Proceedings of ICFTBA 2025 Symposium: Data-Driven Decision Making in Business and Economics
DOI: 10.54254/2754-1169/2025.BL30472

The Influence of Social Media on the Investment Behavior of Stock Investors: An Empirical Analysis Based on Behavioral Finance

Xu Jingchun Water

Nanhang Jingcheng College, Nanjing, China
2471419363@qq.com

Abstract. With the vigorous development of digital finance and the widespread popularity of social media, platforms such as Xueqiu and East Money Stock Bar have become core channels for investors to obtain information and exchange views, profoundly changing the information flow pattern of financial markets and the decision-making environment for investors. Based on the perspective of behavioral finance, this paper focuses on the impact of social media on investors' investment behaviors and conducts an empirical analysis. By collecting 86,420 pieces of discussion data from Xueqiu and East Money Stock Bar between January and June 2024 using Python crawlers, combined with the trading records and characteristic questionnaires of 120 investors, this study employs methods including descriptive statistics, correlation analysis, regression analysis, and mediating effect test to explore the influence of social media discussion popularity and emotional tendency on investors' trading frequency, position change rate, and investment return rate, clarify the action paths, and analyze the behavioral differences among investors with different investment experience. The results show that social media discussion popularity significantly increases investors' trading frequency and position change rate, emotional polarity indirectly affects investment return rate through behavioral biases, and investment experience plays a moderating role in the impact of discussion popularity. The conclusions of this paper provide practical implications for investors' rational decision-making, platform governance optimization, and regulatory authorities' risk prevention and control, as well as empirical support for research on the interaction between social media and financial markets.

Keywords: Social Media, Investment Behavior, Behavioral Finance, Sentiment Polarity, Moderating Effect

1. Introduction

With the rapid development of digital finance, social media has become the core platform for information dissemination, profoundly transforming the information flow structure in financial markets and reshaping investors decision-making environments. Globally, over 50 million finance-related social media posts are generated daily, with Twitter's financial tweets growing at an average

DOI: 10.54254/2754-1169/2025.BL30472

daily rate of 11.3%. Platforms like Xueqiu and East Money Stock Bar have emerged as primary channels for investors to exchange insights and access information. Meanwhile, the proliferation of zero-commission trading platforms has lowered market entry barriers, with retail investors' trading volume rising from 10% in 2010 to 23% in 2023. (These data are derived from relevant industry reports and public statistics, which are not specifically cited in the text and belong to commonly quoted industry data.). The social media-driven "memes stock" phenomenon has further disrupted traditional valuation logic, making the interactive relationship between social media and stock markets a focal point in financial research.

Traditional financial theory, grounded in the Efficient Market Hypothesis (EMH) of the "rational investor" assumption, posits that information is swiftly priced into stock prices, making it difficult to achieve excess returns through public information. However, behavioral finance research reveals that investors exhibit irrational traits such as cognitive biases and emotion-driven decision-making. The information dissemination characteristics of social media may amplify these irrational behaviors, leading to market anomalies like overreactions or delayed adjustments. In this context, exploring how social media influences stock investors' decision-making not only enriches the application scenarios of behavioral finance in the digital era, but also provides practical references for rational investment strategies and optimized market regulation.

Based on the above background, this paper focuses on the following core research questions.

1. How do the popularity and emotional tendency of social media discussions affect the trading decisions of investors (such as trading frequency, position changes, risk appetite)?

2. How do social media affect the investment behavior of investors through which paths (such as information dissemination, emotional contagion, and amplification of behavioral deviation)?

3. Are there significant differences in the behavior of investors with different characteristics (such as investment experience and risk tolerance) under the influence of social media?

2. Literature review

2.1. Core theoretical support

The first is the Behavioral Finance Theory. Breaking through the rational man hypothesis, the Prospect Theory proposed by Kahneman [1] and Tversky [1] reveals the laws of people's deviation from rationality in risky decision-making, and Thaler's [2] Mental Accounting Theory also confirms that investors' decisions are influenced by psychological factors. This theory holds that investors are affected by cognitive biases and emotional factors, leading to irrational behaviors such as overconfidence and disposition effect, and group interactions on social media will further strengthen such biases.

The next is the Information Asymmetry Theory. Akerlof's Lemon Market Theory lays the core framework of this theory, pointing out that information gaps in the market will lead to market imbalances. Social media reduces information acquisition costs, accelerates information dissemination, and changes the traditional information asymmetry structure. However, non-professional investors have limited ability to identify information, which instead exacerbates decision-making biases due to information overload and uneven quality.

The last one is the Limited Attention Theory. Herbert Simon [3] proposed that attention is scarce and a core resource in the information age. Gabaix et al.'s [4] Directed Cognition Model further points out that individuals will rationally allocate attention but achieve poor results due to rational limitations. Algorithm recommendations on social media will forcibly change investors' attention allocation, thereby affecting their judgments on investment targets and trading timing.

DOI: 10.54254/2754-1169/2025.BL30472

2.2. Relevant studies

There is existing research on the correlation between social media and the stock market. For example, Bollen et al. [5] analyzed Twitter sentiment to predict stock market fluctuations, confirming the correlation between social media sentiment and market performance. In addition, it includes industry research from platforms such as CSDN Blog and Douyin (e.g., A New Favorite in Quantitative Investment: Research on the Relationship between Social Media Discussion Heat and Stock Returns, Core Theories and Market Applications of Behavioral Finance). Another example is that Rao Yupei et al. [6] point out in Behavioral Finance that there is a significant correlation between social media popularity and stock returns, with a more pronounced impact on small and mid-cap stocks and retail investors. However, existing studies mostly focus on the macro-level market perspective, with insufficient exploration of the micro-level influence mechanisms on individual investors and a lack of detailed heterogeneous analysis of investors with different characteristics. This gap serves as the research focus of this paper.

The research on the application of sentiment analysis technology in financial texts covers a series of studies from 2011 to 2015 that introduced natural language processing technology into the financial field and constructed sentiment prediction models.

Since 2021, research has focused on the phenomenon of "meme stocks" and explored cutting-edge research on the impact of multi-modal data such as images and videos on investment behavior.

The main findings from the literature are as follows. On the one hand, there is a significant correlation between social media discussion heat and stock returns, especially for small and medium-cap stocks and stocks with high participation of retail investors. On the other hand, existing studies mostly focus on price fluctuations at the market level, with insufficient discussion on the micro-mechanisms influencing individual investors' trading behaviors, and lack detailed analysis of the heterogeneous impacts on different investors. This constitutes the research entry point of this paper.

Research on the relationship between social media and stock markets has evolved through four distinct phases(The stage division of the relationship between social media and the stock market in this study is summarized based on reviewing Bollen (2011), industry reports, such as research from CSDN Blog and Douyin platform, and cutting-edge meme stock research, combined with the evolutionary characteristics of research methods and application scenarios.): The initial exploration phase (2006-2010) identified correlations between social media sentiment and stock market indices. The methodological development phase (2011-2015) applied sentiment analysis techniques to financial texts, establishing preliminary predictive models. The empirical validation phase (2016-2020) verified these correlations through large-scale data analysis while distinguishing the impacts of different emotional types. The application deepening phase (2021-present) focuses on meme-driven stock phenomena, exploring the integration of multi-modal data and real-time trading strategies.

Existing research has demonstrated a significant correlation between social media discussion intensity and stock returns, particularly evident in small and mid-cap stocks and those with high retail investor participation. However, current studies predominantly focus on market-level price fluctuations, with insufficient exploration of micro-level mechanisms influencing individual investor behavior. Moreover, there is a lack of detailed analysis regarding the heterogeneous impacts on different investor groups. These gaps constitute the research focus of this paper.

DOI: 10.54254/2754-1169/2025.BL30472

3. Research design and methods

3.1. Rationale of research design

The rationality of this research design is mainly reflected in the following three dimensions:

Theoretical Consistency is Supported by behavioral finance theory, information asymmetry theory, and limited attention theory, it closely focuses on the research topic of "the impact of social media on individual investors' investment behavior", ensuring a solid theoretical foundation for the research framework. For example, the "cognitive bias" and "herd effect" theories in behavioral finance can effectively explain how social media amplifies irrational trading among individual investors; the information asymmetry theory can illustrate that while social media reduces information acquisition costs, it may exacerbate decision-making biases due to individual investors' insufficient information screening capabilities.

The next is Methodological Adaptability. Adopting a process of "data collection - processing - multi-dimensional analysis" and combining quantitative research methods (descriptive statistics, correlation analysis, regression analysis, mediating effect test), it can systematically and accurately explore the causal relationships and action paths among variables. For instance, Pearson correlation coefficient can clarify the linear relationship between social media heat, sentiment polarity, and individual investors' trading behavior; multiple linear regression can isolate the independent impacts of different variables; and the mediating effect test can reveal the internal mechanism of action.

The third one is Practical Relevance. Focusing on the micro-level of individual investors' investment behavior, it supplements the deficiency of existing research that "focuses more on market-level analysis and less on individual mechanisms"; meanwhile, it pays attention to investor heterogeneity (such as the moderating role of investment experience and risk tolerance), making the research conclusions more practically guiding and able to provide differentiated insights for individual investors, platforms, and regulatory authorities.

3.2. Explanation of sample and data collection

3.2.1. Sample used

This study uses a multi-source integrated sample, including Social media data sample, which is a kind of discussion data of individual investors on Xueqiu and East Money Stock Bar from January to June 2024, with an effective sample size of 86,420 items; The second one is Individual investors' trading behavior sample, which is concurrent trading records of 120 individual investors (obtained through a combination of simulated trading platforms and questionnaire surveys); The third one is Individual investors' characteristic sample, which contains Demographic and behavioral characteristic information of 120 individual investors, such as investment years, risk tolerance, and investment knowledge level (collected through questionnaires).

3.2.2. Data collection process

Social media data collection includes crawling user comments, reposts, likes, topic participation, and other data related to stocks on Xueqiu and East Money Stock Bar using Python crawler tools, with a time window from January 1, 2024, to June 30, 2024.

Another is the collection of individual investors' trading data and characteristic data. Cooperate with simulated trading platforms, invite users to participate voluntarily and authorize access to trading records; meanwhile, design structured questionnaires (including modules such as trading

DOI: 10.54254/2754-1169/2025.BL30472

behavior, personal characteristics, and social media usage habits) and distribute them through online channels (such as financial forums, investor communities), with a total of 120 valid questionnaires collected.

3.2.3. Data cleaning and integration

For social media data, use regular expressions to filter invalid comments such as meaningless characters and advertisements, and perform text word segmentation and sentiment analysis through the NLTK library; For questionnaire data, eliminate invalid questionnaires with incomplete filling and logical contradictions (finally retaining 120 valid samples); Use "user ID + time window" as the matching key to associate social media interaction data, trading data, and individual investors' characteristic data to form an integrated dataset.

3.2.4. Survey time and sample validity

The questionnaire was distributed and collected from July 1, 2024, to July 31, 2024. The data collection period is consistent with that of social media data and trading data (January-June 2024) to ensure temporal consistency. After cleaning, the effective sample of social media data is 86,420 items (after eliminating invalid comments); a total of 120 questionnaires were collected for individual investors' trading and characteristic data, and all were valid samples after validity testing (such as logical consistency and information integrity), with no results eliminated.

3.2.5. Data collection

Social media data is collected by using Python web crawlers. We collected investor discussion data from Xueqiu and East Money Stock Forum (January-June 2024), including stock-related comments, reposts, likes, and topic engagement, yielding 86,420 valid samples. The second one is stock trader transaction data. Through the combination of a simulated trading platform and a voluntary questionnaire survey of stock traders, the transaction records of 120 stock traders in the same period were collected, including structured data such as stock code, transaction time, transaction volume, position change, initial capital and final return. The last one is investor characteristics, which means collecting respondents' demographic and behavioral characteristics such as investment years, risk tolerance, investment knowledge level, and social media usage frequency through questionnaires.

3.2.6. Data processing

According to the Table 1, social Media Data Cleaning includes Utilizing regular expressions to filter out invalid comments (e.g., nonsensical characters, advertisements) and employing Python's NLTK library for text segmentation and sentiment analysis, converting text content into sentiment polarity values (standardized to the [-1,1] range, where-1 indicates completely negative and 1 indicates completely positive). Construct the Social Media Discussion Heat Index (SMDH) using the formula: $SMDH\_i(t) = \alpha \cdot N\_i(t) + \beta \cdot D\_i(t) + \gamma \cdot I\_i(t)$, where $N\_i(t)$ represents the number of discussions for stock i during time period t, $D\_i(t)$ denotes dissemination depth, $I\_i(t)$ indicates interaction intensity, and α, β, γ are weights determined through the Analytic Hierarchy Process (AHP) with values of 0.4, 0.3, and 0.3 respectively.

Data fusion construction means Using "user ID + time window" as the matching key, social media interaction data, transaction data and investor characteristic data are correlated to form a fusion data set containing "social media behavior-investor characteristics-transaction decisions".

DOI: 10.54254/2754-1169/2025.BL30472

Table 1. Definition of research variables

Variable Type	Variable Name	Definition and Measurement Method
Dependent Variable	Trading Frequency	Monthly number of trades (times)
	Position Change Rate	Monthly proportion of changes in the types of stocks held(%)
	Investment Return Rate	Monthly ratio of actual return to initial capital(%)
Independent Variable	Social Media Discussion Heat	Comprehensive index calculated based on the aforementioned formula
	Social Media Sentiment Polarity	Standardized value derived from text sentiment analysis
Moderating Variable	Investment Experience	Investment years (Less than 1 year=1,1-3 years=2,More than 3 years=3)
	Risk Tolerance	Questionnaire score(1-5 points, higher score indicates stronger tolerance)
Control Variable	Gender,Age,Investment Knowledge Level	Gender(Male=1,Female=0); Age(grouped variable); Questionnaire score(1-5 points)

4. Empirical results and analysis

4.1. Descriptive statistics

Looking at the table 2, among the 120 investors surveyed, 35% have less than one year of investment experience, 45% have 1-3 years, and 20% have more than 3 years. 72% have moderate or lower risk tolerance, and 68% spend more than one hour a day using social media to obtain financial information.

Social media data shows that the average monthly discussion volume of popular stocks reaches 1,280, with an average emotional polarity of 0.12, indicating an overall optimistic trend. Investors trade an average of 4.3 times per month, with a 27.6% average holding change rate and 1.8% average investment return rate, showing significant individual differences.

The more investment experience investors have, the weaker the impact of social media discussion popularity on trading frequency. This is consistent with Kahneman's [7] analysis in "Thinking, Fast and Slow" regarding behavioral differences among decision-makers at different cognitive levels.

DOI: 10.54254/2754-1169/2025.BL30472

Table 2. Correlation analysis results

Statistical Dimension	Specific Category/Indicator	Value/Proportion
Investment Tenure Distribution	Less than 1 year	35%
	1-3 years	45%
	More tan 3 years	20%
Risk Tolerance	Medium or below	72%
Social Media Usage	Daily time spent obtaining Financial information is more than 1 hour	68%
Social Media Data	Average monthly discussion volume of popular stocks	1,280 posts
	Average sentiment polarity ([-1,1])	0.12(overall optimistic)
Investors' Trading Behavior	Average monthly trading frequency	4.3 times
	Average monthly position change rate	27.6%
	Average monthly investment return rate	1.8%

4.2. Results of correlation analysis

According to the table 3, the correlation analysis reveals three key relationships: (1) Social media discussion intensity shows a strong positive correlation with stock trading frequency ($r=0.63$, $p<0.01$) and position turnover rate ($r=0.57$, $p<0.01$), while exhibiting a weak negative correlation with investment returns ($r=-0.12$, $p>0.05$); (2) Social media sentiment polarity demonstrates a significant positive correlation with investment returns ($r=0.31$, $p<0.05$), meaning investors exposed to more positive information tend to achieve higher monthly returns.

The impact of social media on the market requires regulatory oversight. Relevant regulatory measures can refer to the practical directions such as digital monitoring and investor education proposed by iFinD [8] in its paper An Analysis of the Relationship Between Social Media Sentiment and Capital Market Regulation.

Table 3. Main effect regression analysis results

Variable Combination	Pearson Correlation Coefficient(r)	P-value	Correlation Conclusion
Social Media Discussion Heat vs. Trading Frequency	0.63	<0.01	Significantly positive correlation
Social Media Discussion Heat vs. Position Change Rate	0.57	<0.01	Significantly positive correlation
Social Media Discussion Heat vs. Investment Return Rate	-0.12	>0.05	Negative correlation (not significant)
Social Media Sentiment Polarity vs. Investment Return Rate	0.31	<0.05	Significantly positive correlation

DOI: 10.54254/2754-1169/2025.BL30472

4.3. Regression analysis results

The results of the multiple linear regression (Table 1) indicate that, after controlling for variables such as gender and age, the intensity of social media discussions significantly positively affects both trading frequency ($\beta=0.58$, $p<0.001$) and position adjustment rate ($\beta=0.51$, $p<0.001$). This suggests that more active social media discussions correlate with higher trading frequency and more frequent position adjustments among investors. Additionally, social media sentiment polarity shows a significant positive correlation with investment returns ($\beta=0.29$, $p<0.01$), while having no statistically significant impact on trading frequency or position adjustment rate. (referring the Table 4)

Table 4. Mediating effect regression analysis results

Dependent Variable	Independent Variable	Coefficient (β)	t-value	p-value	Model R^2
Trading Frequency	Social Media Discussion Heat	0.58	7.24	0.000	0.41
	Social Media Sentiment Polarity	0.08	1.03	0.305	
Position Change Rate	Social Media Discussion Heat	0.51	6.89	0.000	0.36
	Social Media Sentiment Polarity	0.06	0.87	0.386	
Investment Return Rate	Social Media Discussion Heat	-0.09	-1.12	0.267	0.15
	Social Media Sentiment Polarity	0.29	3.45	0.001	

4.4. Moderation effect test

According to the Table 5, when investment experience was introduced as a moderating variable, the interaction term between social media discussion heat and investment experience significantly influenced trading frequency ($\beta=-0.23$, $p<0.05$). This indicates that more experienced investors are less affected by social media discussions and exhibit more rational trading behavior. However, the moderating effect of risk tolerance was not significant ($p>0.05$).

The mediation analysis reveals that information dissemination speed partially mediates the relationship between social media discussion intensity and trading frequency (indirect effect value: 0.18,95% confidence interval excluding 0). Specifically, heightened social media discussions accelerate information spread, thereby improving investors' attention allocation and increasing trading frequency. Meanwhile, behavioral biases (e.g., herd behavior) fully mediate the relationship between social media emotional polarity and investment returns. This indicates that positive emotions indirectly enhance investment returns by reducing irrational behavioral deviations.

This finding corroborates Shiller's [9] assertion in Inefficient Markets: An Introduction to Behavioral Finance regarding cognitive biases mediating the relationship between market information and investment behaviors.

DOI: 10.54254/2754-1169/2025.BL30472

Table 5. Mediating effect test results

Mediating Path	Mediating Variable	Indirect Effect Value	95%Confidence Interval	Mediating Effect Conclusion
Social Media Discussion Heat → Trading Frequency	Information Dissemination Speed	0.18	[0.06,0.31] (excluding 0)	Partial mediating effect
Social Media Sentiment Polarity→ Investment Return Rate	Behavioral Bias (e.g., Herd Behavior)	0.22	[0.08,0.35] (excluding 0)	Full mediating effect

5. Conclusion

Social media discussions significantly enhance stock traders' trading frequency and position turnover rates, yet show no direct impact on investment returns. This indicates that frequent interactions may intensify short-term trading behaviors without necessarily boosting profits. Positive sentiment on social media significantly improves investment returns, while negative emotions could trigger irrational sell-offs.

The influence of social media on investors' investment behavior includes two main paths: "information dissemination-attention allocation-trade decision" and "emotional contagion-behavior deviation-return performance", which correspond to the cognitive and emotional level of the mechanism respectively.

Investment experience has a moderating effect. New investors are more likely to be influenced by the popularity of social media discussions, while experienced investors rely more on their own experience and rational judgment in trading decisions, showing a stronger ability to resist interference

For investors, they approach financial information on social media rationally, avoiding impulsive trading driven by trending topics and minimizing short-term speculation. Develop the ability to discern information, distinguishing factual statements from emotional expressions. Formulate decisions based on personal investment experience and risk tolerance, steering clear of blind herd behavior.

For the platform, it should strengthen the audit of financial information, crack down on false information and malicious guidance, optimize the algorithm recommendation mechanism, balance the diversity and professionalism of information, and provide investors with a better information environment; it can add an investment education section to enhance users' rational investment awareness.

For regulators, they should pay attention to the abnormal market fluctuations caused by social media, establish a monitoring mechanism for "memorandum stocks", improve the information disclosure system, prevent the market risks caused by the spread of false information, and maintain the stability of the financial market.

The limitations of this study include a small sample size (120 stock investors) and a six-month data period, which may affect the universality of conclusions. Additionally, the research did not account for differences in characteristics across social media platforms. Future studies could expand the sample size, extend the data period, and conduct comparative analysis using multi-platform data. Furthermore, incorporating machine learning models could help develop more accurate social media influence prediction models, providing valuable references for quantitative investment strategies.

DOI: 10.54254/2754-1169/2025.BL30472

References

[1] Kahneman, D., & Tversky, A.(1979).Prospect Theory: An Analysis of Decision under Risk [J].Econometrica, 47(2): 263 - 291.
[2] Thaler, R.H.(1980).Toward a positive theory of consumer choice [J].Journal of Economic Behavior & Organization, 1(1): 39 - 60.
[3] Simon, H. A. (1971). Designing organizations for an information-rich world. In M. Greenberger (Ed.), Computers, Communication, and the Public Interest (pp. 37-52). The Johns Hopkins Press.
[4] Gabaix, X., & Laibson, D. (2005). Bounded rationality and directed cognition (Working Paper). Harvard University & MIT.
[5] Bollen, J., Mao, H., & Zeng, X.(2011).Twitter mood predicts the stock market [J].Journal of Computational Science, 2(1): 1 - 8.
[6] Rao, Y., Peng, D., & Sheng, H. (2020). Behavioral Finance (2nd ed.). Beijing: China Machine Press.
[7] Kahneman, D. Thinking, Fast and Slow [M]. Beijing: CITIC Press Group, 2012.
[8] iFinD. An Analysis of the Relationship Between Social Media Sentiment and Capital Market Regulation [EB/OL].https: //m.10jqka.com.cn/20250519/c668250863.shtml, 2025-05-19.
[9] Shiller, R. J. Inefficient Markets: An Introduction to Behavioral Finance [M]. Beijing: China Renmin University Press, 2016.