentiment, showing the profit-seeking characteristics of investors. In addition, the stock market up/down amounts also show a weak but statistically significant positive correlation with investor sentiment from Monday to Thursday. However, the significance decreased significantly on Fridays ( $p = 0.09390$ ). The expectations of holidays may influence the perception of investor sentiment on up/down amounts. Correlation analysis also shows that volume and amount traded are significantly and positively correlated with investor sentiment ( $p < 0.002$ ); however, the amount traded has a stronger effect on investor sentiment than volume, whereby investors tend to dump stocks in better stock market conditions.

The difference between the highest and lowest stock prices of a day shows the magnitude of stock price volatility and, to some extent, the amount of risk in the stock. On Mondays, Tuesdays, Thursdays, and Fridays, stock price volatility showed a significant but weak negative correlation with investor sentiment, indicating risk-averse investors. However, investors do not always see this signal clearly, and on Wednesdays, this correlation becomes very weak and less significant ( $r = -0.07038$ ,  $p = 0.39531$ ).

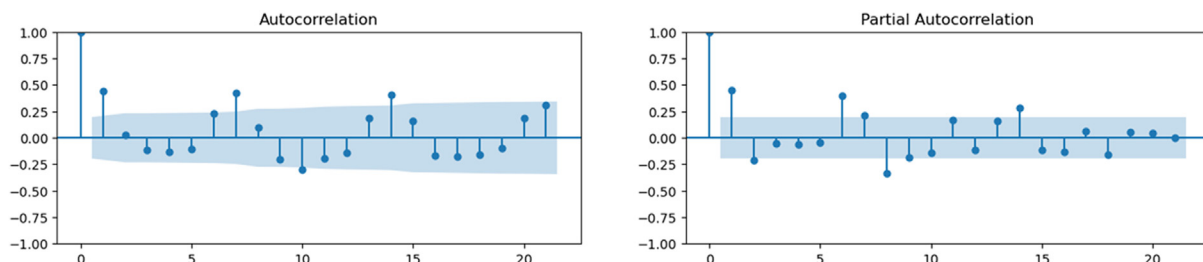
**Table 8.** Centroids of 5-variable clustering.

Classification	PV	RV	CV	ICI	Agree
0	0.05756	0.02083	0.08451	0.38934	0.02302
1	0.22975	0.05070	0.15250	-0.17245	0.00703

**Table 9.** Sample distribution statistics of 5-variable clustering.

	Total days	Classification 0	Classification 1
Trading day	783 (100%)	156 (19.92%)	627 (80.08%)
Holiday	311 (100%)	291 (93.57%)	20 (6.43%)

Note: The percentage of the classification is shown in parentheses.



**Figure 7.** Investor sentiment autoregression curve.



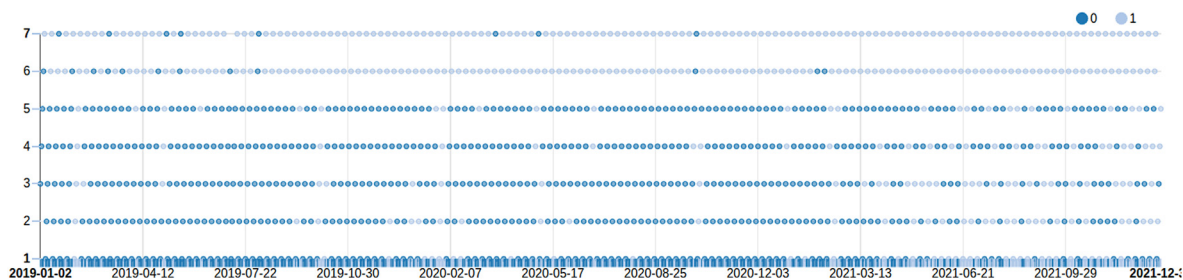


Figure 8. Scatterplot of investor sentiment classification for 5-variable clustering.

Table 10. Centroids of 2-variable clustering.

Classification	ICI	Agree
0	-0.17236	0.00701
1	0.39048	0.02308

Table 11. Sample distribution statistics of 2-variable clustering.

	Total days	Classification 0	Classification 1
Trading day	783 (100%)	624 (79.69%)	159 (20.31%)
Holiday	311 (100%)	24 (7.72%)	287 (92.28%)

Note: The percentage of the classification is shown in parentheses.

The x-axis in Figure 10 represents Monday–Friday. The five curves represent the correlation coefficients between investor sentiment and opening price, closing price, previous trading day's closing price, highest stock price, and lowest stock price, respectively, from which it can be seen that the fluctuation patterns of the five indicators of investor sentiment from Monday to Friday are consistent. The impact of the stock market on investor sentiment varies depending on the day of the week, with the lowest impact on Tuesdays and the highest impact on Thursdays.

In summary, investor sentiment is affected by stock market fluctuations on trading days, showing profit-seeking and risk-averse preferences, reflecting the basic characteristics of “rational people,” but in the five days, the impact of the stock market on investment sentiment will vary depending on the day of the week, there is a day of the week effect of investors' perception of the stock market. This is very similar to the difference in employees' productivity, with double days off depending on

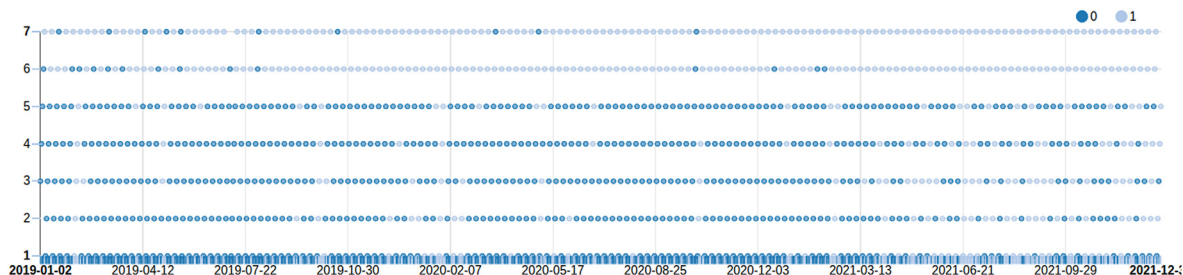


Figure 9. Scatterplot of investor sentiment classification for 2-variable clustering.

Table 12. Statistics of supervised learning classification results.

		Logistic Regression	Decision Tree	Random Forest	SVM	Boosting Strategy	Mean
Classification of 5 Feature Variables	Total	93.87%	90.89%	94.42%	91.45%	90.89%	92.30%
	Holiday	93.55%	84.52%	96.13%	88.39%	84.52%	89.42%
	Trading day	93.99%	93.47%	93.73%	92.69%	93.47%	93.47%
Classification of 2 Feature Variables	Total	76.21%	86.62%	86.99%	76.39%	86.62%	82.57%
	Holiday	67.89%	73.55%	75.48%	97.42%	73.55%	77.58%
	Trading day	96.77%	91.91%	91.64%	67.89%	91.91%	88.02%

Table 13. Correlation analysis between investor sentiment and stock market fundamentals on trading days.

	P <sub>O</sub>	P <sub>C</sub>	P <sub>PC</sub>	P <sub>H</sub>	P <sub>L</sub>	P <sub>H.L</sub>	P <sub>M</sub>	V	MV
ICI <sub>1</sub>	0.3805****	0.4354****	0.3770****	0.3937****	0.4165****	-0.2295***	0.3891***	0.2569***	0.3880***
ICI <sub>2</sub>	0.3297****	0.3614****	0.3345****	0.3355****	0.3547****	-0.1998***	0.2284****	0.2583****	0.3303****
ICI <sub>3</sub>	0.4238****	0.4568****	0.4325****	0.4358****	0.4462****		0.2869****	0.3688****	0.4458****
ICI <sub>4</sub>	0.4582****	0.4967****	0.4557****	0.4688****	0.4851****	-0.2001***	0.4290****	0.3783****	0.4508****
ICI <sub>5</sub>	0.4344****	0.4632****	0.4417****	0.4409****	0.4556****	-0.2050***	0.1406 *	0.3644****	0.4333****

Note: \*, \*\*, \*\*\*, and \*\*\*\* represent significance at the 0.1, 0.05, 0.01, and 0.001 levels, respectively. Only the significant data are presented.

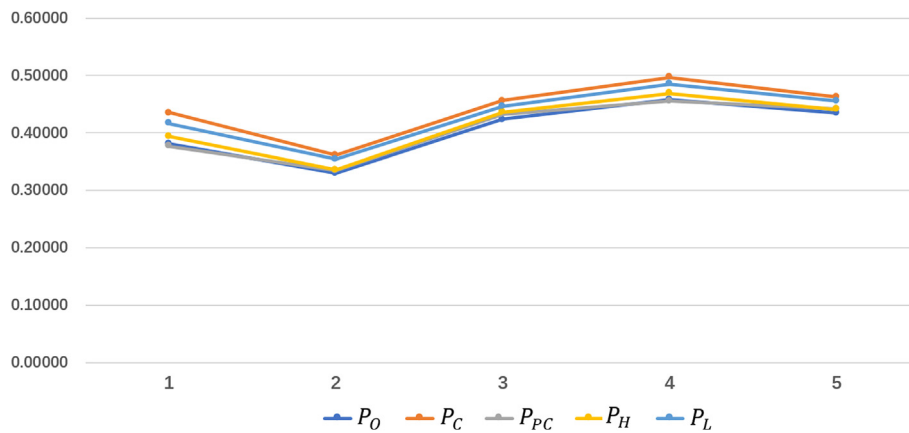


Figure 10. Correlation between investor sentiment and stock market indices on trading days.

the day of the week (Bryson and Forth, 2007), where people are unable to maintain consistent productivity over a longer period (Chung, 2022).

### 3.4. Analysis of investor sentiment over the holiday weekend

Cluster analysis shows that during the holiday weekend, which lacks direct stimulus from stock market conditions, investor sentiment is significantly different from that on the trading day, showing higher investor sentiment and agreement in opinions. To gain insight into the weekend characteristics of investor sentiment, we aggregate the data of 5 trading days and 2-day weekends in a trading cycle into a single message, construct a new sentiment panel data in a group of 7 days, and obtain a total of 129 complete “5 + 2” trading weeks from 2019-01-01 to 2021-12-31.

#### 3.4.1. Holiday effect

Figure 11 shows the scatter plot of investor sentiment for 1094 days, with the X-axis representing Monday to Sunday and the Y-axis representing investor sentiment. **It is clear that investor sentiment is significantly higher on weekends than on Monday to Friday, consistent with the finding of higher investor sentiment on weekends in the cluster analysis.**

Table 14 shows the highest, lowest, average, and fluctuation in investor sentiment for each day of the week. The fluctuation is obtained by subtracting the lowest value from the highest value, and the larger the fluctuation, the more unstable investor sentiment is on that day of the week. By contrast, the opposite indicates that investor sentiment is more stable on that day. The statistical results show that investor sentiment is significantly raised on weekends, especially Saturdays, when its mean value increases from -0.1087 on Fridays to 0.3844, with only Saturdays and Sundays of the week showing positive mean values of investor sentiment. In addition, the minimum value of daily sentiment also shows a significant increase on Saturdays and Sundays, and  $ICI_{MAX-MIN}$  is significantly lower on weekends, indicating that investor sentiment volatility is significantly lower during holidays without stock market stimulation, and the lowest value of volatility is reached on Sundays, where investor sentiment is most stable. In Figure 12, the x-axis is Monday to Sunday, the y-axis is investor sentiment, and the three curves from top to bottom represent the highest, mean, and lowest sentiment values for the days of the week, from which we can review the above conclusions more visually.

Table 15 shows the statistical description of  $ICI_{6.5}$  and  $ICI_{7.6}$  for 129 complete trading weeks. A positive  $ICI_{6.5}$  indicates that investor sentiment was higher on Saturdays than on Fridays for that week. A positive  $ICI_{7.6}$  indicates that investor sentiment was higher on Sundays than Saturdays for that week. Among all 129 trading weeks, 128 weeks had positive  $ICI_{6.5}$ , with only one exception, reaching 99.225% of the total

sample, and 85 trading weeks had positive  $ICI_{7.6}$ , accounting for 65.891% of the total sample.

In summary, during holidays without stock market stimuli, investor sentiment is significantly higher, and the amplitude of sentiment is significantly narrower, showing a typical holiday effect. More specifically, investor sentiment is significantly higher with a narrower amplitude on Saturdays, investor sentiment is more likely to be higher, and the amplitude of sentiment is further narrowed to a minimum on Sundays. The holiday effect on investor sentiment confirms Lakonishok and Levi's (1982) finding that holidays play an important role in the “day of the week effect.”

#### 3.4.2. Correlation analysis between holiday effect and stock market conditions

Suppose that stock market conditions correlate with the holiday effect. The stock market conditions on Fridays closest to the holiday and the current week's stock market fundamentals will impact the holiday effect and vice versa, indicating that stock market conditions do not influence the holiday effect. Table 16 shows the correlation analysis of the magnitude of Saturday's investor sentiment increase with Friday's stock market conditions and fundamentals for that week. Panel A shows that Saturday's sentiment increase is entirely unaffected by Friday's stock market, and is not correlated with Friday's opening price, closing price, lowest stock price, highest stock price, stock price volatility, or up/down amounts ( $p > 0.66$ ). Panel B shows that the increase in sentiment on Saturdays is also not correlated with the highest stock price, lowest stock price, fluctuation, cumulative fluctuation, cumulative up and down, and cumulative absolute up and down of the stock market for that week (cumulative absolute up and down  $p = 0.19658$ , other variables  $p > 0.41$ ). On the first day of the holiday, investor sentiment showed an “irrational” increase, which was unaffected by stock market fluctuations. The investors devoted themselves entirely to the holiday, showing the characteristics of the early stage of the holiday.

Table 17 analyzes the correlation of Sunday's sentiment increase with Friday's stock market movements and fundamentals for that week. Similar to Saturday, Sunday's sentiment increase is also unaffected by Friday's stock market and does not correlate significantly ( $p > 0.24$ ) with Friday's opening price, closing price, minimum stock price, maximum stock price, volatility, or up/down the amount, and similarly does not correlate significantly ( $p > 0.3$ ) with that week's maximum stock price, minimum stock price, volatility, cumulative volatility, cumulative up/down amounts, or cumulative absolute up/down amounts.

Combining the results of the analyses in Tables 15 and 16, Saturday's sentiment increase shows a complete non-correlation with Friday's stock market quotes ( $p > 0.66$ ). However, the correlation between Sunday's sentiment increase and Friday's opening and closing prices, minimum price, maximum stock price, and volatility tended to be more significant

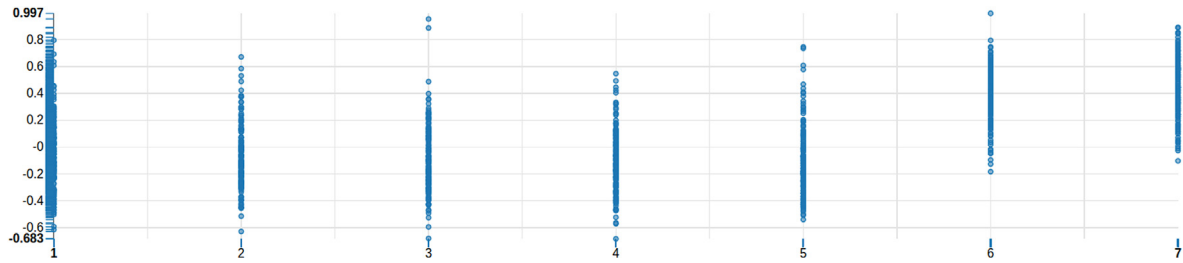


Figure 11. Scatterplot of investor sentiment distribution by day of the week.

Table 14. Statistical description of investor sentiment distribution by day of the week.

	1	2	3	4	5	6	7
ICI <sub>MAX</sub>	0.7965	0.6721	0.9549	0.5475	0.7470	0.9967	0.8926
ICI <sub>MEAN</sub>	-0.0629	-0.0774	-0.0842	-0.0990	-0.1087	0.3844	0.4529
ICI <sub>MIN</sub>	-0.6139	-0.6279	-0.6796	-0.6825	-0.5392	-0.1823	-0.1018
ICI <sub>MAX-MIN</sub>	1.4104	1.3000	1.6345	1.2300	1.2862	1.1790	0.9944

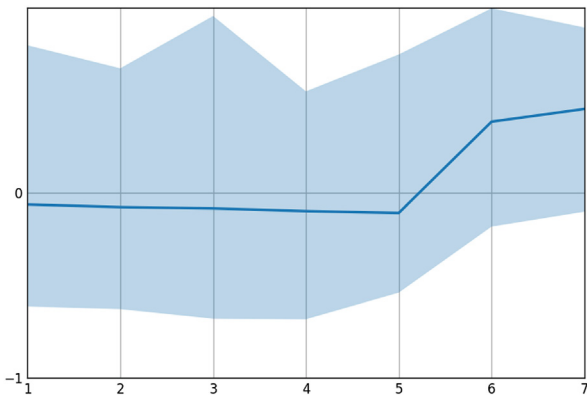


Figure 12. Graph of investor sentiment trends by day of the week.

Table 15. Weekend investor sentiment movement statistics.

	Positive sample size	Negative sample size	Positive sample ratio (%)
ICI <sub>6-5</sub>	128	1	99.225
ICI <sub>7-6</sub>	85	44	65.891

than on Saturdays ( $p > 0.24$ ). Similarly, the correlation between Sunday sentiment increases and the highest stock price ( $r = 0.08760$ ,  $p = 0.32355$ ) and the lowest stock price ( $r = 0.10386$ ,  $p = 0.24147$ ) for that week is also somewhat more significant relative to Saturday.

The holiday effect of investor sentiment is not correlated with the stock market conditions of a week, showing typical “social” characteristics. More specifically, investors show a typical early holiday effect on Saturdays, with a significant increase in sentiment (99.225%), which is not affected by the stock market sentiment of that week; investors remain in the holiday effect on Sundays, with a high probability of an increase in investor sentiment (65.891%). However, towards the end of the holiday, investor sentiment gradually “returns to rationality,” making the elevated investor sentiment present more significant uncertainty.

In summary, investor sentiment cycles repeatedly over a 7-day cycle. Investor sentiment on trading days is significantly correlated with stock market fundamentals, while elevated investor sentiment on holiday weekends is not correlated with stock market fundamentals. Investor sentiment is significantly different on trading days and holiday weekends, giving a 7-day cycle a 5 + 2 volatility pattern. The cyclical stimulus of trading days and holidays leads to 5 + 2 fluctuations in investor sentiment. The above findings support Hypotheses H1, H2, and H3. The regular opening and closing of the stock market may be a potential hypothesis for the “day of the week effect.”

Table 16. Correlation analysis between Saturday's sentiment increase and stock market indexes.

Panel A	P <sub>o5</sub>	P <sub>C5</sub>	P <sub>L5</sub>	P <sub>H5</sub>	P <sub>SH-SL</sub>	P <sub>M5</sub>
ICI <sub>6-5</sub>	-0.03139 (0.72395)	-0.03015 (0.73445)	-0.02777 (0.75478)	-0.03018 (0.73420)	-0.03852 (0.66471)	0.01330 (0.88112)
Panel B	P <sub>HW</sub>	P <sub>LW</sub>	P <sub>HW-LW</sub>	P <sub>DELTA-HW-LW</sub>	P <sub>DELTA-M</sub>	P <sub>DELTA- M </sub>
ICI <sub>6-5</sub>	-0.03415 (0.70081)	0.04845 (0.58557)	-0.07173 (0.41923)	-0.05806 (0.51338)	-0.05296 (0.55114)	-0.11444 (0.19658)

Note: Data in parentheses are p-values corresponding to r-values.

Table 17. Correlation analysis of sundays' sentiment enhancement and stock market indexes.

Panel A	P <sub>o5</sub>	P <sub>C5</sub>	P <sub>L5</sub>	P <sub>H5</sub>	P <sub>SH-SL</sub>	P <sub>M5</sub>
ICI <sub>7-6</sub>	0.09851 (0.26671)	0.09255 (0.29684)	0.09177 (0.30095)	0.09821 (0.26816)	0.10406 (0.24056)	-0.06203 (0.48495)
Panel B	P <sub>HW</sub>	P <sub>LW</sub>	P <sub>HW-LW</sub>	P <sub>DELTA-HW-LW</sub>	P <sub>DELTA-M</sub>	P <sub>DELTA- M </sub>
ICI <sub>7-6</sub>	0.08760 (0.32355)	0.10386 (0.24147)	0.06816 (0.44276)	-0.09058 (0.30731)	-0.03685 (0.67846)	-0.08713 (0.32619)

Note: Data in parentheses are p-values corresponding to r-values.

#### 4. Analysis of atypical trading cycles

In the stock market, in addition to the typical 5 + 2 trading cycle, there are also some special trading cycles where the weekly holidays range from 3-to to seven days. The special trading cycle may be caused by a “holiday” market closure or a market closure caused by a stock market abnormality such as a sustained downward or upward movement. This section supports the conclusions of the previous section with an analysis of investor sentiment during atypical trading cycles.

For the period from 2019-01-01 to 2021-12-31 examined in this study, there were 24 weeks of atypical trading cycles on the Shanghai Stock Exchange with a total of 168 sample days, with a minimum of three days and a maximum of seven days of holidays per week. Table 18 provides statistics on the distribution of trading days and holidays for 24 atypical trading cycles.

Table 19 lists the number of trading days and holiday days in the atypical trading cycle with a total sample size of 168, of which 120 days are from Monday to Friday and 48 days from Saturday to Sunday. Mon-Fri contains 68 trading days and 52 holiday days; all Saturdays and Sundays are holidays.

To determine whether investor sentiment is significantly different between holidays and trading days in atypical trading interim periods, a k-means clustering algorithm was used to perform a cluster analysis on 168 study samples with five indicators: PV, RV, CV, ICI, and Agree, and a cluster number  $k = 2$ . Table 20 shows the centroids of the cluster analysis, from which it can be seen that classification 0 is characterized by higher PV, RV, and CV, whereas ICI and Agree are lower. In contrast to classification 0, classification 1 had lower PV, RV, and CV but higher ICI and Agree values.

Table 21 shows that during the atypical trading cycle, the Monday to Friday sample is discrete and cannot be classified into one classification, with 55% of the sample falling into classification 0 and 45% falling into classification 1. It indicates that when Monday to Friday is no longer a full trading day, investor sentiment cannot be statistically classified into one category, but 91.7% of the weekends fall into classification 1. Even in atypical trading cycles, investor sentiment can still be significantly put into one category during holiday weekends.

The statistical results in Table 21 also show that 92.6% of the trading-day samples belong to category 0 and are characterized by high message

Table 18. Statistics on the distribution of holidays in atypical trading cycles.

Holidays Distribution							Days of holiday	Number of samples	Percentage
1	2	3	4	5	6	7	3	7	29.2%
1	2	3	4	5	6	7	3	4	16.7%
1	2	3	4	5	6	7	4	3	12.5%
1	2	3	4	5	6	7	4	2	8.3%
1	2	3	4	5	6	7	5	1	4.2%
1	2	3	4	5	6	7	5	2	8.3%
1	2	3	4	5	6	7	6	1	4.2%
1	2	3	4	5	6	7	6	2	8.3%
1	2	3	4	5	6	7	7	2	8.3%

White padding represents trading days, light bold padding represents holidays. The numbers 1–7 in the holiday distribution represent Mon-Sun.

Table 19. Statistic of holidays and trading days in the atypical trading cycle.

Day of the week distribution	Total days	Trading days	Holiday days
1–5	120	68	52
6–7	48	0	48
合计act	168	68	100

1-5 refers to Mon-Fri, and 6–7 refers to Sat-Sun.

Table 20. Clustering centroids of 5-variable clusters for atypical trading cycles.

Classification	PV	RV	CV	ICI	Agree
0	0.31586	0.11599	0.14914	-0.12142	0.00535
1	0.05822	0.05353	0.07383	0.39759	0.02399

Table 21. Sample classification statistics for 5-variable clustering of atypical trading cycles.

	Total	Category 0	Category 1
1–5	120 (100%)	66 (55.0%)	54 (45.0%)
6–7	48 (100%)	4 (8.3%)	44 (91.7%)
Trading day	68 (100%)	63 (92.6%)	5 (7.4%)
Holiday	100(100%)	7 (7.0%)	93 (93.0%)

The data in parentheses are the percentage of the sample.

1-5 represents Monday to Friday, and 6–7 represents Saturday to Sunday.

Table 22. Centroids of 2-variable clustering for atypical trading cycles.

Classification	ICI	Agree
0	0.44269	0.02764
1	-0.08002	0.00481

volume, reading volume, and comment volume. However, investor sentiment and sentiment index are low. 93.0 Of the holiday samples, 93.0% are characterized by low message volume, reading volume, and comment volume but high investor sentiment and sentiment index. During the atypical trading cycle, investor sentiment is still clearly distinguishable between trading days and holidays, and their characteristics are consistent with the 5 + 2 typical trading cycle.

To corroborate the above findings, a cluster analysis was performed using the k-means clustering algorithm for the direct emotion indicators ICI and Agree, with the number of clusters  $k = 2$ . Table 22 shows the centroids of the cluster analysis, with higher characteristics of ICI and Agree for classification 0, and the opposite for classification 1. Table 23 lists the sample classification of the 2-variable clustering. 94.1 Of the trading days, 94.1% are classified as 1, and 80.0% of holidays are classified as 0, which indicates that consistent with the results of the 5-variable clustering, trading days and holidays can be clearly distinguished in atypical trading cycles. However, Monday through Friday cannot be clustered into one category because of the mixing of trading days and holidays, consistent with typical trading weeks: during trading days, the investor sentiment and agreement index is low, and on holidays, the investor sentiment and agreement index is high.

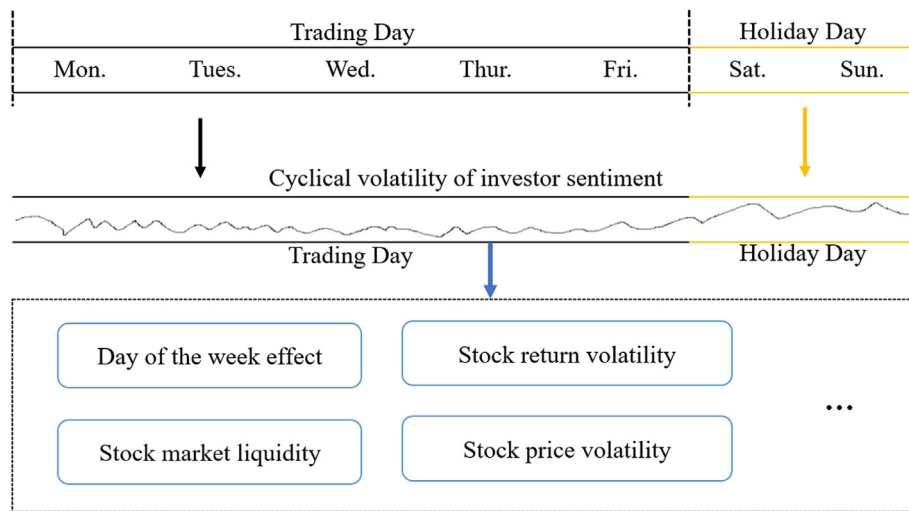
In summary, in the atypical trading cycle, there is a statistically significant difference between investor sentiment on holidays and trading days, with lower investor sentiment and agreement of opinion on trading days and higher investor sentiment and agreement of opinion on holidays, which is consistent with the typical 5 + 2 trading cycle findings. However, there is no statistical consistency in investor sentiment from Monday to Friday, which is different from the typical trading cycle,

Table 23. Sample classification statistics for 2-variable clustering of atypical trading cycles.

	Total	Category 0	Category 1
1–5	120 (100%)	44 (36.7%)	76 (63.3%)
6–7	48 (100%)	40 (83.3%)	8 (16.7%)
Trading day	68 (100%)	4 (5.9%)	64 (94.1%)
Holiday	100 (100%)	80 (80.0%)	20 (20.0%)

The data in parentheses are the percentage of the sample.

1-5 represents Monday to Friday, and 6–7 represents Saturday to Sunday.



**Figure 13.** Diagram of the trading cycle driving the sentiment cycle.

suggesting that it is not the day of the week that affects investor sentiment trading cycles, but whether it is a holiday or a holiday trading day.

Many studies suggest that holiday weekends and investor sentiment may explain the day-of-week effect (Chiah and Zhong, 2019, 2021; Lakonishok and Levi, 1982). Our findings complement these studies, as we show that any holiday can lead to changes in investor sentiment. Gibbons and Hess (1981) studied the Monday effect for 11 indices from nine countries from 1969 to 1992 and found lower returns at the beginning of the week (but not necessarily on Mondays) throughout the cycle. Our study supports Gibbons and Hess's findings. The analysis of investor sentiment for atypical trading cycles corroborates the findings in Section 3 that the repeated cycle of trading days and holidays is directly responsible for constituting investor sentiment in a 5 + 2 cycle.

## 5. Conclusion

By using natural language processing technology, we classified the sentiment of more than 2.66 million messages in the SSE stock bar from 2019-01-01 to 2021-12-31. This is the basis for constructing the investor confidence index (ICI) as a sentiment proxy and combining investor sentiment data and stock market quotes to analyze the fluctuation pattern of investor sentiment and the basic factors affecting sentiment fluctuation. The findings show that investor sentiment in the stock market has a 7-day cycle. In the 5-day trading days, investors show "rational people's" characteristics with profit-taking and risk-averse preferences. During the 2-day holiday, investors show the characteristics of "social people," investor sentiment is significantly higher, and the fluctuations narrow; the magnitude of this increase is not affected by stock market volatility, showing the holiday effect.

Numerous scholars have verified the association between investor sentiment and economic phenomena such as "stock market returns," "stock prices," and "stock market liquidity" (Al-Hajjeh et al., 2011; Teng and Liu, 2013; Thaler, 1987), particularly the "day of the week effect" in the stock market (Gibbons and Hess, 1981; Lakonishok and Levi, 1982). Our findings corroborate these conclusions to some extent. As shown in Figure 13, we find that in a typical trading cycle (Mon-Fri trading days and Sat-Sun holidays), the cyclical stimulation of trading days and holidays leads to 5 + 2 fluctuations in investor sentiment. A typical trading cycle pattern is a potential explanation for many economic phenomena related to investor sentiment.

A trading cycle of five trading days and two holidays is the primary cause of investor sentiment fluctuations. Complex external causes may remove investor sentiment from the basic volatility cycle, but when the external forces disappear, investor sentiment returns to the basic

volatility once again. The cycle of trading days and holidays drives investor sentiment to vary depending on the day of the week, which in turn drives the effect of the day of the week on stock market returns. This is a comprehensive explanation from researchers on the interpretation of holidays and investor sentiment for the DOW effect.

In this study, we use social media data from the Chinese market to study the fluctuation patterns of investor sentiment and the impact of holidays on investor sentiment. Our findings suggest that holidays have an important impact on investor sentiment and that investment sentiment is significantly higher during holidays, which has implications for understanding investor sentiment and the design of stock market systems. As the vast majority of investments in the stock bar are made by individual investors, our analysis of the pattern of investment sentiment fluctuations may only apply to individual investors. In addition, different countries and social cultures may have different online cultures, which in turn may lead to differences in the way investors express their emotions on social media. Future research could use a larger sample size to verify the applicability of the findings of this study to more international markets.

## Declarations

### Author contribution statement

Qing Liu: Analyzed and interpreted the data; Wrote the paper.  
Xinyuan Wang: Conceived and designed the experiments.  
Yamin Du: Conceived and designed the experiments.

### Funding statement

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

### Data availability statement

Data will be made available on request.

### Declaration of interest's statement

The authors declare no competing interests.

### Additional information

No additional information is available for this paper.

## References

- Al-Hajieh, H., Redhead, K., Rodgers, T., 2011. Investor sentiment and calendar anomaly effects: a case study of the impact of Ramadan on Islamic Middle Eastern markets. *Res. Int. Bus. Finance* 25 (3), 345–356.
- Antweiler, W., Frank, M.Z., 2004. Is all that talk just noise? The information content of internet stock message boards. *J. Finance* 59 (3), 1259–1294.
- Arbieu, U., Helsper, K., Dadvar, M., Mueller, T., Niamir, A., 2021. Natural language processing as a tool to evaluate emotions in conservation conflicts. *Biol. Conserv.* 256, 109030.
- Baker, M., Wurgler, J., 2007. Investor sentiment in the stock market. *J. Econ. Perspect.* 21 (2), 129–152.
- Berument, H., Kiyamaz, H., 2001. The day of the week effect on stock market volatility. *J. Econ. Finance* 25 (2), 181–193.
- Boehme, T., Bracewell, R., 1966. *The Fourier Transform and its Applications*.
- Brown, G.W., Cliff, M.T., 2004. Investor sentiment and the near-term stock market. *J. Empir. Finance* 11 (1), 1–27.
- Brown, G.W., Cliff, M.T., 2005. Investor sentiment and asset valuation. *J. Bus.* 78 (2), 405–440.
- Bryson, A., Forth, J., 2007. Productivity and days of the week (No. 4963). In: London School of Economics and Political Science. LSE Library.
- Chiah, M., Zhong, A., 2019. Day-of-the-week effect in anomaly returns: international evidence. *Econ. Lett.* 182, 90–92.
- Chiah, M., Zhong, A., 2021. Tuesday Blues and the day-of-the-week effect in stock returns. *J. Bank. Finance* 133, 106243.
- Chung, H., 2022. A Social Policy case for a four-day week. *J. Soc. Pol.* 1–16.
- Cowan, M.J., Burr, R.L., Narayanan, S.B., Buzaitis, A., Strasser, M., Busch, S., 1992. Comparison of autoregression and fast Fourier transform techniques for power spectral analysis of heart period variability of persons with sudden cardiac arrest before and after therapy to increase heart period variability. *J. Electrocardiol.* 25, 234–239.
- Da, Z., Engelberg, J., Gao, P., 2011. In search of attention. *J. Finance* 66 (5), 1461–1499.
- Da, Z., Engelberg, J., Gao, P., 2015. The sum of all FEARS investor sentiment and asset prices. *Rev. Financ. Stud.* 28 (1), 1–32.
- De Long, J.B., Shleifer, A., Summers, L.H., Waldmann, R.J., 1990. Noise trader risk in financial markets. *J. Polit. Econ.* 98 (4), 703–738.
- Fama, E.F., 1965. The behavior of stock-market prices. *J. Bus.* 38 (1), 34–105.
- Fisher, K.L., Statman, M., 2000. Investor sentiment and stock returns. *Financ. Anal. J.* 56 (2), 16–23.
- Gibbons, M.R., Hess, P.J., 1981. Day of the Week Effects and Asset Returns.
- Goldberg, Y., Levy, O., 2014. word2vec Explained: Deriving Mikolov et al.'s negative-sampling word-embedding method. *ArXiv*.
- Goldblum, C.E., Ritter, R.C., Gillies, G.T., 1988. Using the fast Fourier transform to determine the period of a physical oscillator with precision. *Rev. Sci. Instrum.* 59 (5), 778–782.
- Hirschberg, J., Manning, C.D., 2015. Advances in natural language processing. *Science* 349 (6245), 261–266.
- Hong, H., Wang, J., 2000. Trading and returns under periodic market closures. *J. Finance* 55 (1), 297–354.
- Jackson, A.C., Lacey, S., 2020. The discrete Fourier transformation for seasonality and anomaly detection of an application to rare data. *Data Technologies and Applications*.
- Kanungo, T., Mount, D.M., Netanyahu, N.S., Piatko, C.D., Silverman, R., Wu, A.Y., 2002. An efficient k-means clustering algorithm: analysis and implementation. *IEEE Trans. Pattern Anal. Mach. Intell.* 24 (7), 881–892.
- Kim, K., Ryu, D., 2022. Sentiment Changes and the Monday Effect. *Finance Research Letters*.
- Lakonishok, J., Levi, M., 1982. Weekend effects on stock returns: a note. *J. Finance* 37 (3), 883–889.
- Lee, C.M., Shleifer, A., Thaler, R.H., 1991. Investor sentiment and the closed-end fund puzzle. *J. Finance* 46 (1), 75–109.
- Lemmon, M., Portniaguina, E., 2006. Consumer confidence and asset prices: some empirical evidence. *Rev. Financ. Stud.* 19 (4), 1499–1529.
- Likas, A., Vlasis, N., Verbeek, J.J., 2003. The global k-means clustering algorithm. *Pattern Recogn.* 36 (2), 451–461.
- Liu, Q., Zhou, X., Zhao, L., 2022. View on the bullishness index and agreement index. *Front. Psychol.* 13.
- Liu, S., 2015. Investor sentiment and stock market liquidity. *J. Behav. Finance* 16 (1), 51–67.
- Ljungqvist, A., Nanda, V., Singh, R., 2006. Hot markets, investor sentiment, and IPO pricing. *J. Bus.* 79 (4), 1667–1702.
- López-Cabarcos, M.Á., Pérez-Pico, A.M., Piñero-Chousa, J., Šević, A., 2021. Bitcoin volatility, stock market and investor sentiment. Are they connected? *Finance Res. Lett.* 38, 101399.
- López-Cabarcos, M.Á., Piñero-Chousa, J., Pérez-Pico, A.M., 2017. The impact technical and non-technical investors have on the stock market: evidence from the sentiment extracted from social networks. *Journal of Behavioral and Experimental Finance* 15, 15–20.
- McGurk, Z., Nowak, A., Hall, J.C., 2020. Stock returns and investor sentiment: textual analysis and social media. *J. Econ. Finance* 44 (3), 458–485.
- Mohamad, I.B., Usman, D., 2013. Standardization and its effects on K-means clustering algorithm. *Res. J. Appl. Sci. Eng. Technol.* 6 (17), 3299–3303.
- Naseem, S., Mohsin, M., Hui, W., Liyan, G., Penglai, K., 2021. The investor psychology and stock market behavior during the initial era of COVID-19: a study of China, Japan, and the United States. *Front. Psychol.* 12, 16.
- Neal, R., Wheatley, S.M., 1998. Do measures of investor sentiment predict returns? *J. Financ. Quant. Anal.* 33 (4), 523–547.
- Qadan, M., Nama, H., 2018. Investor sentiment and the price of oil. *Energy Econ.* 69, 42–58.
- Qiu, L., Welch, I., 2004. Investor Sentiment Measures (No. 10794). National Bureau of Economic Research, Inc.
- Sayim, M., Rahman, H., 2015. The relationship between individual investor sentiment, stock return and volatility: evidence from the Turkish market. *Int. J. Emerg. Mark.* 10 (3), 504–520.
- Schadner, W., 2021. On the persistence of market sentiment: a multifractal fluctuation analysis. *Phys. Stat. Mech. Appl.* 581, 126242.
- Sinaga, K.P., Yang, M.-S., 2020. Unsupervised K-means clustering algorithm. *IEEE Access* 8, 80716–80727.
- Singer, E., 2002. The use of incentives to reduce nonresponse in household surveys. *Survey nonresponse* 51 (1), 163–177.
- Song, Y., Ji, Q., Du, Y.-J., Geng, J.-B., 2019. The dynamic dependence of fossil energy, investor sentiment and renewable energy stock markets. *Energy Econ.* 84, 104564.
- Sprenger, T.O., Tumasjan, A., Sandner, P.G., Welpe, I.M., 2014. Tweets and trades: the information content of stock microblogs. *Eur. Financ. Manag.* 20 (5), 926–957.
- Teng, C.C., Liu, V.W., 2013. The pre-holiday effect and positive emotion in the Taiwan Stock Market, 1971–2011. *Invest. Anal. J.* 42 (77), 35–43.
- Thaler, R.H., 1987. Anomalies: weekend, holiday, turn of the month, and intraday effects. *J. Econ. Perspect.* 1 (2), 169–177.
- Wang, G.-J., Xiong, L., Zhu, Y., Xie, C., Foglia, M., 2022. Multilayer network analysis of investor sentiment and stock returns. *Res. Int. Bus. Finance* 62, 101707.
- Xiong, X., Chunchun, L.U.O., Ye, Z., 2017. Stock BBS and trades: the information content of stock BBS. *J. Syst. Sci. Math. Sci.* 37 (12), 2359.
- Zouaoui, M., Nouyrigat, G., Beer, F., 2011. How does investor sentiment affect stock market crises? Evidence from panel data. *Financ. Rev.* 46 (4), 723–747.