Available online at www.sciencedirect.com

Borsa İstanbul Review

Borsa İstanbul Review 23-1 (2023) 76–92
http://www.elsevier.com/journals/borsa-istanbul-review/2214-8450

Full Length Article

Synergy between stock prices and investor sentiment in social media

Qing Liu [a,c], Woon-Seek Lee [c], Minghao Huang [c], Qingjun Wu [b,*]

[a] College of Economics and Management, Huainan Normal University, Huainan, China
[b] Shanghai Lixin University of Accounting and Finance, Shanghai, China
[c] Graduate School of Management of Technology, Pukyong National University, Busan, South Korea

Received 25 June 2022; revised 8 September 2022; accepted 10 September 2022
Available online 24 September 2022

Abstract

The underlying assumption of using investor sentiment to predict stock prices, stock market returns, and liquidity is that of synergy between stock prices and investor sentiment. However, this synergistic relationship has received little attention in the literature. This paper investigates the synergistic pattern between stock prices and investor sentiment using social media messages from stock market investors and natural language processing techniques. At the macro level, we reveal extremely significant positive synergy between investor sentiment and stock prices. That is, when a stock price rises, investor sentiment rises, and when a stock price falls, investor sentiment falls. However, this synergy may be reversed or even disappear over a specific time period. Through a segmented measurement of the synergy between stock prices and investor sentiment over the course of a day, we also find that investor sentiment on social media is forward looking. This provides theoretical support for using investor sentiment in stock price prediction. We also examine the effect of lockdowns, the most draconian response to COVID-19, on synergy between stock prices and investor sentiment through causal inference machine learning. Our analysis shows that external anxiety can significantly affect synergy between stock prices and investor sentiment, but this effect can promote either positive or negative synergy. This paper offers a new perspective on stock price forecasting, investor sentiment, behavioral finance, and the impact of COVID-19 on the stock markets.

JEL classification: G40; G41

Keywords: COVID-19; Investor sentiment; Lockdowns; Social media; Synergy

1. Introduction

Numerous studies have shown that investor sentiment can be used to forecast stock returns (Brown & Cliff, 2004; Finter et al., 2012; Kim et al., 2019; Lee et al., 2002). In addition, in the literature, investor sentiment is linked to a variety of financial phenomena, including “the day of the week effect” (Berument & Kiymaz, 2001), “asset pricing” (Brown & Cliff, 2005; Ljungqvist et al., 2006), “stock market liquidity” (Liu, 2015), and even “stock market crises” (Liu, 2015). All of these studies are predicated on the hypothesis that interaction exists between the stock market and investor sentiment, that is, a synergistic effect exists between the stock market and investor sentiment. However, the relationship between stock prices and investor sentiment is seldom discussed in academic papers. This paper analyzes and confirms the synergistic effect between stock prices and investor sentiment, using natural language processing techniques to quantify investor sentiment embedded in social media. This research has significant implications for the forecasting of stock returns and the interpretation of numerous financial phenomena.

Sentiment is complex and difficult to measure, and Baker (Baker & Wurgler, 2007) states: “The question is no longer whether investor sentiment affects stock prices, but how to measure investor sentiment and quantify its effects.”

* Corresponding author. No. 995, Shangchuan Road, Caolu Town, Pudong, Shanghai, China.
E-mail address: wuqingjun888@139.com (Q. Wu).
Peer review under responsibility of Borsa İstanbul Anonim Şirketi.

https://doi.org/10.1016/j.bir.2022.09.006

Traditional sentiment proxy approaches can be divided into three main categories: those based on market indicators, those based on survey indices, and those based on 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. Sentiment proxies based on survey indices quantify investor sentiment by collecting investors' optimistic or pessimistic expectations about the stock market through surveys, for example, the Consumer Confidence Index (Brown & Cliff, 2005), the UBS/GALLUP Investor Optimism Index (Lemmon & Portniaguina, 2006), and the Investment Newsletter (Qiu & Welch, 2004). The approach based on special events, often using special social events as sentiment proxies, of which COVID-19 is perhaps the most representative current example. Naseem et al. (2021) analyze the impact of COVID-19 on investor psychology and the impact of this impact on the stock market. The basic characteristics of these three sentiment proxy approaches are detailed in the literature (Baker & Wurgler, 2007; Hu et al., 2021).

Because of limitations in data acquisition, investor sentiment obtained by these methods has a certain lag and cannot be measured in real time for investor sentiment, which makes it difficult to apply these methods to real-time research on investor sentiment and high-frequency forecasting of the stock market. Natural language processing (NLP) technology offers a new opportunity for quantifying investor sentiment. NLP can tap into investor sentiment embedded in text and social networks and has the advantages of easy data availability, real-time access, and high credibility, providing a research basis for basing investor sentiment research on social media.

Social media plays a central role in bridging the "information asymmetry" between markets and investors (Ali, 2018). Based on an analysis of over 1.5 million messages on Yahoo!, Antweiler and Frank (2004) conclude that social media messages help predict market volatility and that their effect on stock returns is statistically significant but economically small. This was one of the first studies to examine the relationship between investor sentiment and the stock market based on social media. To date, the use of investor sentiment on social media to aid in the prediction of stock market returns remains an important topic in finance (McGurk et al., 2020). High-frequency investor sentiment measurement is a unique advantage of measuring investor sentiment based on social media and NLP. According to Renault (2017) and Sun et al. (2016), prediction of stock market returns can be facilitated by examining social media a half-hour ahead and a half-hour behind investor sentiment. This suggests a causal chain between the stock market, investor sentiment, and social media, as well as interaction between investor sentiment and the stock market. Behrendt and Schmidt (2018), using a higher-frequency measure of sentiment at 5-min intervals, conclude that investor sentiment is statistically significantly related to stock returns but cannot be applied to stock return prediction in practice (Behrendt & Schmidt, 2018). Numerous studies have documented that investor sentiment can be used to achieve better predictions of stock market returns, yet the underlying hypothesis that underpins these studies—the synergistic relationship between stock prices and investor sentiment—has received little scholarly attention. Confirming the presence of synergy between stock markets and investor sentiment based on a larger number of markets is important in the study of the predictive role of investor sentiment on stock market returns and the explanation of multiple financial phenomena associated with investor sentiment.

In this paper, which uses reliable NLP techniques and a sample size of over 3 million, we investigate the synergistic pattern between stock prices and investor sentiment in the stock market. We use crawler technology to obtain investor messages on social media from January 1, 2019, to May 31, 2022, and a convolutional neural network to divide investor messages into "positive," "neutral," and "negative" groups based on the strength of confidence in the stock market revealed in investor messages. We construct an investor confidence index (ICI) following Antweiler's method, and then we analyze the synergistic relationship between investor sentiment and stock prices by using ICI as a proxy for sentiment, along with panel data on investor sentiment and stock market prices.

Moreover, in the context of the continuing prevalence of COVID-19, numerous scholars have studied the complex effects of COVID-19 on the stock market and stock market investment sentiment (Eachempati et al., 2021; Fallahgoul, 2021; Hoang & Syed, 2021; Naseem et al., 2021; Sun & Shi, 2022). However, few studies have mentioned the effect of COVID-19 on the synergistic relationship between stock prices and investor sentiment. In particular, a gap remains in the literature on the impact of urban lockdowns, the most severe response to COVID-19, on investor sentiment synergy. We anticipate that urban blockades will exacerbate investor apprehension. This anxiety will be manifested on social media, affecting the causal relationship between investor sentiment, stock prices, and stock market returns. We use causal inference machine learning to look at how the Chinese government's two urban lockdowns (in Wuhan and Shanghai) in response to COVID-19 affected the synergy related to investor sentiment. This fills a gap in the literature on the topic.

Using a novel controlled experiment, Wang et al. (2022) investigate the effect of online message board mood on stock returns. They discover a significant causal impact between social media mood and same-day stock returns. The sentiment has no effect on stock returns on future days, and this effect is primarily driven by messages with positive sentiment. The study of Wang et al. and our study are quite complementary. In this paper, we investigated the relationship between stock price, stock returns and investor sentiment in three ways: general investor sentiment, investor sentiment at various times of the day, and investor sentiment during specific time periods.

Our research provides new evidence for the idea that investor sentiment can be used to predict stock prices. The most important points made in this paper are:

- At the macro level, weak but extremely significant positive synergy exists between stock prices, stock returns, and investor sentiment.

- Investor sentiment on social media is a comprehensive expectation of stock market quotes, rather than a real-time perception of the stock market by investors, and investor sentiment on social media is forward looking.
- The emotional tension caused by the lockdown of cities can significantly affect the synergy between stock prices and investor sentiment, and this effect can be manifested as either a positive or a negative synergy.
- Social media has both positive synergistic and negative synergistic signals of investor sentiment, and the synergistic relationship between stock price and in the interaction of the two signals in investor sentiment can be represented in three ways: positive, negative, and neutral. This finding helps to explain the paradox of how investor sentiment can have either a positive or negative effect on stock returns, which has been confirmed by scholars.

The research in the paper is organized as shown in Fig. 1.

Following this introduction, the paper is structured as follows. Section 2 describes the hypothesis, data, and methodology. Section 3 examines the pattern of synergy between stock prices and investor sentiment. Section 4 focuses on the periods of lockdown in Wuhan and Shanghai and investigates the synergy of investor sentiment in the short term and the impact of unusual events on the synergy between stock prices and investor sentiment. Section 5 is the conclusion.

2. Hypothesis, Data, and Methodology

2.1. Settings and Hypotheses

Broadly speaking, investor sentiment is a belief about future cash flows and investment risk (Baker & Wurgler, 2007), whereas investors' assessment of future cash flows and investment risk in the stock market is derived from fundamental judgments about the past and current stock market. Thus, a causal relationship is found between the stock market and investor sentiment, from stock market sentiment to investor sentiment. And investor sentiment also affects stock market returns, stock prices, and cash flows (Baker & Wurgler, 2007; McGurk et al., 2020; Sayim & Rahman, 2015). As a result,

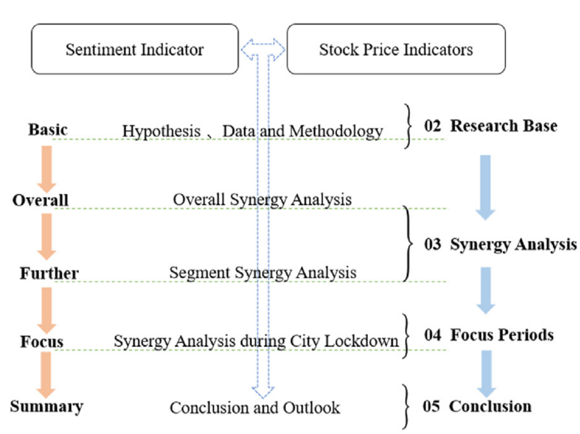

Fig. 1. Research structure.

investor sentiment and stock market sentiment are causally related, from sentiment to stock market sentiment. In light of this discussion, we create our first research setting as follows:

S1. *Investor sentiment and stock prices influence each other and are mutually causal (as shown in Fig. 2).*

Stock market fundamentals affect investor sentiment, and investor sentiment spreads on social media (Paris et al., 2015; Stieglitz & Dang-Xuan, 2013). In turn, the sentiment spread by social media has an impact on investor sentiment (Utz, 2019), and investor sentiment reacts to the stock market. Based on this logic, the stock market is indirectly influenced by social media (Ge et al., 2020), and Fig. 3 illustrates this mutual influence. Based on this discussion, we propose the second research setting as follows:

S2. *Stock market prices, investor sentiment, and social media make up a complex system of mutual influence.*

Based on S1 and S2 and the bidirectional prediction of investor sentiment and the stock market documented by numerous scholars, we posit the first hypothesis as follows:

Hypothesis 1. Positive synergy is created between investor sentiment and stock prices.

Higher stock prices are expected to lead to higher investor confidence, which drives higher stock prices; conversely, lower stock prices are expected to lead to lower investor confidence, which drives lower stock prices.

The explosion of social media and the development of NLP techniques have made high-frequency sentiment measurement possible (Kleinnijenhuis et al., 2013; Sun et al., 2016; Xing et al., 2020). The underlying logic of high-frequency sentiment measurement is similar to that of S2 and considers this cyclical process rapidly responsive. These studies argue that more real-time sentiment has a higher impact on stock prices and that higher real-time stock prices have a greater impact on investor sentiment. Based on this logic, we propose the second hypothesis as follows:

Hypothesis 2. Investor sentiment is a real-time view of how people feel about the stock market. On trading days, investor sentiment is more in line with the most recent stock price indexes.

Many scholars use investor sentiment to predict stock market returns and cash flows, and these studies are highly relevant to S1. To make the findings more reliable, some scholars first use a Granger-causality analysis and information theory to demonstrate that investment sentiment contains statistically significant ex ante information about future stock prices, before using investor sentiment for stock market prediction (Zhang et al., 2017). However, scholars argue that when researchers do not have a priori knowledge of the phenomenon under study, the results of Granger-causality tests might not be useful because of the different possible explanations (Maziarz, 2015). Zhang's findings might need more validation. For these reasons, we propose the third hypothesis as follows:

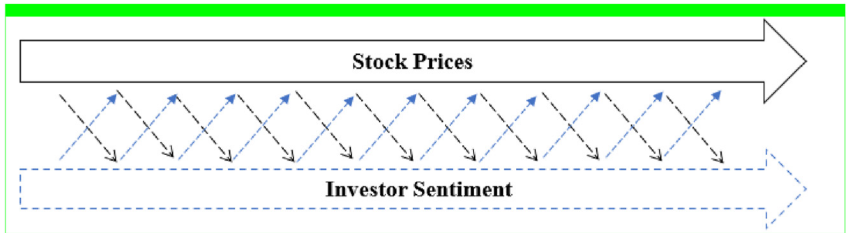

Fig. 2. Mutual cause and effect between stock prices and investor sentiment.

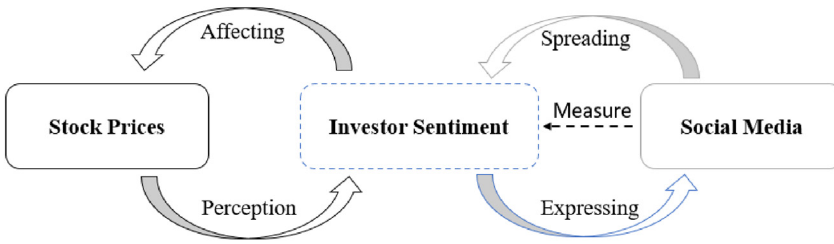

Fig. 3. Interaction of stock prices, investor sentiment, and social media.

Hypothesis 3. Investor sentiment expressed on social media is prospective and contains statistically significant ex ante information about future stock prices.

The COVID-19 pandemic severely affected daily life, disrupting world trade and trade flows (Haleem et al., 2020), as well as the stock market (Cox et al., 2020; Fernandez-Perez et al., 2021; Wagner, 2020). The most extreme response to COVID-19, urban lockdowns, can significantly increase anxiety. This concern is reflected on social media, altering the relationship between social media sentiment, stock prices, and stock market returns. A clear explanation of this discovery is necessary in order to understand the influence of important events on financial markets. In view of this, we propose the fourth hypothesis as follows.

Hypothesis 4. Urban lockdowns affect the synergy between stock prices and investor sentiment.

2.2. Data

The social media data in this study come from investor messages on China's largest social media platform for stock market investors (Dongfang Fortune stock bar). Examples of investor messages are shown in Fig. 4. Each message has a title, author, date, number of times read, number of comments, text, and so on. Like Antweiler's text messages, investor messages are usually short and mostly express personal views.

TITLE Insufficient downward momentum. May stock index will be 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 have been missed. Stop losing money at first. Coupled with the recent emergence of several slogan-style stabilization of stock measures, stock indices are expected to continue to fall in mid to 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.

Fig. 4. Examples of investor messages.

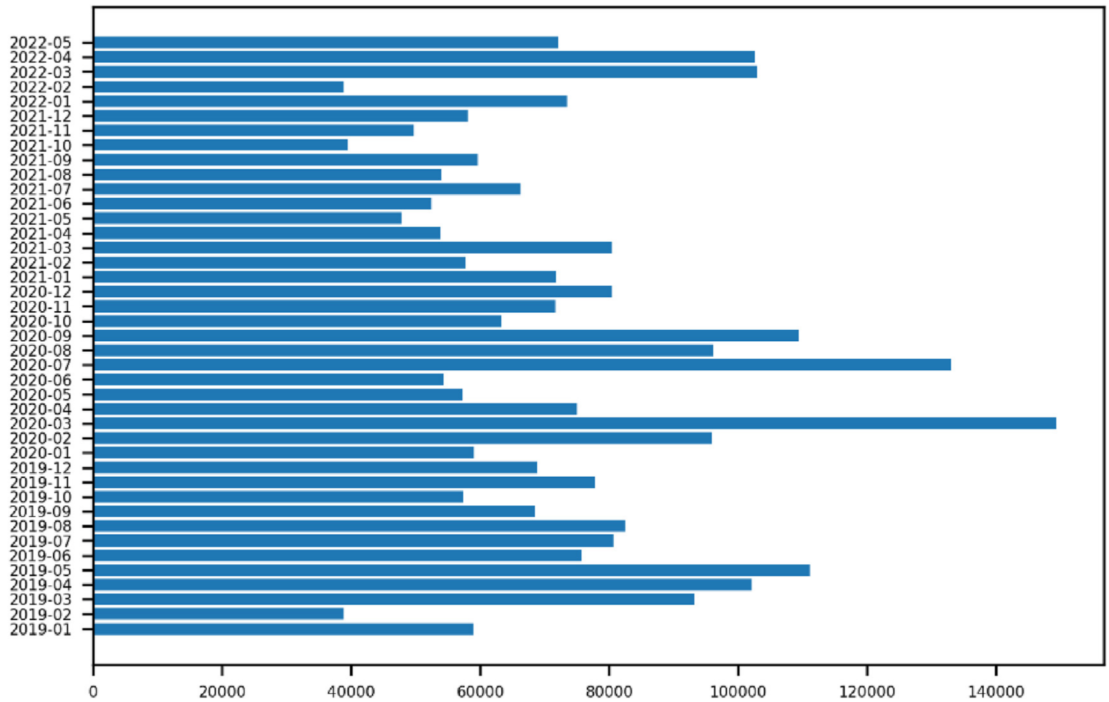

Fig. 5. Monthly distribution of investor messages.

Using crawling techniques, we obtained all investor messages on the SSE Index Bar from January 1, 2019, to May 31, 2022. Fig. 5 shows the distribution of investor messages over these 41 months, showing the heterogeneity in monthly investor messages. The total number of investor messages is about 3.05 million, with a total reader count of 2.039 billion and a total comment count of 6.9268 million. The highest daily message volume is 19,204; the average daily message volume is 2,452, and the lowest daily message volume is 13. Detailed statistics are shown in Table 1.

SSE Index Bar is the social media platform about the SSE Index, the sentiment on social media is about the SSE Index. Therefore, we use the SSE Composite Index as a proxy for the composite stock price, and the study period on stock prices is the same as the social media data, comprising 826 trading days from January 1, 2019, to May 31, 2022. The opening price, closing price, highest stock price, lowest stock price, and previous closing price of the SSE Index are used as the stock price panel data. The statistical description of the data is in Table 2.

Table 1
Statistical description of social media texts.

Variables	Daily message volume	Daily volume read	Daily comment volume
Mean	2452	1,637,761	5564
Std	2070	1,935,995	3869
Min	13	24,426	22
25%	597	741,113	3155
50%	2226	1,269,664	4825
75%	3359	1,973,761	6763
Max	19,204	44,924,350	43,562

Note: 25%, 50%, are 75% are quartile samples.

2.3. Distinguishing the Sentiment in Messages

Given that the number of investor messages is as high as 3.05 million, it is difficult to manually determine the sentiment in all the messages, so we use supervised learning to solve this problem. The basic logic of supervised learning is to let the network model learn the manual classification rules and then evaluate the network model. If the model can achieve the expected classification performance, the network model can be used instead of manual classification (Kotsiantis et al., 2007).

The text data generated by investor messages comprises multiple elements. Content other than text, which is called "noise" in text processing, does not contribute to the sentiment classification of the text. In text preprocessing, we omit numbers, punctuation marks, extra spaces, and other characters that are difficult to understand, such as parentheses, brackets, and ampersands. We also replace web links in the text with URL and usernames with USERNAME and then start encoding the text. The text data needs to be encoded before the model is trained and a deep learning model is used for

Table 2
Statistical description.

Variables	Opening price	Closing price	Highest price	Lowest price	Previous closing price
Mean	3204.52	3207.10	3225.67	3183.66	3206.26
Std	299.10	298.02	299.01	296.68	299.05
Min	2446.02	2464.36	2488.48	2440.91	2464.36
25%	2933.14	2937.86	2951.66	2918.35	2937.66
50%	3240.14	3243.37	3265.97	3213.67	3243.37
75%	3475.41	3482.86	3493.01	3455.81	3482.86
Max	3721.09	3715.37	3731.69	3692.82	3715.37

Note: 25%, 50%, and 75% are quartile samples. RMB is the price unit.

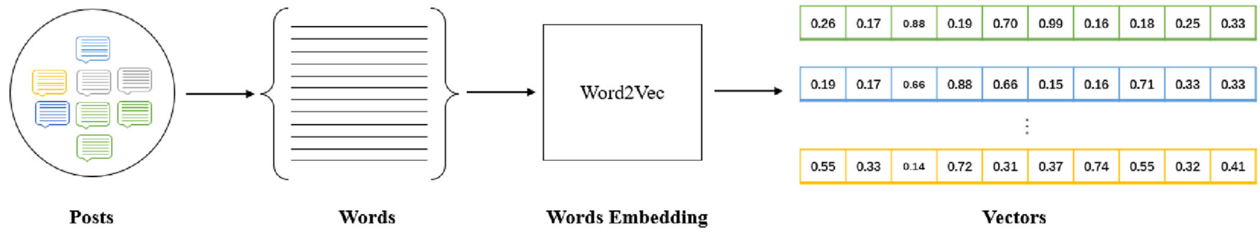

Fig. 6. Flowchart for encoding text.

sentiment classification. The encoded text generates a set of vectors, which represent words in the dataset that can be understood and learned by the network model. Word2vec is a two-layer neural network for processing text and representing words as vectors, and the model can transform the input text data into a set of vectors as output (Goldberg & Levy, 2014). Because of the peculiarities of the Chinese language, text needs to be subsumed before text is encoded. As shown in Fig. 6, we split the sentence into words and then use Word2vec for encoding to transform the text into word vectors.

As shown in Fig. 7, we first randomly sampled all the messages and extracted 1 percent of the messages (about 30,200) for manual classification. The messages were divided into three groups—"positive," "negative," and "neutral"—based on the level of confidence in the market expressed in the investor messages. Multiple convolutional neural networks are then trained using 90 percent of the manually marked data as the training data set. The other 10 percent of the marked data is used to evaluate the classification performance of the neural network and to obtain the best-performing classification network. Finally, this network model is used to classify the unclassified text for sentiment.

The classification performance of the deep learning model used in this study is shown in Table 3. The accuracy of classification is 90.99 percent for samples with positive sentiment, 94.62 percent for samples with negative sentiment, and 81.83 percent for samples with neutral sentiment. The combined accuracy of the network model in classifying the sentiment of unlearned investor messages is 89.14 percent, which is higher than 88.1 percent by Antweiler and Frank and 85.4 percent by Xiong et al. (2017) In particular, the classification accuracy of positive and negative samples, which are the most important for constructing the ICI, is 90.99 percent and 94.62 percent, respectively. It can be said that there is a high level of confidence in the classification model and that there is no systemic error.

2.4. Measurement of Investor Sentiment and Synergy Index

To measure the combined sentiment of investors expressed by the daily average of 2452 messages, the ICI is defined following Antweiler's method of defining bullish sentiment index. $N_p^{positive}$ is the number of messages with confidence in the stock market at time p, and $N_p^{negative}$ is the number of messages without confidence in the stock market at time p. The ICI is constructed as follows.

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

When the percentage of social media messages expressing market confidence during a given time period is higher, the ICI for that time period is higher, indicating that investors have more confidence in the market. When the proportion of messages expressing confidence in the market on social media is lower, the ICI for that time period, indicating that investors lack confidence in the market, is lower. In this paper, ICI is used as a proxy for investor sentiment, with a high ICI representing high investor sentiment and a low ICI indicating low investor sentiment. We follow Liu et al. (2022) by not using the Agreement Index constructed by Antweiler to co-construct the sentiment indicators.

The correlation of two or more sets of serial data can be measured in various ways (Taylor, 1990). We tried to measure

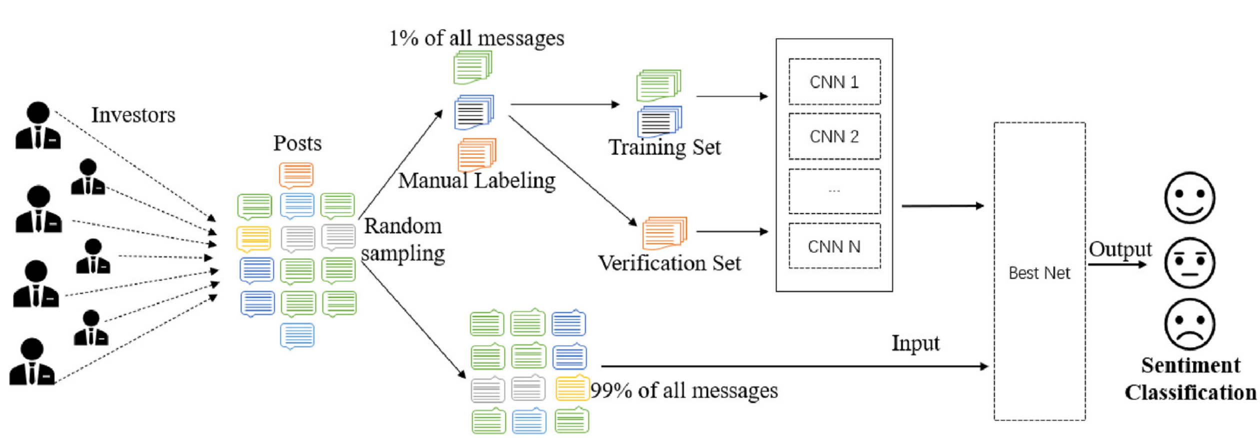

Fig. 7. Flowchart of text sentiment classification.

Table 3
Confusion matrix for network performance.

True classification	Prediction classification			
	Positive	Negative	Neutral	Accuracy
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 by referring to the confusion matrix commonly used in the field of machine learning.

Table 4
Definition of the variables.

Variables	Description
ICI_{Day}	Daily investor sentiment
ICI_A	Investor sentiment in the period 0:00–9:30
ICI_B	Investor sentiment in the period 9:30–11:30
ICI_C	Investor sentiment in the period 11:30–13:00
ICI_D	Investor sentiment in the period 13:00–15:00
ICI_E	Investor sentiment in the period 15:00–24:00
OP	Opening price of the day
CP	Closing price of the day
HP	Highest share price of the day
LP	Lowest share price of the day
PCP	Previous closing price
R	Stock returns
SI_p	Synergy index of stock price and investor sentiment at time p

the synergy between stock prices and investor sentiment using various methods, such as regression coefficients, Pearson's correlation coefficients, Spearman's correlation coefficients, and Kendall correlation coefficients. However, the results of these methods differ only slightly and have no impact on the conclusions of our analysis. As a result, we employ the efficient and simple Pearson's correlation coefficient to assess the relationship between investor sentiment and stock prices. The synergy index between stock prices and investor sentiment at time p can be expressed as

$$SI_p = pearson(ICI_p, Price_p) \tag{2}$$

In addition, following previous literature (Mallikarjuna & Rao, 2019; Nimal, 1997), we define stock returns as follows:

$$R_t = \frac{(P_t - P_{t-1})}{P_{t-1}} \times 100 \tag{3}$$

The variables used in this paper are defined in Table 4.

3. Synergistic Analysis

3.1. Daily Synergy and Intraday Time Synergy

Table 5 shows the synergistic relationship between investor sentiment and stock price indicators for 826 trading days from January 1, 2019, to May 31, 2022. Weak but extremely significant positive synergy is found between the opening price, closing price, daily high stock price, daily low stock price, previous closing price, stock returns, and investor sentiment. However, it appears that investors are more sensitive to stock prices than stock returns. This result confirms H1. Higher stock prices drive higher sentiment, and higher sentiment drives higher stock prices. Similarly, lower stock prices lead to pessimistic sentiment, and pessimistic sentiment leads to lower stock prices.

In Fig. 8, we divide the 24 h in a trading day into 5 periods (labeled A, B, C, D, and E), divided between the morning and afternoon trading sessions, in which B and D are trading periods, A is the 9 h and 30 min before the opening, C is the 1.5 h of the noontime break, and E is the 9 h after the closing. The synergy between investor sentiment and stock prices is further analyzed for each of these periods. The length of the bar indicates the length of the period.

Fig. 9 shows the distribution of the number of investor messages over five time periods, showing the time of day at which investors prefer to express their opinions. The opening price, closing price, highest stock price, and lowest stock price describe the fundamentals of the stock price in different dimensions during the day. It is impossible to estimate at which moment the highest and lowest stock prices will appear, which can be highly contingent, but the point at which the opening and closing prices will appear is certain. The opening price appears at the beginning of period B, while the closing price appears at the end of period D. We examine the robustness of H1 as well as test H2 by measuring the relationship between investor sentiment and OP, CP, HP, LP, and PCP for periods A, B, C, D, and E, respectively.

Table 6 statistically shows the synergy between investor-segmented sentiment and stock price fundamentals. In all time segments, investor sentiment has a weak but extremely significant correlation with OP, CP, HP, LP, and PCP of the stock market, further confirming H1. However, the synergy index of opening prices with investor sentiment in period B is 0.25237, which is not much higher than the correlation between OP and ICI_C, ICI_D, and ICI_E. The synergy index between closing prices and investor sentiment for periods D and E is 0.25178 and 0.23537, respectively. This is not much higher than the correlation between CP and ICI_A, ICI_B, and ICI_C. The results of the analysis in Table 6 show that the stock price indexes at specific times of the day do not have higher synergy with investor sentiment in adjacent periods. This finding fails to support H2. If H2 does not hold, two derived hypotheses, at least one of which should be true, can be derived to explain why H2 is false.

Table 5
Analysis of synergy in social media sentiment and stock price fundamentals.

	OP	CP	HP	LP	PCP	R
ICI_{Day}	0.35409 (0.00000)	0.38630 (0.00000)	0.36512 (0.00000)	0.37312 (0.00000)	0.35716 (0.00000)	0.22014 (0.00000)

Notes: The values in parentheses are p-values.

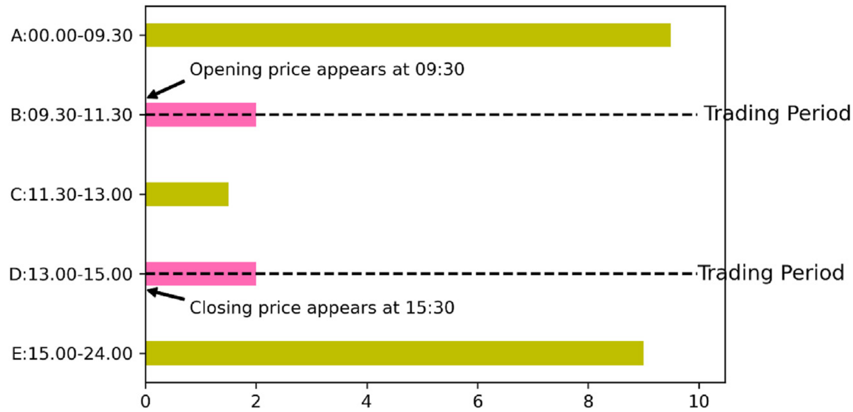

Fig. 8. Investor sentiment segmented into daily periods.

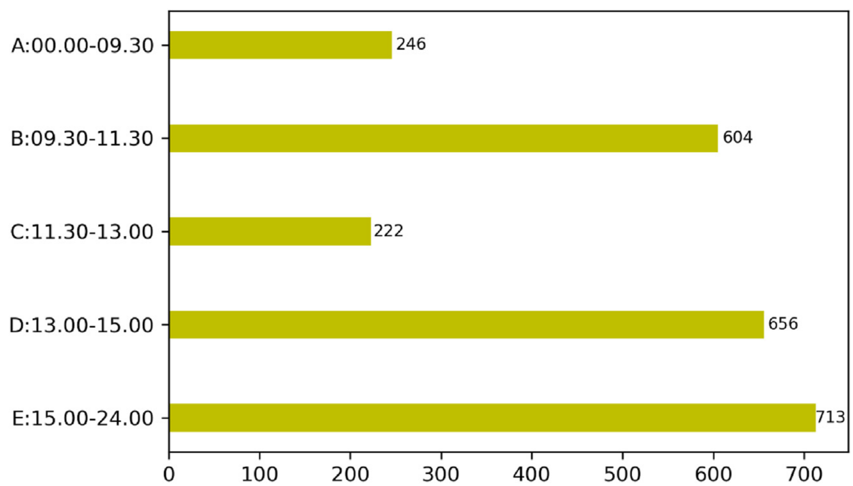

Fig. 9. Average message volume per time period.

Table 6
Segmented investor sentiment synergy analysis.

	OP	CP	HP	LP	PCP	R
ICI_A	0.39051 (0.00000)	0.38630 (0.00000)	0.36512 (0.00000)	0.37312 (0.00000)	0.35716 (0.00000)	0.06875 (0.04823)
ICI_B	0.25237 (0.00000)	0.27132 (0.00000)	0.26608 (0.00000)	0.26379 (0.00000)	0.26138 (0.00000)	0.06866 (0.04854)
ICI_C	0.18881 (0.00000)	0.20951 (0.00000)	0.19466 (0.00000)	0.20783 (0.00000)	0.19025 (0.00000)	0.14816 (0.00002)
ICI_D	0.22769 (0.00000)	0.25178 (0.00000)	0.23717 (0.00000)	0.23382 (0.00000)	0.23427 (0.00000)	0.13242 (0.00014)
ICI_E	0.19499 (0.00000)	0.23537 (0.00000)	0.20369 (0.00000)	0.21962 (0.00000)	0.19252 (0.00000)	0.33668 (0.00000)

Note: *P*-values in parentheses correspond to the *R*-values. Mean column is the mean value of ICI synergy with OP, CP, HP, LP, and PCP.

Hypothesis 2-E1. Investor sentiment on social media is a composite sentiment based on fluctuations in stock prices, rather than real-time feedback on a specific stock price. The impact of investor sentiment on stock prices is also composite, rather than being driven purely by a specific recent stock price.

Hypothesis 2-E2. Investor sentiment on social media feeds back based on a particular metric, but the interaction between stock price, investor sentiment, and social media is not in real time and is lagged.

If H2-E1 were confirmed, then high-frequency sentiment measures would fail to predict future stock prices and returns, which contradicts the findings of many studies (Kleinnijenhuis et al., 2013; Xing et al., 2020). If H2-E2 held, then, the results of the synergy analysis in Table 6 would show a lag in the synergy between specific stock price indicators and investor sentiment. For example, the opening price would show higher synergy with investor sentiment sometime after period B, but we do not observe this phenomenon. For the reasons mentioned above, H2-E2 is not confirmed or is only partially confirmed. However, based on the exclusion method, H2-E1 is confirmed. Investor sentiment on social media gives a full picture of how people feel about the stock market as a whole, not how they feel about a single stock price indicator.

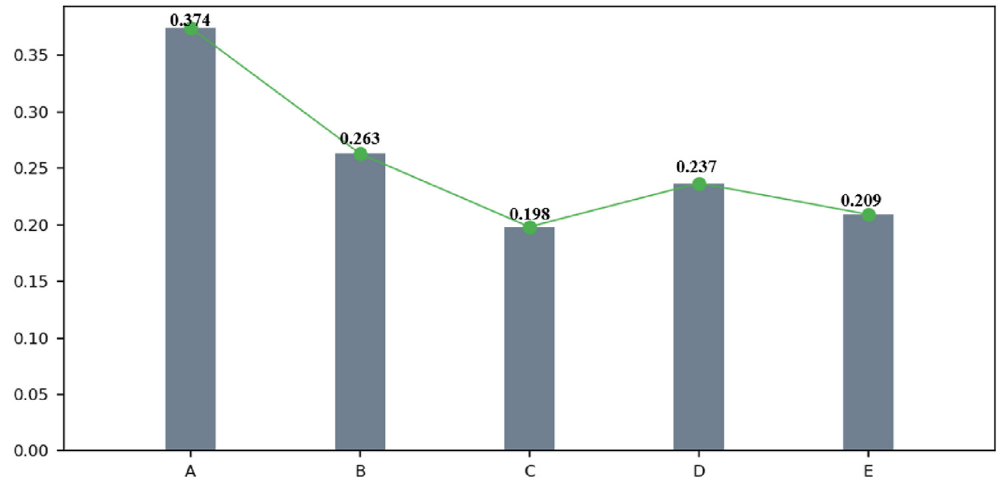

Fig. 10. Correlation between the mean stock price and segmented investor sentiment.
Note: The stock price uses an average index price with the equation.$Mean\ Price = (OP + CP + HP + LP)/4$.

The results in Table 6 show that the correlation between investor sentiment and the opening price, closing price, highest stock price, and lowest stock price is much higher in period A than in other periods. Fig. 10 depicts the synergy index between stock price and investor sentiment in periods A, B, C, D, and E, enabling us to reach this conclusion more easily. However, the emergence of investor sentiment in period a precedes the emergence of OP, CP, HP, and CP. This suggests that investor sentiment in the stock market is forward looking and contains statistically significant ex ante information about future prices, which confirms H3 and Zhang's conclusions. It also shows that, after the market opens, investor sentiment is affected by various external factors, which reduces this forward-looking attitude, or more noise is generated in investor sentiment, which supports our rejection of H2 from another perspective.

Table 6 shows an interesting and commonsense finding. The impact of media sentiment on stock returns shows a gradual increase over the course of the day, from period A to period E, as illustrated in Fig. 11. In the earlier periods, A and B, it was difficult for investors to estimate the daily returns, and the correlation between investor sentiment and returns was very weak. In contrast, in periods C and D, investors have some confidence about the daily return, and the correlation between investor sentiment and return increases over that in periods A and B. After the stock market closes for the day (period E), investors already have a handle on the daily returns. Based on investors' preference for profit taking, good returns lead to positive sentiment, and bad returns lead to negative sentiment. These sentiments are reflected on social media, thus creating a causal relationship between social media sentiment and stock market returns.

Behrendt and Schmidt (2018) suggest that the correlation between high-frequency sentiment measures and stock market returns only has statistical significance but no practical application. This implies that high-frequency investor sentiment measures should not be applied to extremely recent stock price forecasts. Our findings are in line with those of Behrendt and Schmidt, who say that investor sentiment after the opening of the stock market is complex and integrated and does not have

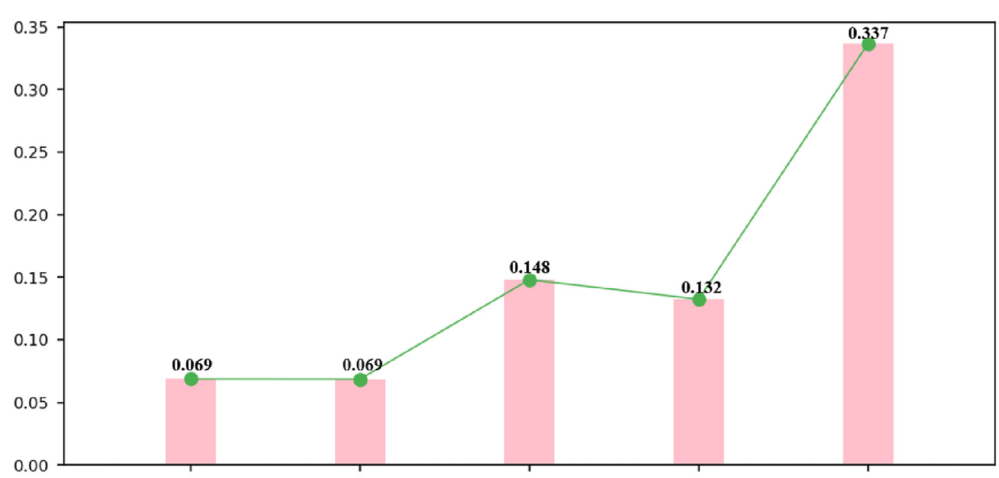

Fig. 11. Correlation between stock returns and segmented investor sentiment.

greater synergy with the stock market in the near term. At the same time, our findings provide new ideas for stock price forecasting based on investor sentiment. That is, investor sentiment in social networks is forward looking and can be used for stock price prediction, whereas investor sentiment before the stock market opens is clearer, more readable, and more valuable to study.

3.2. Regression Analysis of Stock Returns Using Investor Sentiment

We attempt to forecast stock returns using investor sentiment, following Baker and Wurgler (2006) and Wang et al. (2022), to confirm the positive correlation between investor sentiment and returns. To forecast returns, we combine investor sentiment for the entire day (ICI_{Day}) with five periods of sentiment (ICI_A, ICI_B, ICI_C, ICI_D, and ICI_E). The predictive power of intraday composite sentiment and sentiment at various times of the day on stock returns is examined separately. The outcomes of the regression analysis employing investor sentiment for stock return predictions are displayed in Table 7. Panel A shows the results using data from all 826 trading days, and Panel B uses a selection of trading days with serial numbers 600–700. The goal is to show the differences in the impact of social media sentiment on stock returns over time.

The statistical results in Table 7 show that the results of ICI_A on investor returns are insignificant, which indicates that investors cannot effectively predict the returns of the day until the opening of the market. In Panel A, all models except ICI_A have Prop (F-statistic) values that are less than 0.05 and F-statistics that are large enough to indicate a linear relationship between stock returns and investor sentiment that can be explained. Furthermore, the value of $P>|t|$ is less than 0.05, and the equation successfully explains the relationship between investor sentiment and stock returns, except for ICI_A. The connection between investor sentiment and stock returns becomes obvious after the stock market opens. In particular, when ICI_E is used to regress returns, the significant coefficient is nearly 0, the F-statistic value is as high as 105.1, and the *t*-value is 10.25. The mean square error likewise falls to its lowest value, obtaining the best predictive performance. When full-day sentiment (ICI_{Day}) is used in the analysis, R^2 is 0.0525, and the model can account for only 5.25 percent of the data. The highest R^2 when predicting stock returns with ICI_E is as little as 0.1164.

The sample data used in Panel B comprise a continuous segment of the full sample, and the findings presented in Panel B are consistent with the findings in the full sample analysis, but the data in this segment have a better fit. When ICI_{Day} is used for stock return prediction, R^2 is 0.3502, which indicates that 35.02 percent of the sample is revealed by the model. When ICI_B, ICI_C, and ICI_D are used for stock return prediction, R^2 is 0.1039, 0.0806, and 0.1624, respectively. In line with the findings in Panel A, R^2 reaches a maximum of 0.4862 when ICI_E is used to forecast stock returns. Fig. 12 shows the fit of the regression models using ICI_A, ICI_B, ICI_C, ICI_D, ICI_E, and ICI_{Day} to forecast stock market returns.

In conclusion, there is a significant and interpretable linear relationship between investor sentiment and returns, and this relationship is clearest in period E after the market's midday break. Individual statistical intervals (shown in Panel B), 48.62 percent of the sample can be explained by the model, but, overall, the model explains less of the sample. We therefore conclude that sentiment and stock returns have a significant positive correlation in social media, but little overall economic significance. This finding is consistent with that of Antweiler and Frank (2004). In addition, the regressors based on the five forecasting models (ICI_A, ICI_B, ICI_C, ICI_D, and ICI_E) are significantly different. After the market closes, investors transmit the sentiment on accurate returns to social media, which leads to the strongest relationship between ICI_E and returns.

Table 7
Ordinary least squares (OLS) regression analysis of investor sentiment on stock returns.

Panel A. OLS regression with the full sample

	Coeff.	t	$P>\|t\|$	MSE	F-statistic	P (F-statistic)	R^2
ICI_{Day}	1.1031	6.6473	0.0000	1.2283	44.1861	0.0000	0.0525
ICI_A	0.2360	1.7668	0.0776	1.2913	3.1216	0.0776	0.0039
ICI_B	0.3334	2.0564	0.0401	1.2895	4.2288	0.0401	0.0053
ICI_C	0.5047	4.1674	0.0000	1.2687	17.3674	0.0000	0.0213
ICI_D	0.6055	3.9138	0.0001	1.2719	15.3180	0.0001	0.0188
ICI_E	1.3379	10.2516	0.0000	1.1455	105.0956	0.0000	0.1164

Panel B. Partial sample OLS regression, sample series number: 600-700

	Coeff.	t	$P>\|t\|$	MSE	F-statistic	P_(F-statistic)	R^2
ICI_{Day}	2.6218	7.2678	0.0000	0.4941	52.8215	0.0000	0.3502
ICI_A	0.1241	0.3981	0.6914	0.7591	0.1585	0.6914	0.0016
ICI_B	1.3504	3.3706	0.0011	0.6814	11.3609	0.0011	0.1039
ICI_C	0.7197	2.9311	0.0042	0.6991	8.5913	0.0042	0.0806
ICI_D	1.4569	4.3590	0.0000	0.6369	19.0009	0.0000	0.1624
ICI_E	2.4839	9.6291	0.0000	0.3907	92.7202	0.0000	0.4862

Note: The shaded row indicates that the data are not significant. Panel B depicts a subset of the total sample.

4. The Impact of Urban Lockdowns on Synergy

As COVID-19 has affected people's lifestyles (Bennett et al., 2021; Cartenì et al., 2020; Molarius & Persson, 2022), it has also severely affected economic and financial markets (Bashir et al., 2020; Boissay & Rungcharoenkitkul, 2020). Urban lockdowns are the most extreme tool employed in response to COVID-19. We focus on the periods of lockdown in Wuhan and Shanghai to examine how particular events affect the relationship between stock prices and how investors feel about them. By looking at and understanding how these events affect the relationship among stock prices, stock market returns, and investor sentiment, we can get a better idea of how investor sentiment, stock prices, and stock returns all affect one another in social media.

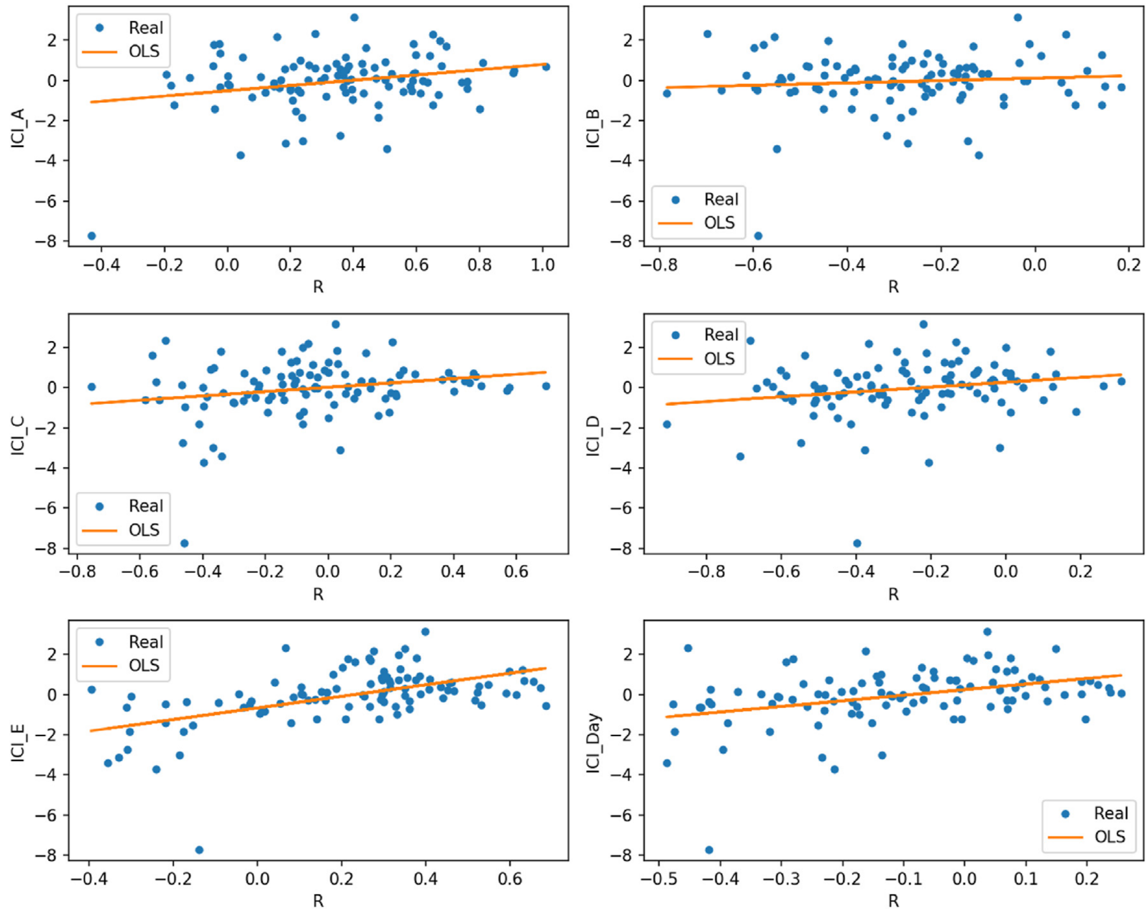

Fig. 12. Fitted graphs of OLS regression analysis of investor sentiment and stock returns.

4.1. Investor Sentiment Synergy during the Lockdown in Wuhan

Between January 23 and April 7, 2020 (the blockade was lifted on April 8), Wuhan implemented strict COVID-19 control measures (Cao et al., 2020), with a lockdown lasting 76 days, comprising 47 stock market trading days. As shown in Fig. 13, we use p_0 to denote the 47 trading days during the blockade as the treatment group, p_{-1} to denote the 47 trading days before the lockdown as the pre-control group, and p_1 to denote the 47 trading days after the lockdown was lifted as the post-control group.

Table 8 shows the synergy indexes of stock prices and investor sentiment for periods p_0, p_{-1}, and p_1. The synergy index between investor sentiment and the closing price is 0.37252 and 0.39939 for the periods p_{-1} and p_1, respectively, and 0.61069 for the period p_0, which is remarkably higher than that for p_{-1} and p_1. The synergy index between investor sentiment and the lowest stock price is 0.29064 and 0.31740 for p_{-1} and p_1, respectively, and 0.49941 for p_0. The synergy index is higher and statistically significant in the 47 trading days during the lockdown in Wuhan than in the 47 trading days before and after the lockdown. In addition, the synergistic index and statistical significance of investor sentiment and stock returns are higher during the city's closure than before and after the closure.

Because of the large difference between stock prices and investor sentiment values, we normalized investor sentiment and stock prices using a z-score (Cheadle et al., 2003) in order to observe investor sentiment and stock prices together. Fig. 14 shows the relationship between investor sentiment and stock price volatility in p_{-1}, p_0, and p_1, a total of 141 trading days.

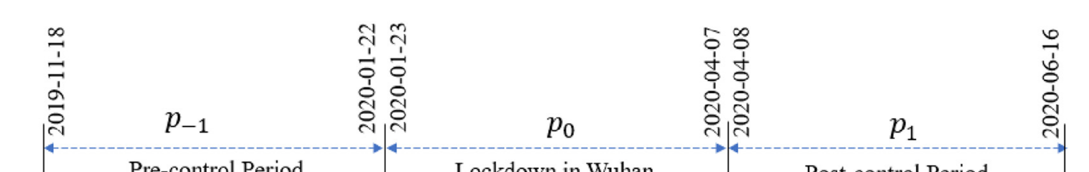

Fig. 13. Period of lockdown in Wuhan and control periods.

Table 8
Synergistic statistics for the period of lockdown in Wuhan and control periods.

	OP	CP	HP	LP	PCP	R
ICI_{Day}	in p_{-1} 0.29261[b]	0.37252[c]	0.28965[b]	0.29064[b]	0.33723[b]	0.32044[b]
ICI_{Day}	in p_0 0.42726[c]	0.61069[d]	0.37972[c]	0.49941[c]	0.53039[d]	0.43485[d]
ICI_{Day}	in p_1 0.19497	0.39939[b]	0.23580[a]	0.30500[b]	0.31740[b]0.34688[c]	

[a] Significant at 0.1.
[b] Significant at 0.05.
[c] Significant at 0.01.
[d] Significant at 0.001.

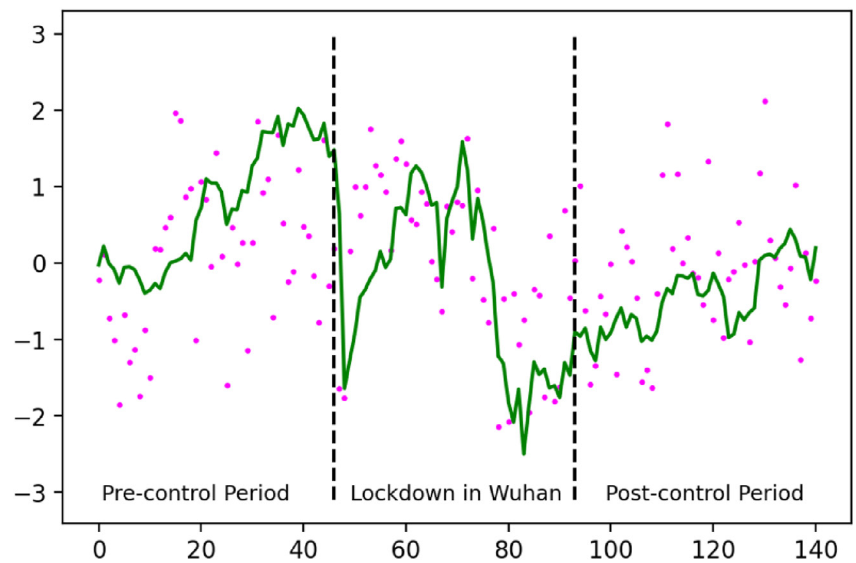

Fig. 14. Synergy during the lockdown in Wuhan and control periods.
Note: The *x*-axis indicates the 141 (3 × 47) trading days. The pink dots represent investor sentiment on a day, and the green curve shows composite stock volatility. The stock price uses an average index price based on the equation: *Mean Price* = (*OP* + *CP* + *HP* + *LP* + *PCP*)/5.

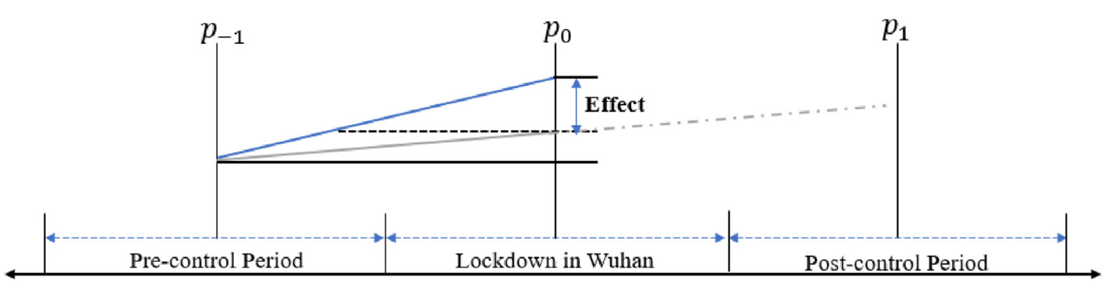

Fig. 15. Diagram of DID method for lockdown effect in Wuhan based on linear assumptions.

Fig. 12 shows that the relationship between investor sentiment and stock price volatility is stronger during the lockdown.

As shown in Fig. 15, based on the difference-in-difference (DID) model, the causal effect of an urban lockdown on investor sentiment synergy can be calculated by subtracting the before-and-after change in the treatment group (D1) from the before-and-after change in the control group (D2) (De Chaisemartin & d'Haultfoeuille, 2018). D1 can be obtained by subtracting the synergy index (SI_{p0}) in period p_0 from the synergy index (SI_{p-1}) in period p_{-1}. Assuming that the synergy index SI between investor sentiment and stock prices has a linear relationship in the local interval that we are examining, D2 can be written as

$$D2 = \frac{\left(SI_{p1} - SI_{p-1}\right)}{2}$$

The effect of the lockdown in Wuhan on SI is

$$E_{Wuhan} = D1 - D2$$

$$= SI_{p0} - SI_{p-1} - \frac{SI_{p1} - SI_{p-1}}{2} \tag{4}$$

$$= \frac{SI_{p0} - SI_{p-1}}{2} + \frac{SI_{p0} - SI_{p1}}{2}$$

Under the assumption that SI is linearly distributed in the local interval, the effect of the lockdown in Wuhan on the synergistic effect between investor sentiment and closing prices, with reference to Equation (4), is

$$E^{CP}_{Wuhan} = \frac{0.61069 - 0.37252}{2} + \frac{0.61069 - 0.39939}{2} = 0.224735$$

The impact of the lockdown in Wuhan on the synergistic effect of stock returns is as follows:

$$E^{R}_{Wuhan} = \frac{0.43485 - 0.32044}{2} + \frac{0.43485 - 0.34688}{2} = 0.10119$$

Similarly, E^{HP}_{Wuhan}, E^{LP}_{Wuhan}, and E^{PCP}_{Wuhan} are calculated as 0.20159, 0.20308, and 0.116995, respectively. In period p_{-1} the correlation between ICI and OP in the control group was not significant, making E^{OP}_{Wuhan} impossible to calculate.

During the Wuhan lockdown, the synergistic index between stock prices and investor sentiment rises significantly. We calculated the effect of the city's closure on SI based on the assumption that the SI index is linearly distributed in the local interval, referring to DID. However, as Bertrand states, the evaluation results of the DID method might not be meaningful if the underlying assumptions do not hold (Bertrand et al., 2004). If it is assumed that the SI of p_{-1}, p_0, and p_1 are not linear when the city's closure has no effect, the measure of lockdown in Wuhan above is not meaningful, and in fact we cannot confirm this assumption. However, we believe that the idea of measuring the effect of city lockdown based on the DID model is still informative and therefore document this process in the paper.

4.2. Investor Sentiment Synergy during the Lockdown in Shanghai

The Shanghai government implemented zoning classification and grid management for Pudong, Pune, and adjacent areas at five o'clock on April 1, 2022, and permitted full restoration of normal production and living conditions as of June 1. Shanghai's two-month lockdown was the second large-scale urban lockdown by the Chinese government in response to COVID-19. As shown in Fig. 16, we examine the change in SI before and after Shanghai's closure, with p_0 denoting the 38 trading days during the closure as the treatment group and p_{-1} denoting the 38 trading days before the closure as the control group.

Table 9 presents the synergy indexes between stock prices and investor sentiment during the lockdown in Shanghai and the control period. The synergistic index is statistically very significant in p_{-1} and statistically less significant in p_0. Even the correlation between CP, LP, and ICI is statistically insignificant in p_0, which is the opposite of what happened while Wuhan was closed. From the perspective of SI, the synergy index changed from positive to negative after the closure of Shanghai. This implies that investor confidence decreases when stock prices rise, whereas investor sentiment rises when stock prices fall, showing an unconventional reversal in investor sentiment. Fig. 17 shows the fluctuations in investment sentiment and composite stock prices for a total of 76 trading days during p_{-1} and p_0, showing that investor sentiment and stock price volatility tended to diverge while the city was closed. The results of the analysis in Table 9 illustrate that, although, investor sentiment shows an extremely positive and significant synergy with stock price volatility in the long run, not all local areas are consistent with this finding. However, just as in Wuhan while it was closed, synergy between investor sentiment and stock returns is significantly higher in Shanghai while it was closed than beforehand.

Because our study ends just after the lockdown in Shanghai, post-control group data for the 38 trading days after the city was reopened are not available. As a result, we could not estimate the effect of the lockdown in Shanghai on synergy between stock prices and investor sentiment using a DID model based on an assumption of a linear distribution of SI values.

As shown in Fig. 18, we employ an alternative method by projecting 38 trading days backward as the p_{-2} control group based on p_{-1}. p_{-2} , p_{-1} , and p_0 are formed to estimate the effect of the lockdown in Shanghai. The correlation between stock prices and investor sentiment in p_{-2} has no statistical significance ($p > 0.2$).

In summary, although at the macro level, investor sentiment and stock prices have extremely significant positive synergy, at local intervals, investor sentiment and stock prices might have negative synergy or no synergy. During the lockdown in Shanghai, we see a negative correlation between investor sentiment and stock prices. However, we cannot tell from the data whether the Shanghai lockdown had an effect on the synergy index.

4.3. Evaluation on the Effect of City Lockdowns

The assumption that the synergy index is linearly distributed across local time periods used in the calculation of the effect of closing Wuhan on the synergy effect, combined with the analysis in Sections 4.1 and 4.2, might not hold. Thus, H4 is not supported. We address this issue using alternative methods.

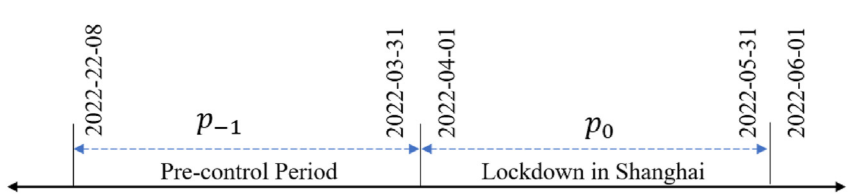

Fig. 16. Period of lockdown in Shanghai and control period.

Table 9
Synergistic statistics on the lockdown in Shanghai and control period.

	OP	CP	HP	LP	PCP	R
ICI_{Day} in	p_{-1} 0.46727[c]	0.59474[d]	0.49138[c]	0.56535[d]	0.45664[c]	0.35201[b]
ICI_{Day} in	p_0 −0.31033[b]	−0.11543	−0.29481[a]	−0.20728	−0.35406[b]	0.50331[d]

[a] Significant at 0.1.
[b] Significant at 0.05.
[c] Significant at 0.01.
[d] Significant at 0.001.

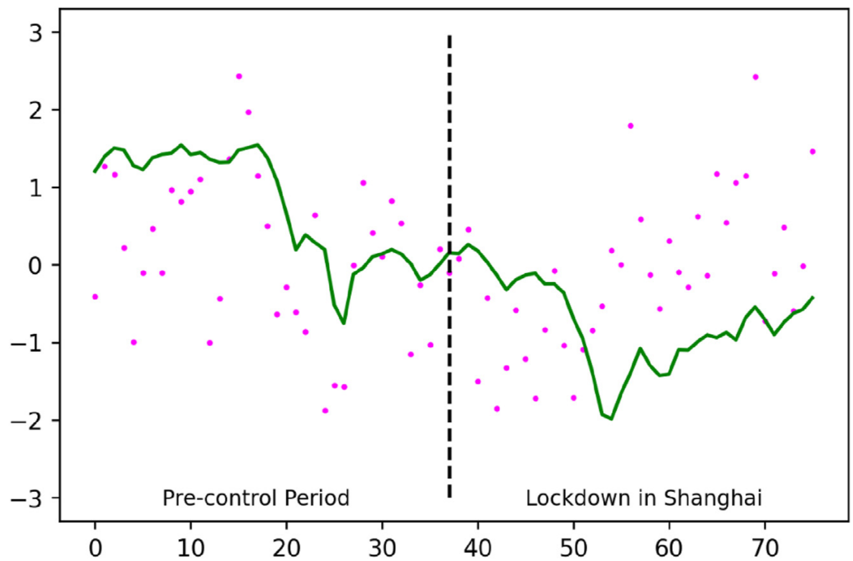

Fig. 17. Synergy between lockdown in Shanghai and control period.
Note: The x-axis indicates the 76 (2 × 38) trading days. The pink dots represent the investor sentiment on a day, and the green curve shows the composite stock volatility. The stock price uses an average index price based on the equation: $Mean\ Price = (OP + CP + HP + LP + PCP)/5$.

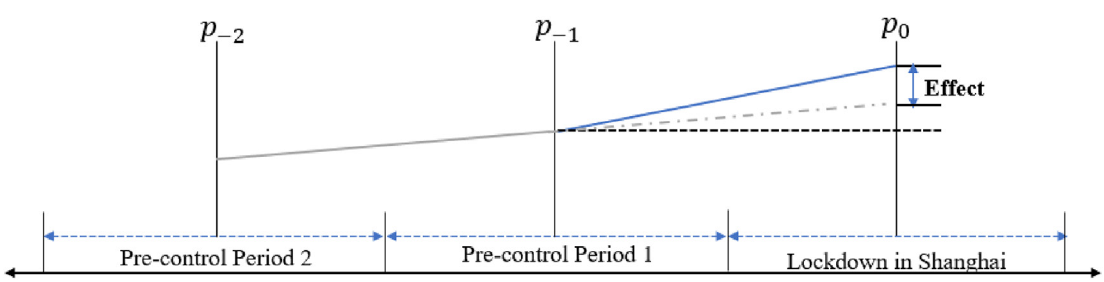

Fig. 18. Numerical estimation of D2 based on linear assumptions.

Causal ML is a Python package that provides a suite of uplift modeling and causal inference methods using machine learning algorithms based on recent research. Essentially, it estimates the causal impact of intervention T on outcome Y for users with observed features X, without strong assumptions on the model form. Considering that the relationship between stock prices and investor sentiment might be complex and nonlinear, we do not use the ordinary ML model. Instead, we use the XGBoost algorithm, which has excellent performance in the field of categorical regression and a multilayer neural network model that can simulate any function to measure the effect of urban lockdowns on SI.

Based on the official Causal ML website, the assessment process consists of two steps as follows:

Step 1. Estimate the average outcomes μ0(x) and μ1(x):

$$\mu_0(x) = E[Y(0)|X = x]$$
$$\mu_1(x) = E[Y(1)|X = x]$$

using ML models.

Step 2. Define the conditional average treatment effect (CATE) estimate as:

$$\hat{\tau}(x) = \hat{\mu}_1(x) - \hat{\mu}_0(x)$$

We evaluated the causal effect of urban lockdowns on SI by evaluating the difference in the effect of stock prices on investor sentiment during an urban lockdown period and the control period. In the same way, we also assess the causal

Table 10
The impact of lockdowns on investor sentiment synergy.

	XGBoost	MLP	Mean
E_{Wuhan}	0.70 (0.53, 0.87)	0.39 (0.05, 0.73)	0.545
$E_{Shanghai}$	−0.77 (−0.97, −0.58)	−0.76 (−1.07, −0.46)	−0.765
E^R_{Wuhan}	0.22 (−0.02, 0.46)	0.08 (−0.26, 0.42)	0.15
$E^R_{Shanghai}$	−0.57 (−0.76, −0.37)	−0.55 (−0.87, −0.24)	−0.56

Note: The values in parentheses are the maximum and minimum values of the single sample impact. The mean column gives is the mean value of the treatment effect obtained by the two algorithms.

effect of a city's closure on synergy between investor sentiment and stock returns. To cover the period of lockdowns in Wuhan and Shanghai, we use the two periods before and after the city’s closure, p_{-1} and p_1, as the control groups and p_0 as the treatment group.

Table 10 shows the results of the causal inference ML analysis indicating the effect of the city's closure on SI. The average treatment effect (ATE) of closing Wuhan on SI is 0.545, which implies that investor sentiment has a higher positive causal association with stock prices during the city’s closure, consistent with the results in Section 4.1. The ATE of Shanghai’s closure on synergy between stock prices and investor sentiment is −0.765, which implies that this closure causes investor sentiment to have reverse synergy. Based on the range of values of investor sentiment before and after the city’s closure (see notes to Table 9), the impact of this closure on IS is significant at the time it occurs. The absolute value of the treatment effect demonstrates that the lockdown in Shanghai has a greater synergistic impact on investor sentiment, although it is not clear whether the reversal of sentiment is a reflection of the greater impact.

In addition, Table 10 shows that the city's closure greatly affects interaction between stock returns and investor sentiment. The average disposition effect of the closure of Wuhan is 0.15, indicating that the same stock returns lead to higher investor sentiment during this closure. However, the average disposition effect of the lockdown in Shanghai is −0.56, which leads investor sentiment to become counterintuitive and has a more significant disposition effect.

The results of this analysis confirm our hypothesis. Emotional anxiety among investors due to the city's closure is reflected on social media, with impacts on the stock market. Thus, it enhances interaction between the stock market and investor sentiment. In summary, the anxiety caused by the urban lockdowns can significantly affect the synergy between stock prices and investor sentiment and between stock returns and investor sentiment. This effect might promote the positive synergy of SI or to reverse the positive synergy of SI.

4.4. The Three Representations of Synergy

As stated in H1, investor sentiment has extremely significant positive synergy with stock prices statistically. However, in particular periods, the synergy between investor sentiment and stock prices may not be significant (e.g., p_{-2} in Fig. 18) and may even be reversed from positive to negative. In general, SI could have positive or negative synergy. In this section, we try to explain this phenomenon.

In this paper, we construct investor sentiment following Antweiler's method, illustrated in Fig. 19. Specifically, we construct the ICI index from the positive and negative sentiment expressed by investors on social media. So, does positive or negative sentiment promote positive or negative synergy between stock prices and investor sentiment? When stock prices as a whole are rising, some investors might lose confidence in the stock market and express untimely negative sentiment because of the decline of individual stocks. When stock prices fall as a whole, not all investors lose money or get frustrated. Social media is also full of lies with ulterior motives or rebellious sentiment due to anxiety. In short, social networks are full of disagreement. As shown in Table 11, when stock prices are rising, positive sentiment drives positive sentiment synergy, and negative sentiment drives negative sentiment synergy. When stock prices are falling, negative sentiment drives positive sentiment synergy, while positive sentiment drives negative sentiment synergy. So, stock price synergy with investor sentiment is not a single signal but, rather, is composed of symbiotic positive synergy and negative synergy. The combination of positive synergistic signals and negative synergistic signals gives SI a variety of possibilities in local time periods.

In general, a positive synergy signal has a greater impact than a negative synergy signal, and stock prices and investor sentiment have extremely significant positive correlation statistically. However, at a smaller time scale, synergy between stock prices and investor sentiment can have positive, negative, or no synergy under the combined effect of positive and negative synergistic signals. This finding helps to explain the contradiction documented by scholars that the effect of investor sentiment on stock market returns can be positive (Hengelbrock et al., 2013; Perez-Liston et al., 2016) or negative (Grigaliūnienė & Cibulskienė, 2010; Schmeling, 2009).

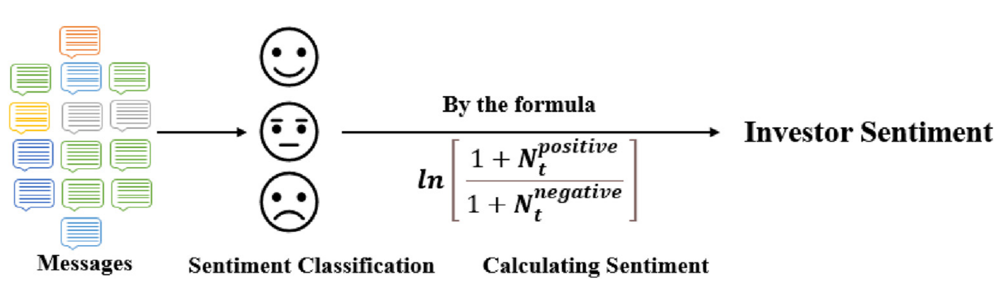

Fig. 19. How investor sentiment is created.

Table 11
The effect of sentiment on synergy.

	Stock prices rise	Stock prices fall
$N_t^{positive}$	1	0
$N_t^{negative}$	0	1

Notes: In the table, 1 = sentiment drives positive synergy, and 0 = sentiment drives negative synergy.

5. Conclusion

Using messages on social media by stock market investors and NLP techniques, we measure investor sentiment on social networks and investigate the basic patterns in synergy between stock prices and investor sentiment. We find that, in general, investor sentiment has a weak but extremely significant positive correlation with stock prices. However, in a specific periods (e.g., one or two months), the synergy between investor sentiment and stock prices be positive, negative, or neutral under the combined effect of positive and negative synergy signals. Many studies have discussed the positive and negative impacts of investor sentiment on current and expected future stock returns. Our findings help to explain the contradictory results in these studies.

Using a segmented measure of investor sentiment, we find that specific stock price indexes do not have higher synergy with real-time social media sentiment. This implies that high-frequency social media sentiment measures might not be useful for stock price prediction, which is consistent with the findings by Behrendt and Schmidt. We also find that pre-opening investor sentiment has greater sentiment synergy with stock prices. This suggests that social media is noisier after the opening and that investor sentiment on social media before the opening might be more relevant for stock price prediction. It also suggests that investor sentiment in the stock market is forward looking and contains ex ante information indicating that future prices are statistically significant, which is similar to Zhang's findings.

We also study changes in the synergistic relationship between stock prices and investor sentiment before and after the lockdowns in Wuhan and Shanghai. We measure the impact of these lockdowns on investor sentiment synergy with causal inference machine learning.

The Chinese government's closure of these two cities in response to the outbreak of COVID-19 had different effects on the synergy of investor sentiment. The closure of Wuhan led to positive synergy between stock prices and investor sentiment, as these investors were more likely to "tell the truth," whereas the closure of Shanghai had a negative effect, as the investors started to "tell lies." This finding helps to fill a gap in the literature on the effect of COVID-19 on investor sentiment in the stock market.

Conflict of interest

The authors have nothing to disclose.

References

Ali, H. (2018). *Twitter, investor sentiment and capital markets: What do we know?*. SSRN 3230304.

Antweiler, W., & Frank, M. Z. (2004). Is all that talk just noise? The information content of internet stock message boards. *The Journal of Finance, 59*(3), 1259–1294.

Baker, M., & Wurgler, J. (2006). Investor sentiment and the cross-section of stock returns. *The Journal of Finance, 61*(4), 1645–1680. https://doi.org/10.1111/j.1540-6261.2006.00885.x

Baker, M., & Wurgler, J. (2007). Investor sentiment in the stock market. *The Journal of Economic Perspectives, 21*(2), 129–152.

Bashir, M. F., Ma, B., & Shahzad, L. (2020). A brief review of socio-economic and environmental impact of Covid-19. *Air Quality, Atmosphere & Health, 13*(12), 1403–1409. https://doi.org/10.1007/s11869-020-00894-8

Behrendt, S., & Schmidt, A. (2018). The Twitter myth revisited: Intraday investor sentiment, Twitter activity and individual-level stock return volatility. *Journal of Banking & Finance, 96*, 355–367. https://doi.org/10.1016/j.jbankfin.2018.09.016

Bennett, G., Young, E., Butler, I., & Coe, S. (2021). The impact of lockdown during the COVID-19 outbreak on dietary habits in various population groups: A scoping review. *Frontiers in Nutrition, 8*. https://www.frontiersin.org/article/10.3389/fnut.2021.626432.

Bertrand, M., Duflo, E., & Mullainathan, S. (2004). How much should we trust differences-in-differences estimates. *Quarterly Journal of Economics, 119*(1), 249–275. https://doi.org/10.1162/003355304772839588

Berument, H., & Kiymaz, H. (2001). The day of the week effect on stock market volatility. *Journal of Economics and Finance, 25*(2), 181–193.

Boissay, F., & Rungcharoenkitkul, P. (2020). *Macroeconomic effects of Covid-19: An early review. No. 7*. Bank for International Settlement.

Brown, G. W., & Cliff, M. T. (2004). Investor sentiment and the near-term stock market. *Journal of Empirical Finance, 11*(1), 1–27.

Brown, G. W., & Cliff, M. T. (2005). Investor sentiment and asset valuation. *Journal of Business, 78*(2), 405–440.

Cao, S., Gan, Y., Wang, C., Bachmann, M., Wei, S., Gong, J., Huang, Y., Wang, T., Li, L., Lu, K., Jiang, H., Gong, Y., Xu, H., Shen, X., Tian, Q., Lv, C., Song, F., Yin, X., & Lu, Z. (2020). Post-lockdown SARS-CoV-2 nucleic acid screening in nearly ten million residents of Wuhan, China. *Nature Communications, 11*(1), 5917. https://doi.org/10.1038/s41467-020-19802-w

Cartenì, A., Di Francesco, L., & Martino, M. (2020). How mobility habits influenced the spread of the COVID-19 pandemic: Results from the Italian case study. *Science of the Total Environment, 741*, Article 140489. https://doi.org/10.1016/j.scitotenv.2020.140489

Cheadle, C., Cho-Chung, Y. S., Becker, K. G., & Vawter, M. P. (2003). Application of z-score transformation to Affymetrix data. *Applied Bioinformatics, 2*(4), 209–217.

Cox, J., Greenwald, D. L., & Ludvigson, S. C. (2020). *What explains the COVID-19 stock market? Working paper 27784*. National Bureau of Economic Research. https://doi.org/10.3386/w27784

De Chaisemartin, C., & d'Haultfoeuille, X. (2018). Fuzzy differences-in-differences. *The Review of Economic Studies, 85*(2), 999–1028.

Eachempati, P., Srivastava, P. R., & Panigrahi, P. (2021). Sentiment analysis of COVID-19 pandemic on the stock market. *American Business Review*. https://doi.org/10.37625/ABR.24.1.141-165

Fallahgoul, H. A. (2021). Inside the mind of investors during the COVID-19 Pandemic: Evidence from the StockTwits data. https://doi.org/10.3905/jfds.2021.1.058.

Fernandez-Perez, A., Gilbert, A., Indriawan, I., & Nguyen, N. H. (2021). COVID-19 pandemic and stock market response: A culture effect. *Journal of Behavioral and Experimental Finance, 29*, Article 100454. https://doi.org/10.1016/j.jbef.2020.100454

Finter, P., Niessen-Ruenzi, A., & Ruenzi, S. (2012). The impact of investor sentiment on the German stock market. *Zeitschrift für Betriebswirtschaft, 82*(2), 133–163.

Ge, Y., Qiu, J., Liu, Z., Gu, W., & Xu, L. (2020). Beyond negative and positive: Exploring the effects of emotions in social media during the stock market crash. *Information Processing & Management, 57*(4), Article 102218. https://doi.org/10.1016/j.ipm.2020.102218

Goldberg, Y., & Levy, O. (2014). word2vec explained: Deriving Mikolov et al.'s negative-sampling word-embedding method. *arXiv preprint arXiv*, Article 1402.3722.

Grigaliūnienė, Ž., & Cibulskienė, D. (2010). Investor sentiment effect on stock returns in Scandinavian stock market. *Economics and Management, 15*, 929–940.

Haleem, A., Javaid, M., & Vaishya, R. (2020). Effects of COVID-19 pandemic in daily life. *Current Medicine Research and Practice, 10*(2), 78–79. https://doi.org/10.1016/j.cmrp.2020.03.011

Hengelbrock, J., Theissen, E., & Westheide, C. (2013). Market response to investor sentiment. *Journal of Business Finance & Accounting, 40*(7–8), 901–917.

Hoang, T., & Syed, Q. (2021). Investor sentiment and volatility prediction of currencies and commodities during the COVID-19 pandemic. https://doi.org/10.46557/001C.18642.

Hu, J., Sui, Y., & Ma, F. (2021). The measurement method of investor sentiment and its relationship with stock market. *Computational Intelligence and Neuroscience, 2021*. https://doi.org/10.1155/2021/6672677

Kim, K., Ryu, D., & Yang, H. (2019). Investor sentiment, stock returns, and analyst recommendation changes: The KOSPI stock market. *Investment Analysts Journal, 48*(2), 89–101.

Kleinnijenhuis, J., Schultz, F., Oegema, D., & van Atteveldt, W. (2013). Financial news and market panics in the age of high-frequency sentiment trading algorithms. *Journalism, 14*(2), 271–291. https://doi.org/10.1177/1464884912468375

Kotsiantis, S. B., Zaharakis, I., & Pintelas, P. (2007). Supervised machine learning: A review of classification techniques. *Emerging artificial intelligence applications in computer engineering, 160*(1), 3–24.

Lee, W. Y., Jiang, C. X., & Indro, D. C. (2002). Stock market volatility, excess returns, and the role of investor sentiment. *Journal of Banking & Finance, 26*(12), 2277–2299.

Lemmon, M., & Portniaguina, E. (2006). Consumer confidence and asset prices: Some empirical evidence. *Review of Financial Studies, 19*(4), 1499–1529.

Liu, S. (2015). Investor sentiment and stock market liquidity. *The Journal of Behavioral Finance, 16*(1), 51–67.

Liu, Q., Zhou, X., & Zhao, L. (2022). View on the bullishness index and agreement index. *Frontiers in Psychology, 13*. https://doi.org/10.3389/fpsyg.2022.957323

Ljungqvist, A., Nanda, V., & Singh, R. (2006). Hot markets, investor sentiment, and IPO pricing. *Journal of Business, 79*(4), 1667–1702.

Mallikarjuna, M., & Rao, R. P. (2019). Evaluation of forecasting methods from selected stock market returns. *Financial Innovation, 5*(1), 1–16. https://doi.org/10.1186/s40854-019-0157-x

Maziarz, M. (2015). A review of the Granger-causality fallacy. *The Journal of Philosophical Economics: Reflections on Economic and Social Issues, 8*(2), 86–105.

McGurk, Z., Nowak, A., & Hall, J. C. (2020). Stock returns and investor sentiment: Textual analysis and social media. *Journal of Economics and Finance, 44*(3), 458–485.

Molarius, A., & Persson, C. (2022). Living conditions, lifestyle habits and health among adults before and after the COVID-19 pandemic outbreak in Sweden—results from a cross-sectional population-based study. *BMC Public Health, 22*(1), 171. https://doi.org/10.1186/s12889-021-12315-1

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. *Frontiers in Psychology, 12*, Article 626934.

Nimal, P. D. (1997). Relationship between Stock Returns and Selected Fundamental Variables; Evidence from Sri Lanka. *Sri Lankan Journal of Management, 2*(3), 268–286. https://doi.org/10.1186/s40854-019-0157-x

Paris, C., Christensen, H., Batterham, P., & O'Dea, B. (2015). Exploring emotions in social media. In *2015 IEEE Conference on Collaboration and Internet Computing (CIC)* (pp. 54–61). https://doi.org/10.1109/CIC.2015.43

Perez-Liston, D., Huerta, D., & Haq, S. (2016). Does investor sentiment impact the returns and volatility of Islamic equities? *Journal of Economics and Finance, 40*(3), 421–437.

Qiu, L., & Welch, I. (2004). *Investor sentiment measures*. National Bureau of Economic Research. https://doi.org/10.3386/w10794. Working Paper 10794.

Renault, T. (2017). Intraday online investor sentiment and return patterns in the U.S. stock market. *Journal of Banking & Finance, 84*, 25–40. https://doi.org/10.1016/j.jbankfin.2017.07.002

Sayim, M., & Rahman, H. (2015). The relationship between individual investor sentiment, stock return and volatility: Evidence from the Turkish market. *International Journal of Emerging Markets, 10*(3), 504–520. https://doi.org/10.1108/IJoEM-07-2012-0060

Schmeling, M. (2009). Investor sentiment and stock returns: Some international evidence. *Journal of Empirical Finance, 16*(3), 394–408.

Stieglitz, S., & Dang-Xuan, L. (2013). Emotions and information diffusion in social media—sentiment of microblogs and sharing behavior. *Journal of Management Information Systems, 29*(4), 217–248. https://doi.org/10.2753/MIS0742-1222290408

Sun, L., Najand, M., & Shen, J. (2016). Stock return predictability and investor sentiment: A high-frequency perspective. *Journal of Banking & Finance, 73*, 147–164. https://doi.org/10.1016/j.jbankfin.2016.09.010

Sun, L., & Shi, W. (2022). Investor sentiment and stock market reactions to COVID-19: Evidence from China. *Discrete Dynamics in Nature and Society, 2022*, Article e8413916. https://doi.org/10.1155/2022/8413916

Taylor, R. (1990). Interpretation of the correlation coefficient: A basic review. *Journal of Diagnostic Medical Sonography, 6*(1), 35–39. https://doi.org/10.1177/875647939000600106

Utz, S. (2019). Utz, S. (2019). Social media as sources of emotions. In *Social psychology in action* (pp. 205–219). Cham: Springer. https://doi.org/10.1007/978-3-030-13788-5_14.

Wagner, A. F. (2020). What the stock market tells us about the post-COVID-19 world. *Nature Human Behaviour, 4*(5), 440. https://doi.org/10.1038/s41562-020-0869-y

Wang, X., Xiang, Z., Xu, W., & Yuan, P. (2022). The causal relationship between social media sentiment and stock return: Experimental evidence from an online message forum. *Economics Letters, 110598*. https://doi.org/10.1016/j.econlet.2022.110598

Xing, F., Hoang, D. H., & Vo, D.-V. (2020). High-frequency news sentiment and its application to forex market prediction. SSRN Scholarly Paper 3711227 https://papers.ssrn.com/abstract=3711227.

Xiong, X., Luo, C., & Ye, Z. (2017). Stock BBS and trades: The information content of stock BBS. *Journal of Systems Science and Mathematical Sciences, 37*(12), 2359.

Zhang, G., Xu, L., & Xue, Y. (2017). Model and forecast stock market behavior integrating investor sentiment analysis and transaction data. *Cluster Computing, 20*(1), 789–803. https://doi.org/10.1007/s10586-017-0803-x