Empirical Study on Influences of Investors' Cognitive Differences on Stock Returns — Based on Text Information of "Taoguba"

Zeyu Chen

Department of Financial Technology, Inner Mongolia University of Finance and Economics, Hohhot, Inner Mongolia, China

Abstract: This paper constructs a unique indicator of domestic individual investors' cognitive differences using "Taoguba" forum text data and the Naive Bayes method. Through Granger causality tests, instantaneous Granger causality tests, and intertemporal regression analysis, it examines whether cognitive differences influence or predict stock returns. Empirical results show cognitive differences impact current stock returns but lack predictive ability. Non-trading hour forum posts predict opening prices, while trading hour posts significantly influence closing prices and daily yields. This study enhances understanding of online trader behavior and its impact on stock returns, offering insights into market dynamics and investor sentiment.

***Keywords*:** text information analysis, investors' cognitive differences, stock returns, group behavior

1. Introduction

Traditional financial economics has evolved from single-factor models like CAPM to multi-factor models, such as the Fama-French three-factor and five-factor models, typically treating stock return factors as exogenous. Behavioral finance, however, emphasizes endogenous influences, such as investor emotions and concerns, on stock returns (Zhang, 2018)[1]. This paper explores the relationship between individual investors' cognitive differences and stock returns, focusing on China's stock market, where individual investors dominate in number and trading volume. By analyzing cognitive differences, this research provides practical insights into how individual investor behavior impacts stock returns, contributing to understanding market dynamics in a retail-driven context.

The paper may be innovative in the following aspects: (1) taking attempts to break through the obstacles of insufficient data mining of online text information at present, and using text mining methods and new logical analysis methods to discuss the influence of localized investors' cognitive differences on stock returns; (2) using Granger test, instantaneous causality test and intertemporal regression analysis to verify influences of investors' cognitive differences on stock returns; (3) using SVM, Light GBM, TF-IDF and other interdisciplinary methods such as applied statistics, computer autonomous learning and big data analysis to construct a more reasonable logical analysis model and studying the cognitive differences of investors.

2. Literature Review and Research Hypotheses

The definitions of investors' cognitive differences in the existing literature are not the same. The concept of investors' cognitive differences contains some basic connotations. Specifically, the process of investors' receiving and analyzing market information will lead to differences due to investors' cognitive differences. Then, it results in the heterogeneity of investment behaviors. By reviewing and combing, the existing specialized literature can be roughly divided into the following two categories.

2.1 Influences of Investors' Cognition on Stock Returns

Research on investors' cognition in China began relatively late. Pei (2004)[2] used questionnaire surveys to confirm widespread cognitive bias among Chinese investors. Liu (2016)[3] found that increased cognition reduces market liquidity, and investors often lack sufficient understanding of new information. This paper argues that cognitive differences and stock returns interact: high returns reduce cognitive differences, boosting investor expectations and shareholding, while low returns increase cognitive differences, worsening expectations and prompting divergent behaviors like holding or selling, further influencing stock returns. This dynamic highlight the interplay between investor cognition and market performance in China's stock market. Accordingly, this paper puts forward hypothesis 1: Investors' cognitive differences have influences on stock returns.

2.2 Prediction of Stock Returns by Investors' Cognition

Todd (2000)[4] found excessive analysis of existing information reduces attention to new information, hindering timely responses to price changes. Schmeling (2009)[5] and Ylva (2018)[6] revealed individual investors rely on institutional analysis and past data, showing weak stock prediction abilities. Jia (2016)[7] and Cui (2017)[8] noted domestic investors' improved cognition doesn't enhance return prediction or market fluctuation recognition. This paper argues that investors' inherent and acquired limitations in information processing lead to inaccurate future earnings judgments. Seeking more information in online forums amplifies cognitive differences among investors, perpetuating market influence. Accordingly, this paper puts forward hypothesis 2: Investors' cognitive differences cannot predict market returns.

3. Investors' Cognitive Indicators Based on Text Information and Text Cognitive Classification Model

3.1 Investors' Cognitive Indicators based on Text Information

Concerning the method used by Antweiler (2004)[9]: constructing an analysis indicator by network text information, namely:

$$B_t = \frac{M_t^{pos} - M_t^{neg}}{M_t^{pos} + M_t^{neg}} \tag{1}$$

Where: $M_t^c = \sum_{i \in D(t)} w_i x_i^c$ is the sum of the weighted numbers of messages of type $C \in \{pos, neu, neg\}$ in a period of time D(t). x_i^c is an indicator variable, indicating the source type of information. The bullish indicator B_t of investors is between -1 and 1, which expresses the relative bullish degree of investors. Antweiler (2004)[9] also defined another indicator:

$$B_t^* = \ln[\frac{1+M_t^{pos}}{1+M_t^{neg}}] \tag{2}$$

In addition, $B_t^* \approx B_t \ln(1+\left(M_t^{pos} + M_t^{neg}\right))$ was given. As believed, B_t^* considers the degree and quantity of posts at the same time, which can better reflect the characteristics of indicators. It is better than that when B_t^* only considers emotions, but does not consider the text information contained in the post. This paper holds that investor posting is a form of investor cognition, and the text information contained in the post is valuable. In view of this, this paper puts forward the investors' cognitive difference index based on text information, B_t^{ICD}:

$$B_t^{ICD} = B_t \ln(1+M_t) \tag{3}$$

Where: $M_t = M_t^{APP} + M_t^{DAP} + M_t^{SID}$ is the degree of investors' cognitive differences expressed by the total number of posts. M_t^{APP} is the number of posts that investors agree with. M_t^{DAP} is the number of posts that investors disagree with. M_t^{SID} is the number of posts that investors are neutral about. To reduce the extremely irrational information in text information, this paper standardizes the moving average of the cognitive difference indicator and constructs the investor cognitive difference index:

$$ABN\left(S_t\right) = \frac{S_t - average\left(S_{(t-k,t-1)}\right)}{average\left(S_{(t-k,t-1)}\right)} \tag{4}$$

Where: $ABN\left(S_t\right)$ is the investors' cognitive difference index. S_t is the investors' cognitive difference, which indicates the difference degree of individual investors' cognitive differences. At the same time, considering the cognitive differences of individual investors, the moving average $average\left(S_{(t-k,t-1)}\right)$ is selected to analyze the differences between extremely irrational posts. To distinguish the periodicity of cognitive differences, different time units are selected. Different trading periods are set: k=12 (month); k=52 (weeks); k=252 (days).

3.2 Classification Model of Cognitive Expectation based on Naive Bayes

The cognitive classification of texts mainly uses dictionaries and classifiers. Chinese has free expression grammar and few linear features, and thus it is not suitable for using the text dictionary method. Therefore, this paper uses a classifier to classify, that is, Naive Bayesian Model (NBM). The model is simple and robust. It still has high training and use efficiency on massive investor data. Specifically, firstly, the information in this paper is classified and part-of-speech filtered. Only words with large information are reserved, which are expressed as $d=\left(w_1,\cdots,w_i,\cdots,w_n\right)$, where w_i is the i-th non-empty eigenvalue. Its experience distribution in the class C_j is as follows:

$$p(w_i \mid c_j)=\frac{TF\left(w_i,c_j\right)+1}{\sum_q\left(TF\left(w_i,c_j\right)+1\right)} \tag{5}$$

Where: $TF\left(w_i,c_j\right)$ is the number of times the feature w_i appears in the class c_j. Laplacian smoothing is used in $\sum_q\left(TF\left(w_i,c_j\right)+1\right)$ for smoothing processing. Therefore, the probability that c belongs to the class c_j can be estimated as:

$$p(c_j \mid d)=p\left(c_j\right)p(d \mid c_j)/p\left(d\right)\propto p\left(c_j\right)\prod_{i=1}^{n}p(w_i \mid c_j) \tag{6}$$

Therefore, the final category of text information d is:

$$C_d=\underset{c_j\in c}{arg}\max\{p\left(c_j\right)\prod_{i=1}^{n}p(w_i \mid c_j)\} \tag{7}$$

The experimental data of this paper was collected from Taoguba. To prepare the training data set, 10000 pieces of investor data filtered by garbage were randomly extracted from the database and then labeled cognitively. Three kinds of classification models were established. A 5-fold cross validation was adopted, as shown in Fig. 1.

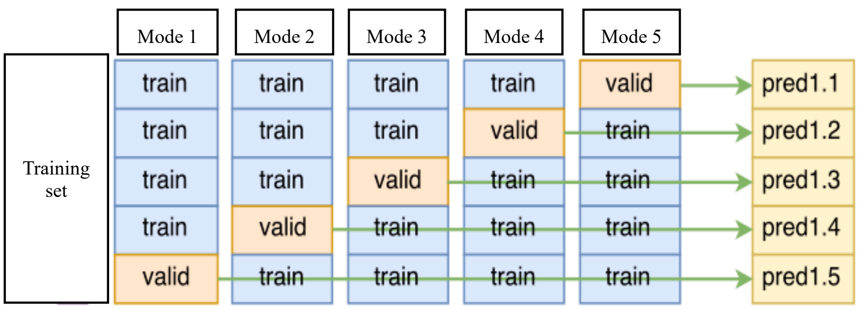

Figure 1. Cross Validation

After cross validation, the results are shown in Table 1. The accuracy of measuring investors' cognitive differences could reach 84.45%. Compared with some foreign literature that used the Naive Bayes method to analyze the accuracy of autonomous learning, the method used in this paper and the result were relatively more reliable. For example, Das (2007) [10] used the Naive Bayes method to classify Yahoo investor data, while the test accuracy was 50%. The sample test results showed that the average recall rate of approval and disapproval was 72.3%. Kim (2014)[11] classified investor data into positive and negative categories under the assumption that all investors had been eliminated, while the classified recall rate was 62.7%.

Table 1. Cross Validation

Cross Validation	APP			NAP			DIS		
	p (%)	R (%)	F (%)	p (%)	R (%)	F (%)	p (%)	R (%)	F (%)
1	67.37	89.33	76.81	79.88	75.84	77.81	85.52	63.93	73.16
2	70.37	85.93	77.38	81.52	76.92	79.16	82.32	68.53	74.79
3	63.11	84.15	72.13	80.66	82.64	76.44	80.92	63.73	71.33
4	69.51	83.78	75.98	76.54	75.27	75.39	87.14	69.71	77.46
5	71.19	90.32	79.62	83.03	78.29	80.57	86.35	66.28	75.31
Average	68.31	86.72	76.38	80.33	85.79	77.98	84.45	66.43	74.36

Note: p-accuracy, r-recall rate, f=2pr/(p+r), all of which are positive indicators.

4. Data and Research Method

4.1 Data and Statistical Description

In this paper, the stocks of listed companies in Shanghai Stock Exchange were selected for the whole description of the stock market. Monthly, weekly and daily frequency data of 2 million main posts and more than 20 million replies from November 9, 2007 to January 21, 2023 was selected as research samples. To correspond to the closing price of the Shanghai Composite Index, different moments were set as cut-off points when calculating investors' cognitive differences.

Table 2. Descriptive Statistics

	APP			NAP			DIS		
	B_t	B_t^*	B_t^{ICD}	B_t	B_t^*	B_t^{ICD}	B_t	B_t^*	B_t^{ICD}
Sample number	700	700	700	3000	3000	3000	1400	1400	1400
Average	0.0046	-0.265	-3.239	0.009	-0.278	-2.964	0.002	-0.268	-2.484
Median	0.003	-0.265	-3.239	0.004	-0.271	-2.976	0.004	-0.271	-2.484
Maximum	0.165	-0.179	-2.214	0.095	-0.101	-0.953	0.646	-0.285	-0.803
Minimum	0.277	-0.339	-4.154	0.091	-0.386	-4.218	-0.073	-0.426	-3.857
Standard deviation	0.077	0.036	1.458	0.031	0.041	0.463	0.014	0.053	0.492
Skewness	-0.454	-0.006	-0.112	0.125	0.249	0.211	-0.295	0.244	0.152
Kurtosis	4.638	2.392	2.429	3.067	3.334	3.922	5.191	2.988	3.007
JB statistics	9.789 (0.007)	1.084 (0.571)	1.097 (0.577)	0.833 (0.659)	4.454 (0.107)	12.787 (0.017)	302.05 (0.000)	14.031 (0.009)	5.442 (0.065)

Table 2 shows descriptive statistics of relevant data. Statistical natures of stock yield B_t, trading volume B_t^* and investors' cognitive differences B_t^{ICD} were analyzed. It was found that the number of neutral posts in investors' cognitive differences was the highest. It indicates that domestic investors' blind obedience to the market has declined during market development. At the same time, the cognitive differences of investors were still large. Hence, there is still some information asymmetry in the trading process of investors, in which APP reaches 0.454 and DIS reaches 0.295.

Table 3. Stability Test

	B_t	B_t^*	B_t^{ICD}	B_t	B_t^*	B_t^{ICD}	B_t	B_t^*	B_t^{ICD}
ADF test	-8.532^{**} (0.000)	-4.134^{**} (0.016)	-4.304^{**} (0.009)	-16.908 (0.000)	-6.012^{**} (0.000)	-5.927^{**} (0.000)	-6.835^{**} (0.000)	-5.328^{**} (0.000)	-5.318^{*} (0.000)

Note: *, * * and * * indicate that they are credible at 1%, 5% and 10% respectively, the same below.

The ADF unit root test results are shown in Table 3. The stock yield B_t, trading volume B_t^* and investors' cognitive differences B_t^{ICD} are all stationary sequences.

4.2 Research Methods

Firstly, the Granger causality test was selected to explore the relationship between investors' cognitive differences and stock returns. Then, the influences of investors' cognitive differences on stock returns were mainly discussed.

$$R_t = \alpha_1 + \sum_{i=1}^{m} \beta_{1,i} R_{t-i} + \sum_{j=1}^{n} \gamma_{1,j} S_{y-j} + \varepsilon_{1,t} \tag{8}$$

$$S_t = \alpha_2 + \sum_{i=1}^{m} \beta_{2,i} R_{t-i} + \sum_{j=1}^{n} \gamma_{2,j} S_{t-j} + \varepsilon_{2,t} \tag{9}$$

Where: S_t is the value indicating the cognitive differences of investors at time t (expressed by B_t, B_t^*, B_t^{COD}, and abnormal cognitive indicator $Abn(S_t)$). R_t is the logarithmic yield of the closing price at time t, that is $R_t = \ln(P_t) - \ln(P_{t-1})$. P_t is the closing price of today. P_{t-1} is the closing price of the previous day. Two hypothesis tests were carried out by using the estimated results of the above equations. The original hypothesis in the first pair formula (8) test is $H_{0,1}: \gamma_{1,j} = 0, j = 1, 2, \cdots, n$. The original hypothesis of the second pair formula (9) test is $H_{0,2}: \gamma_{2,j} = 0, j = 1, 2, \cdots, n$. Hypothesis tests were judged by Wald chi-square statistics.

To further test the influence of investors' cognitive differences on stock returns, the Granger causality test of formula (8) was extended to the instantaneous Granger causality test:

$$R_t = \alpha_3 + \sum_{i=1}^{m} \beta_{3,i} R_{t-i} + \gamma_{3,0} S_t + \sum_{j=1}^{n} \gamma_{3,j} S_{t-j} + \varepsilon_{3,t} \quad (10)$$

The original hypothesis of formula (10): $H_{0,3}: \gamma_{3,j} = 0, j = 1, 2, \cdots, n$.

5. Empirical Test and Analysis

5.1 Empirical Test of Influences of Investors' Cognitive Differences on Stock Returns

Table 4 shows that investor cognitive difference indicators are highly like stock returns but not their Granger cause, whereas returns are the Granger cause of cognitive differences. Daily returns have a positive impact, while monthly and weekly returns have negative effects. Cognitive difference indicators are positively correlated with return persistence.

Table 4. Granger Causality Test Results of Investors' Cognitive Difference Indicators and Stock Yield

	Monthly frequency				Weekly frequency				Daily frequency		
	B_t	B_t^*	B_t^{ICD}		B_t	B_t^*	B_t^{ICD}		B_t	B_t^*	B_t^{ICD}
α_1	-0.0375 (-0.52)	-0.359 (-0.52)	-0.014 (-0.22)	α_1	0.007 (0.54)	0.007 (0.52)	0.015 (1.01)	α_1	-0.002 (-1.06)	-0.002 (-1.05)	-0.001 (-0.68)
R_{t-1}	-0.06 (-0.05)	-0.05 (-0.05)	-0.18 (-0.14)	R_{t-5}	-0.122 (-2.19)**	-0.122 (-2.19)**	-0.119 (-2.15)**	R_{t-1}	0.026 (0.09)	0.026 (0.09)	0.24 (0.84)
S_{t-1}	-0.152 (-0.57)	0.071 (-0.25)	-0.005 (-0.25)	S_{t-5}	0.148 (2.32)**	0.147 (2.32)**	0.146 (2.04)**	S_{t-1}	-0.026 (-1.12)	-0.004 (-1.12)	-0.006 (-0.76)
				S_{t-1}	0.026 (0.51)	0.011 (0.49)	0.004 (1.08)				
R^2	-0.024	-0.024	-0.028	R^2	0.026	0.026	0.031	R^2	-0.074	-0.064	-0.057
DW	2.032	2.033	2.034	DW	1.938	1.937	1.969	DW	1.996	1.996	1.995
X^2	0.325	0.332	0.061	X^2	0.263	0.244	1.156	X^2	1.262	1.261	0.582
P	0.567	0.564	0.804	P	0.607	0.621	0.282	P	0.261	0.263	0.445

By comparing the above investors' cognitive indicators, it can be concluded that the explanatory effect of the investors' cognitive difference index B_t^{ICD} on stock yield is better than that of the other two indicators B_t and B_t^*. This shows that the search frequency of text information has a positive correlation with the yield. Hence, the revised investors' cognitive difference indicator is more accurate.

Table (5) was summarized from Formula (10). As found, the self-lag variable of the yield was not significant, and thus the equation revealed by the empirical results is listed as follows:

$$\hat{R}_t = c + c_1 S_t + c_2 S_{t-1} \quad (11)$$

The empirical results reveal that $c_1 > 0$ and $c_2 < 0$. Both were significant at a 1% significance level. As the absolute values of investors' cognitive indicators S_t and S_{t-1} and test results were similar, it can be considered that the change of stock returns was caused by investors' cognitive differences. In other words, $s_t - s_{t-1}$. Hence, the correlation function between investors' cognitive differences and stock returns is:

$$\hat{R}_t = c + c_1 \Delta S_t \quad (12)$$

The B_t^{ICD} influence of investors' cognitive changes ΔS_t constructed based on daily investors' cognitive indicators on daily yield:

$$\hat{R} = 0.0002 + 0.0083 \Delta B_t^{ICD} \quad (13)$$

Table 5 shows that ΔB_t^{ICD} was significant at a 1% significance level, which verifies hypothesis 1 of this paper: investors' cognitive differences have influences on stock returns.

Table 5. Granger Causality Test Results of Investors' Cognitive Difference Indicators and Stock Yield (1)

	Monthly frequency				Weekly frequency				Daily frequency		
	B_t	B_t^*	B_t^{ICD}		B_t	B_t^*	B_t^{ICD}		B_t	B_t^*	B_t^{ICD}
α_1	-0.037 (-0.52)	-0.359 (-0.52)	-0.014 (-0.22)	α_1	0.007 (0.54)	0.007 (0.52)	0.015 (1.01)	α_1	-0.002 (-1.06)	-0.002 (-1.05)	-0.001 (-0.68)
R_{t-1}	-0.06 (-0.05)	-0.05 (-0.05)	-0.18 (-0.14)	R_{t-5}	-0.122 (-2.19)**	-0.122 (-2.19)**	-0.119 (-2.15)**	R_{t-1}	0.026 (0.09)	0.026 (0.09)	0.024 (0.84)
S_{t-1}	-0.152 (-0.57)	0.071 (-0.25)	-0.005 (-0.25)	S_{t-5}	0.148 (2.32)***	0.147 (2.32)***	0.146 (2.04)***	S_{t-1}	-0.026 (-1.12)	-0.004 (-1.12)	-0.006 (-0.76)
				S_{t-1}	0.026 (0.51)	0.011 (0.49)	0.004 (1.08)				
R^2	-0.024	-0.024	-0.028	R^2	0.026	0.026	0.031	R^2	-0.074	-0.064	-0.057
DW	2.032	2.033	2.034	DW	1.938	1.937	1.969	DW	1.996	1.996	1.995
X^2	0.325	0.332	0.061	X^2	0.263	0.244	1.156	X^2	1.262	1.261	0.582
P	0.567	0.564	0.804	P	0.607	0.621	0.282	P	0.261	0.263	0.445

Because investors behave differently in market trading hours and non-trading hours, this paper divided the analysis time of text information according to market trading hours. The investor's cognitive indicators calculated in the two periods are recorded as $B_{t,pre-open}^{ICD}$ and $B_{t,trading}^{ICD}$ respectively. Correspondingly, the daily stock yield difference was divided into the opening yield $R_t^{open} = \ln(p_t^{open}) - \ln(P_{t-1}^{close})$ and trading day yield $R_t^{trading} = \ln(p_t^{close}) - \ln(P_t^{open})$. The instantaneous Granger causality test shown in Formula (10) was estimated by using the above indicators. Then, the influences of investors' cognitive indicators on the opening and closing prices of stocks were discussed. The specific results are shown in Table 6.

Table 6. Granger Causality Test Results of Investors' Cognitive Difference Indicators and Stock Yield (2)

	Monthly frequency				Weekly frequency				Daily frequency		
	B_t	B_t^*	B_t^{ICD}		B_t	B_t^*	B_t^{ICD}		B_t	B_t^*	B_t^{ICD}
α_2	-0.099 (-3.67)*	-0.185 (-3.67)*	-1.185 (-3.95)*	α_2	-0.072 (-5.26)	-0.147 (-5.17)	-0.686 (-4.42)*	α_2	-0.032 (-4.64)	-0.067 (-4.66)**	-0.368 (-5.01)**
R_{t-1}	-0.103 (-2.28)	-0.218 (-2.31)	-1.255 (-2.91)	R_{t-1}	0.121 (1.97)**	0.259 (1.96)*	-0.679 (-0.96)	R_{t-1}	1.057 (15.18)	2.293 (15.22)**	9.314 (14.89)**
R_{t-2}	-0.149 (3.59)**	-0.323 (-3.62)	-1.956 (-3.77)	R_{t-2}	-0.159 (-2.66)	-0.347 (-2.69)	-2.157 (-3.12)	R_{t-2}	0.252 (10.01)*	0.254 (11.07)**	0.308 (5.36)***
S_{t-1}	0.657 (7.15)**	0.658 (7.18)***	0.633 (6.92)***	S_{t-1}	0.494 (8.04)**	0.494 (8.03)***	0.503 (8.08)***	S_{t-1}	0.129 (5.34)**	0.103 (5.35)***	0.133 (5.36)***
				S_{t-2}	0.293 (4.04)**	0.241 (4.06)***	0.268 (4.38)***	S_{t-5}	0.232 (9.43)	0.229 (9.34)***	0.204 (8.33)***
								S_{t-10}	0.166 (7.18)**	0.167 (7.24)***	0.146 (6.49)***
								S_{t-15}	0.098 (7.18)**	0.097 (4.44)***	0.087 (3.98)***
R^2	0.474	0.477	0.483	R^2	0.453	0.455	0.441	R^2	0.539	0.549	0.552
DW	1.798	1.793	1.829	DW	2.061	2.029	2.073	DW	1.887	1.888	1.904
X^2	17.589	17.813	22.32	X^2	10.781	11.213	11.306	X^2	230.03	231.67	221.08
P	0.0002	0.0001	0.0000	P	0.0037	0.0035	0.0044	P	0.0000	0.0000	0.0000
Effect	(-)***	(-)***	(-)***	Effect	(-)***	(-)***	(-)***	Effect	(-)***	(-)***	(-)***

5.2 Empirical Test of Investors' Cognitive Differences on the Ability to Predict Stock Returns

Based on the above analysis, this paper holds that investors' cognitive differences have a positive correlation with the yield in the same period. Investors' cognitive differences in trading hours have a positive influence on trading returns. Considering that the formation of investors' cognitive differences depends on the previous stock returns, investors' cognitive differences before the daily opening can predict the opening price of stocks.

Table 7. Instantaneous Granger Causality Test Results of Investors' Cognitive Difference Indicators on Stock Yield

Monthly frequency				Weekly frequency				Daily frequency			
	B_t	B_t^*	B_t^{ICD}		B_t	B_t^*	B_t^{ICD}		B_t	B_t^*	B_t^{ICD}
α_3	0.067 (0.69)	0.063 (0.69)	0.081 (1.14)	α_3	0.042 (2.94)***	0.047 (3.61)***	0.053 (3.85)***	α_3	0.009 (3.85)***	0.008 (3.83)	0.001 (4.16)
R_{t-1}	0.098 (0.56)	0.099 (0.57)	0.112 (0.89)	R_{t-5}	-0.069 (-1.36)	-0.068 (-1.34)	-0.064 (-1.27)	R_{t-1}	-0.058 (-1.75)*	-0.052 (-1.77)*	-0.059 (-2.01)*
				R_{t-6}	0.155 (2.07)***	0.154 (2.69)***	0.166 (2.92)***				
S_t	1.144 (3.93)***	0.533 (3.94)***	0.084 (3.44)***	S_t	0.362 (7.52)***	0.169 (7.61)***	0.034 (7.61)***	S_t	0.086 (9.03)***	0.045 (9.02)***	0.016 (9.48)***
S_t	-0.948 (-2.56)*	-0.422 (-2.66)*	-0.068 (-2.37)*	S_t	-0.206 (-3.22)*	-0.097 (-3.35)*	-0.016 (-2.79)*	S_t	-0.025 (-6.01)*	-0.024 (-6.12)*	-0.065 (-6.62)*
R^2	0.147	0.148	0.116	R^2	0.146	0.166	0.188	R^2	0.061	0.061	0.069
DW	2.001	2.001	2.015	DW	1.991	1.989	1.796	DW	2.005	2.006	2.006
X^2	17.84	18.43	11.87	X^2	58.38	59.62	61.71	X^2	84.85	84.82	92.93
P	0.0001	0.0001	0.0026	P	0.0000	0.0000	0.0000	P	0.0000	0.0000	0.0000
Effect	(+)***	(+)***	(+)***	Effect	(+)***	(+)***	(+)***	Effect	(+)***	(+)***	(+)***

Kim (2014)[11] and Antweiler (2004)[9] found investors couldn't predict stock returns. This study analyzes trading vs. non-trading hour posts, revealing non-trading hour texts predict opening prices, while trading hour texts predict closing prices, as non-trading hours allow deeper discussion and trading hour posts reflect investment logic.

Table 8. Granger Causality Test and Instantaneous Granger Causality Test Results of Investors' Cognitive Difference Indicators on Stock Opening Price and Intraday Rise

Opening Rise					Intraday Rise				
C	0.024 (6.53)***	0.009 (1.05)	-0.001 (-1.28)	-0.001 (-1.09)	C	-0.002 (-0.16)	-8.23 (0.03)	0.029 (11.01)**	0.029 (7.19)***
R_{t-1}^{tra}	-0.059 (-1.88)*	-0.032 (-1.01)	-0.067 (-2.11)**	-0.053 (-1.63)	R_{t-1}^{tra}	-0.025 (-0.77)	-0.036 (-1.19)	0.015 (0.32)	0.022 (0.65)
$B_{t,pre-ope}^{ICD}$	0.002 (8.059)**		0.028 (9.24)***	0.029 (9.29)***	$B_{t,tra}^{ICD}$			0.014 (12.78)**	0.016 (14.05)***
$B_{t-1,tra}^{ICD}$		0.006 (1.85)	-0.001 (-4.55)**	-0.004 (-3.42)**	$B_{t-1,pre-ope}^{ICD}$	-0.007 (-1.01)		-0.006 (-8.78)**	-0.004 (-5.94)***
$B_{t-1,pre-ope}^{ICD}$				-0.007 (-2.36)**	$B_{t-1,tra}^{ICD}$		-0.004 (-0.42)		-0.006 (-5.91)***
R^2	0.058	0.021	0.063	0.066	R^2	0.001	0.003	14.36	0.163
DW	1.906	1.996	1.981	2.012	DW	2.002	1.998	1.932	2.025
X^2	64.949	4.412	86.228	88.133	X^2	1.013	0.175	171.954	207.981
P	0.0000	0.0647	0.0000	0.0000	P	0.3141	0.6755	0.0000	0.0000
Effect	(+)***	(+)*	(+)***	(+)***	Effect	(+)***	(+)***	(+)***	(+)***

Table 8 shows investors' cognitive differences are influenced by prior market prices: negative over long periods, positive over short ones. Formula (9) indicates a two-week inertia, marking the long/short-term divide. Investors are trend followers, optimistic short-term but reverting long-term, unable to predict returns confirming hypothesis 2.

5.3 Robustness Test

The Granger causality test is very sensitive to the choice of the length of the lag period. In addition, influences caused by sample size length on stationarity of the tested variables should be avoided. Hence, a stability test is needed. By selecting different lag periods for testing, whether the random disturbance term in the model affects the stability of the model can be tested. It can also test whether there is false regression in the model variables. In this paper, three variable periods with different lengths were selected to test the stationarity of the above Granger test, as shown in Table 8.

Table 9. Stability Test of Influences of Investors' Cognitive Difference Indicators on Stock Yield

	B_t			B_t^*				B_t^{ICD}		
α_0	1.096 (1.09)*	-0.015 (-0.03)	0.432 (0.64)	1.649 (2.37)**	-0.107 (-0.16)	0.677 (0.92)	0.714 (1.72)*	0.301 (0.68)	0.392 (0.08)	1.178 (1.91)*
V_{t-1}	0.582 (16.36)**	0.578 (16.34)**	0.583 (16.42)**	0.579 (16.43)**	0.577 (16.32)**	0.579 (16.42)**	0.578 (16.17)**	0.579 (16.39)	0.581 (16.21)	0.582 (16.55)
V_{t-2}	0.193 (5.51)***	0.192 (5.47)***	0.191 (5.44)***	0.194 (5.55)***	0.194 (5.51)***	0.195 (5.58)***	0.194 (5.51)***	0.191 (5.47)**	0.195 (5.42)**	0.189 (5.42)**
V_{t-3}	0.115 (3.61)***	0.116 (3.64)***	0.114 (3.06)***	0.113 (3.57)***	0.117 (3.68)***	0.114 (3.59)***	0.115 (3.63)***	0.115 (3.62)**	0.116 (3.61)**	0.112 (3.49)**
R_{t-1}	33.155 (11.15)*	31.535 (10.23)*	31.408 (10.02)*	33.796 (11.21)**	32.103 (10.76)***	33.536 (11.13)***	32.885 (11.04)***	32.404 (9.68)***	31.025 (9.12)***	31.302 (9.68)***
ICD_t		1.084 (1.28)	1.865 (1.86)*		1.255 (1.33)	2.484 (2.31)**		0.601 (1.08)	0.769 (1.17)	1.503 (2.11)**
ICD_{t-1}	-0.458 (-0.58)		-1.143 (-1.05)	-1.009 (-1.16)		-2.364 (-2.04)***	0.067 (0.13)		-0.289 (-0.46)	-2.298 (-2.03)
R^2	0.749	0.747	0.749	0.746	0.747	0.755	0.743	0.745	0.749	0.752
DW	2.013	2.005	2.017	2.012	2.005	2.019	2.009	2.008	2.012	2.017

Table 10. Empirical Results of Influences of Investors' Cognitive Differences on Daily Stock Volatility

	B_t			B_t^*			B_t^{ICD}		
R_{t-1}	-0.006 (-4.38)***	-0.008 (-4.09)***	-0.006 (-4.09)***	-0.007 (-4.51)***	-0.007 (-4.41)***	-0.007 (-4.05)***	-0.006 (-4.36)***	-0.005 (-3.33)***	-0.006 (-3.46)***
R^2	0.965	0.968	0.965	0.969	0.959	0.969	0.965	0.955	0.966
DW	2.003	2.033	2.032	2.033	2.035	2.036	2.037	2.034	2.035

The stability test supports that investors' cognitive differences serve as a Granger causality of stock return changes, and verifies the previous hypothesis.

6. Conclusion

In this paper, the Granger causality test, instantaneous causality test and intertemporal regression analysis were used to analyze the influences of investors on stock returns at different times and frequencies. As found, investors' cognitive differences have a significant influence on stock returns in the current period. However, the investors' cognitive differences have no predictive ability on stock returns. In different trading periods, it is found that investors' cognitive differences before the opening can predict the opening price. The investor's cognition in the trading period after the opening has a positive influence on the closing price of that very day.

References

[1] Zhang Xu and Chen, Management power and R&D catering investment of listed companies, Securities Market Herald, 2018, (7), 38-47.

[2] Pei and Zhang, Empirical test of the cognitive bias of China stock investors, Management World, 2004, (12), 12-22.

[3] Liu, Zhang, Gu and Yao, How does investor behavior affect the liquidity of the stock market? -Analysis based on investor sentiment, information cognition and short selling constraints, Journal of Management Sciences in China, 2016, (10), 87-100.

[4] Todd Houge, Tim Loughran, Cash Flow Is King? Cognitive Errors by Investors? Journal of Behavioral Finance, 2000, 1(3-4), 161-175.

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

[6] Ylva Backstrom, Jo Silvester and Rachel A. J. Pownall, Millionaire investors: financial advisors, attribution theory and gender differences, The European Journal of Finance, 2018, 24, 1333-1349.

[7] Jia, Zhu and Chen, Company name, investor cognition and company value - behavioral finance research based on company name evaluation indicator system, Journal of Financial Research, 2016, (5), 173-190.

[8] Cui and Hong, Will investors' cognitive bias on economic fundamentals affect securities prices? -Comparative Analysis

of Chinese and American Securities Markets, Economic Research Journal, 2017, (7), 94-109.
[9] Antweiler, Werner, and Murray Z, Frank. “Is All That Talk Just Noise? The Information Content of Internet Stock Message Boards”, Journal of Finance, 2004, 59(3): 1259-1295.
[10] Das Sanjiv and Mike Y. Chen, “Yahoo! for Amazon: Sentiment Extraction from Small Talk on the Web”, Management Science, 2007, 53(9), 1375-1388.
[11] Kim S H, Kim D. Investor sentiment from internet message postings and the predictability of stock returns. Journal of Economic Behavior & Organization, 2014, 107, 708-729.