Does Individual Investor Sentiment Have an Impact on Stock Returns?—Empirical Research Based on Stock Comments for CSI 300 Index

Zhekun Xiao

Quantitative Finance, HSBC Business School, Peking University, Shenzhen, China
Email: 2201212426@stu.pku.edu.cn (Z.K.X.)
Manuscript received April 8, 2024; revised April 30, 2024; accepted May 10, 2024; published August 9, 2024.

***Abstract*—This study investigates the impact of individual investor sentiment on stock returns in China's growing stock market. By utilizing sentiment metric data from Cathay Pacific stock forum views, this paper validates the relationship between investor sentiment and stock returns. In terms of methodology, this study discards traditional control variables and adopts an innovative approach combining historical investor sentiment to fit stock returns through a higher-order autoregressive model. In the empirical analysis, panel data of 144 stocks among the constituents of CSI 300 index from January 16, 2019 to December 1, 2021 are used. The homoskedasticity and heteroskedasticity problems, as well as the current relationship and endogeneity problems in the panel data analysis are solved by employing the Feasible Generalized Least Square and Least Squares Dummy Variable techniques. The results show that the long index can effectively serve as a proxy variable for individual investor sentiment in the Chinese stock market at the individual stock level. In addition, investor sentiment at the individual stock level has a significant impact on stock returns: current investor sentiment greatly enhances contemporaneous stock returns, while the lagged term of sentiment (delayed by one to three days) has a significant negative impact on stock returns, which is an important finding for better protecting the rights and interests of individual investors and maintaining the stability of the financial market.**

***Keywords*—Least Squares Dummy Variable (LSDV), Feasible Generalized Least Square (FGLS), individual investor sentiment, stock return**

I. Introduction

Studying the relationship between stock returns and individual investor sentiment in the Chinese stock market has important theoretical and practical implications. Stock return, as an important indicator of market performance, directly affects investors' asset values and investment decisions. Individual investor sentiment, as an irrational factor affecting stock market volatility, often reveals its important role in abnormal market fluctuations. Studies have shown that individual investor sentiment may cause the stock market to overreact or underreact, thereby affecting stock prices and yields.

China's stock market has now developed into a diversified and comprehensive stock market covering a wide range of sectors. At the same time, China's economy continues to grow steadily, with its GDP ranking second in the world and residents' disposable income rising. With the growing sophistication of China's stock market and the increasing number of investors, the demand for financial management among Chinese individual investors has been increasing. According to China Securities Depository & Clearing Corporation (CSDC), as of April 17, 2021, the number of registered A-share accounts reached 198 million, with one-person-one-household as the mainstream, and the total market value of shares held by individual investors is 1.4 times that of institutional investors. This shows that individual investors are still the main players in the Chinese stock market, so the study of individual investors in the Chinese stock market is of great significance.

This paper uses the public opinion data of the stock bar of Guotai' an Database to measure individual investor sentiment. After verifying that this index can measure investor sentiment, it carries out empirical analysis on individual investor sentiment and stock return rate, and explores the impact of individual investor sentiment on the stock market under high-frequency data from the level of individual stocks, with a view to better protecting individual investors and maintaining financial market stability.

II. Literature Review

DeLong *et al.* (1990) were the first to conceptualize investor sentiment as investors' biased expectations about a stock's value. Han *et al.* (2018) argue that investor sentiment represents the collective bias that emerges in the investor community's predictions about market asset price movements. Building on this, this article defines investor sentiment as the discrepancy between investors' valuation judgments and the true value of assets. To measure investor sentiment, Wurgler (2006) utilized principal component analysis to construct an investor sentiment index, while Fernandes, Goncalves, and other scholars (2013) considered the Economic Sentiment Index and Consumer Confidence Index, collected through surveys in the European Union, as explicit indices reflecting investor sentiment. Bu & Forrest (2024) conclude that sentiment shock is a more accurate indicator of the relationship between investor sentiment and contemporaneous stock returns. There are also scholars who directly measure investor sentiment through text data mining and machine learning. For instance, Bartov *et al.* (2018) employed machine learning and deep learning algorithms for text sentiment classification. Recognizing that textual information is a direct carrier of shareholder sentiment, which can immediately reflect shareholders' views and expectations about individual stocks, this paper seeks proxies for investor sentiment at the individual stock level in the field of text mining. With the continuous development of behavioral finance, an increasing number of researchers are examining the impact of investor sentiment on the stock market. Behrendt, Schmidt (2018) discovered a feedback effect between investor sentiment and intraday volatility based on a

doi:10.18178/joebm.2024.12.3.809

per-minute investor sentiment index at the individual stock level. Gao *et al.* (2019) constructed a weekly country-specific investor sentiment index using Google search frequencies for sentiment terms and found that investor sentiment has a negative predictive effect on weekly market returns. Wu and Yang (2022) found that investor sentiment exerts positive effects on the deviation of stock prices from their fundamental value, with the direction and magnitude of the price deviation depending on the relative strength of the sentiment signal.

In this paper, this research uses a novel approach to explore the relationship between stock returns and investor sentiment. Different from the traditional method of adding control variables, this research uses only the past period's investor sentiment to regress the stock returns, which can effectively solve the pseudo-regression problem of control variables and avoid the interference of other factors to fully portray the relationship between stock returns and investor sentiment.

III. Model and Methods

A. Investor Sentiment Indicator

In this study, I selected the public opinion data from Cathay Pacific China stock bars, and specifically used the average daily optimism index of individual stock investors collected and organized by this database as a proxy measure of individual investor sentiment. The Cathay Pacific China Stock Bar Public Opinion Research Database employs a deep learning model to analyze the sentiment of stock review texts from online stock bars, to sort out the sentiment attitudes of listed companies' stock reviews, and to screen, quantify and count them, providing quantitative opinion data for the listed companies categorized by time and posters' characteristics.

This research uses the following Bullish index:

$$E_{\mathrm{t}} = \frac{M_t^{buy} - M_t^{sell}}{M_t^{buy} + M_t^{sell}} \quad (1)$$

where M_t^{buy} represents the total number of positive posts in a day and M_t^{sell} represents the total number of negative posts in a day.

B. Research Design

The central objective of this empirical investigation is to verify the relationship between stock returns and past investor sentiment. Specifically, it is to determine whether current investor sentiment acts as a driver of stock returns, while conversely, a long lag in investor sentiment tends to have a negative impact on stock returns.

C. Innovation Points

In this paper, I wish to explore the relationship between stock returns and investor sentiment, but unlike traditional models with added control variables, this paper wishes to test the relationship between stock returns and investor sentiment from a novel perspective. This paper discards the traditional regression model with control variables and replaces it with time-series data on investor sentiment, and utilizes a higher-order autoregressive model of investor sentiment to explore the relationship between stock returns and investor sentiment.

D. Dependent Variables

This study takes an innovative approach by dispensing with the traditional reliance on control variables. Instead, it introduces a series of Sentiment Indicators, ranging from SI-1 through SI-10, as the core variables. For the purpose of this study, 'RET' denotes the daily return on stocks, adjusted for cash dividends reinvested into the system. Meanwhile, 'SI' encapsulates the investor sentiment as gleaned from an analysis of the textual comments on Guba websites. The term 'SI-t' refers to the Sentiment Indicator (SI) offset by t trading days to reflect the lagged impact of investor sentiment.

RET: The daily stock return (%) considered cash dividends reinvested.

SI: Investor sentiment calculated from Guba comment text analysis.

SI-t: Investor sentiment (SI) lagged t trading days.

E. Higher-order Autoregressive Models

In this paper, a fixed effects model is selected based on the results of the Hausmann test. But different from the traditional regression method using control variables, I propose a new modeling idea: regressing stock returns on investor sentiment only for the current period and for the past several periods, so as to better explore the impact of investor sentiment on stock returns in different periods, and finally I obtain a higher-order autoregressive model by setting the order to 10, which is due to the fact that the two-Iek trading day is usually 10 days.

The model is crafted with the following specifications designed to ensure dimension consistency across variables:

$$RET_{\mathrm{it}} = \beta_0 SI_{\mathrm{i,t}} + \beta_1 SI_{\mathrm{i,t\text{-}1}} + \ldots + \beta_{10} SI_{\mathrm{i,t\text{-}10}} + u_i + \varepsilon_{i,t} \quad (2)$$

where i = l, 2, 3, ..., n; t = l, 2, 3, ..., T, denoting stocks and time, respectively, u_i is the intercept term representing individual heterogeneity, $\varepsilon_{i,t}$ is a disturbance term that varies across individuals and time, and $SI_{\mathrm{i,t\text{-}j}}$ is a j-th order lag term for investor sentiment.

Given that changes in macroeconomic policies are relatively modest over short spans, both monthly and annual dummy variables are integrated into the model to administer control over time-related fixed effects. The Least Squares Dummy Variable (LSDV) method is engaged to estimate the bidirectional fixed effects model.

The analysis of regression results requires a comprehensive comparison between LSDV Estimator and Feasible Generalized Least Square (FGLS) Estimator.

F. Least Squares Dummy Variable

Least Squares Dummy Variable (LSDV) is a widely used technique in multiple linear regression analysis, mainly for dealing with unobserved heterogeneity in Panel Data. The method ensures the consistency of model estimation by introducing Dummy Variables to capture unobservable individual or time effects.

The specific process is as follows:

1) Determine the existence of individual or time effects and decide whether to introduce individual or time dummy variables;

2) Add N-1 dummy variables (for N individuals) to the regression model, each representing an individual in addition to a reference group;
3) Estimate the model parameters, including the coefficients of the dummy variables, by least squares;
4) Perform statistical tests, such as F-tests, based on the regression results to test the overall significance of the dummy variables.

The main advantage of the LSDV method is its intuition and simplicity. By introducing dummy variables, the effects of categorical variables on dependent variables can be easily identified and explained. In addition, LSDV is particularly useful when dealing with Fixed Effects Models (FEM), where it is able to control for the effects of individual characteristics that do not vary over time on the model.

G. Feasible Generalized Least Square

The Feasible Generalized Least Squares (FGLS) method is an estimation technique in econometrics that deals with the problems of heteroskedasticity and autocorrelation. It improves the accuracy of the estimator by exploiting information about the structure of the heteroskedasticity and autocorrelation of the error terms. It first builds a model with reasonable assumptions about the structure of the variance and covariance of the error terms and then uses this model to transform the raw data so that the transformed model satisfies the classical assumptions of OLS estimation.

The specific process is as follows:

1) Estimate original model using OLS and obtain residuals;
2) Estimate the heteroskedasticity or autocorrelation structure of the residuals, usually by testing or modelling the residuals;
3) Construct a weight matrix or transformation matrix based on the results in step 2;
4) Transforming the data using the constructed matrix such that the error terms of the transformed data have homoskedasticity and no autocorrelation;
5) Apply OLS estimation to the transformed data.

The main advantage of the FGLS method is that it can improve the efficiency of the estimator by making reasonable use of the residual information. Theoretically, the FGLS estimator has a smaller standard error and is more reliable than the OLS estimator.

IV. Empirical Analysis and Result

In the empirical analysis, panel data of 144 stocks among the constituents of CSI 300 index from January 16, 2019 to December 1, 2021 are used. The homoskedasticity and heteroskedasticity problems, as well as the current relationship and endogeneity problems in the panel data analysis are solved by employing the FGLS and LSDV techniques.

A. Data Source

Given the incomplete data collection of the CSMAR bullish index prior to 2019 and the higher incidence of missing data for certain stocks under consideration, this study sets its starting point in 2019.

The CSI 300 index, a broad-based benchmark reflecting China’s stock market performance, selects its components by thoroughly weighing factors like liquidity and market capitalization. This study prioritizes liquidity and aims for a wide representation, hence the choice of the CSI 300 index constituents as the subjects for analysis.

The constituent stocks of CSI 300 Index change every six months, so this paper selects the stocks that are always the constituent stocks of the index during the research period, and excludes the stocks that have been suspended during the research period, and finally obtains 144 stocks as the research object.

Data Source: CSMAR database
Time range: 2019.01.16–2021.12.01
Stock range: 144 stocks that belongs to CSI 300

B. Results

In this body of research, I have constructed a fixed effects model with the primary dependent variable identified as the stock return (RET), and the principal explanatory variable defined as the investor sentiment indicators ranging from the current period’s SI to the sentiment delayed by ten periods. my approach to solving the model entailed the application of two distinct methodologies: Least Squares Dummy Variable (LSDV) and Feasible Generalized Least Squares (FGLS).

1) Descriptive statistics

Table 1. Descriptive statistics result

Variable	Obs	Mean	Std.Dev.	Min	Max
SI	41576	−0.3970445	0.7641641	−1	1
SI_1	41503	−0.3972607	0.7641889	−1	1
SI_2	41424	−0.3976478	0.764024	−1	1
SI_3	41339	−0.3977764	0.7641379	−1	1
SI_4	41265	−0.3984172	0.763945	−1	1
SI_5	41192	−0.3986846	0.7638008	−1	1
SI_6	41134	−0.3988777	0.7637119	−1	1
SI_7	41068	−0.3988371	0.7637419	−1	1
SI_8	40996	−0.3991473	0.7636997	−1	1
SI_9	40920	−0.3995335	0.7635883	−1	1
SI_10	40839	−0.3997418	0.7635911	−1	1
RET	82859	0.0008797	0.0266439	−0.199956	0.178635

The mean value of the sentiment index is around −0.397 with a standard deviation of about 0.764, which indicates that investor sentiment tends to be slightly negative and volatile. In addition, the mean of the daily return on equity (RET) is weakly positive at 0.0008797, with a standard deviation of 0.0266349, a minimum value of −19.965%, and a maximum value of 17.8635%, indicating a more significant range of volatility in returns. This range of volatility is indicative of the uncertainty faced by the market during the sample period, as well as the potential investment risks and rewards.

2) Relevance analysis

The correlation between the variables is shown in the table below:

Table 2. Correlation coefficient

	SI	SI_1	SI_2	SI_3	SI_4	SI_5	SI_6	SI_7	SI_8	SI_9	SI_10	RET
SI	1.000											
SI_1	0.375	1.000										
SI_2	0.284	0.375	1.000									
SI_3	0.236	0.285	0.374	1.000								
SI_4	0.208	0.236	0.284	0.374	1.000							
SI_5	0.192	0.208	0.236	0.283	0.373	1.000						
SI_6	0.177	0.191	0.208	0.235	0.282	0.372	1.000					
SI_7	0.166	0.177	0.191	0.208	0.235	0.282	0.372	1.000				
SI_8	0.152	0.166	0.176	0.191	0.207	0.234	0.281	0.372	1.000			
SI_9	0.144	0.153	0.166	0.176	0.191	0.207	0.234	0.281	0.372	1.000		
SI_10	0.139	0.144	0.153	0.165	0.176	0.190	0.207	0.234	0.281	0.372	1.000	
RET	0.209	0.023	0.013	0.010	0.010	0.010	0.010	0.010	0.010	0.012	0.009	1.000

SI-1, SI-2, and SI-3, show higher correlation coefficients with RET, indicating that short-term investor sentiment may have a more immediate impact on stock returns. As the lag increases, the correlation tends to diminish, which could suggest that the impact of investor sentiment on stock returns decreases over time. This decay in correlation may reflect the market's absorption and subsequent integration of sentiment into stock prices. Interestingly, certain Sentiment Indicators, such as SI-7 through SI-10, still maintain moderate correlations with RET despite the extended lag period. This may indicate that some aspects of investor sentiment have a more enduring effect on stock returns.

3) LSDV regression results

The LSDV regression analysis, as shown in Table 3, statistically validates the significant impact of investor sentiment on stock returns up to three lagged periods, with a significance level less than 0.01. The regression coefficient for the current period investor sentiment (SI) is notably positive, suggesting that contemporaneous investor sentiment has an immediate and beneficial influence on stock returns. This reflects the market's prompt response to prevailing sentiment as captured through Guba comment text analysis.

Table 3. LSDV regression results

Variable	**Coef**	**Std.Err**	**t**	**P>|t|**	**[95% Conf. Interval]**	
SI	0.01609	0.00022	71.77	0.000	0.01565	0.01653
SI_1	−0.00300	0.00023	−12.95	0.000	−0.00346	−0.00254
SI_2	−0.00155	0.00023	−6.66	0.000	−0.00201	−0.00109
SI_3	−0.00095	0.00023	−4.09	0.000	−0.00141	−0.00049
SI_4	−0.00046	0.00023	−2.00	0.046	−0.00092	−8.7e-06
SI_5	−0.00047	0.00023	−2.01	0.045	−0.00092	−0.00001
SI_6	−0.00028	0.00023	−1.20	0.229	−0.00073	0.00018
SI_7	−0.00028	0.00023	−1.23	0.219	−0.00074	0.00017
SI_8	−0.00010	0.00023	−0.44	0.660	−0.00056	0.00035
SI_9	0.00011	0.00023	0.51	0.608	−0.00033	0.00057
SI_10	−0.00022	0.00022	−0.97	0.333	−0.00065	0.00022
_cons	0.00201	0.00009	22.46	0.000	0.00183	0.00218
Sigma_u	0.00136				Sigma_e	0.02586
R^2	0.0492				Corr(u_i,Xb)	−0.0686
F(11,99273)	473.9				Prob>F	0.0000
F(143,99273)	1.71				Prob>F	0.0000
rho	0.00276					

The coefficients for lagged sentiment indicators (SI_1, SI_2, and SI_3) are negative, denoting an inverse relationship between past sentiment and future stock returns. This indicates that while investor sentiment can initially drive stock prices up, this effect tends to reverse in subsequent periods. The negative coefficients for SI_1 and SI_2 are particularly significant, pointing to a potential overreaction in the market to sentiment changes, which is then corrected in the days following the sentiment expression.

The diminishing magnitude and significance of coefficients for further lagged periods (SI_4 through SI_10) suggest that the impact of sentiment on stock returns weakens over time and becomes statistically insignificant after the third lag. This attenuation of influence may be due to the market's assimilation of sentiment information or the emergence of other market-driving forces that overshadow the effect of investor sentiment with time.

The R-squared value of the model is relatively low, indicating that while investor sentiment does have a predictive influence on stock returns, it accounts for only a small proportion of the variation in returns. This underscores the multifaceted nature of stock price movements, influenced by a multitude of factors beyond investor sentiment alone. The F-statistic values are highly significant, supporting the overall model's validity. The rho statistic, while small, is also significant, suggesting that there is some degree of autocorrelation in the panel data that the LSDV model is accounting for.

The coefficients' standard errors are small, lending credibility to the precision of the estimates. The 95% confidence intervals further reinforce the significance of the results, especially for the immediate sentiment indicator (SI) and the first two lags (SI_1 and SI_2), which do not cross the zero threshold, indicating that the effects are not due to random chance. This reinforces the conclusion that investor sentiment has a measurable and immediate impact on stock returns, followed by a period of adjustment where the initial effects are partially or wholly negated.

4) Heteroscedasticity

Table 4. Heteroscedasticity result

Source	chi2	df	p
Heteroskedasticity	471.95	65	0.0000
Skewness	434.39	10	0.0000
Kurtosis	1300.57	1	0.0000
Total	2206.91	76	0.0000
Chi2(65)	471.95	Prob > chi2	0.0000

The heteroscedasticity test results encapsulated in Table 4 provide a stark indication of the variability in the regression model's residuals. The Prob > chi2 statistic is observed at an unequivocal 0.0000, prompting the rejection of the null hypothesis and confirming the presence of heteroscedasticity. This means that the error terms' variances are not constant across the dataset, implying that the model's predictive capability varies at different levels of the independent variables, specifically the range of Sentiment Indicators (SI) examined. The chi2 values for heteroscedasticity, skewness, and kurtosis are considerably high, indicating non-constant variance, non-symmetric distribution of residuals, and non-normality in their distribution with significant heavy or light tails. These factors suggest that the classic assumptions necessary for standard OLS regression–constant variance and normality of residuals–do not hold for our data. Such conditions can lead to underestimation of the standard errors and, consequently, to erroneous conclusions about the significance of the model's coefficients. Consequently, the FGLS method is being considered for a more robust solution to the presented heterogeneity.

5) FGLS regression results

Table 5. FGLS regression results

Variable	Coef	Std.Err	z	P>|z|	[95% Conf. Interval]	
SI	0.01594	0.00022	71.36	0.000	0.01550	0.01638
SI_1	−0.00312	0.00023	−13.46	0.000	−0.00357	−0.00266
SI_2	−0.00165	0.00023	−7.10	0.000	−0.00201	−0.00109
SI_3	−0.00104	0.00023	−4.49	0.000	−0.00150	−0.00059
SI_4	−0.00056	0.00023	−2.39	0.017	−0.00101	−0.0001
SI_5	−0.00056	0.00023	−2.40	0.017	−0.00101	−0.0001
SI_6	−0.00037	0.00023	−1.59	0.111	−0.00082	0.00009
SI_7	−0.00038	0.00023	−1.63	0.104	−0.00083	0.00008
SI_8	−0.00020	0.00023	−0.87	0.386	−0.00065	0.00025
SI_9	5.7e-06	0.00023	0.02	0.980	−0.00044	0.00046
SI_10	−0.0004	0.00022	−1.64	0.102	−0.00079	−0.00022
_cons	0.00190	0.00009	21.55	0.000	0.00173	0.00007
Wald chi2(11)	5157.08				Prob > chi2	0.0000
Log likelihood	222286.5					

The Feasible Generalized Least Squares (FGLS) regression results in Table 5 provide an insightful reassessment of the impact of investor sentiment on stock returns after accounting for heteroscedasticity detected in previous OLS estimates. Consistent with the LSDV results, the FGLS method reveals a statistically significant impact of current and lagged investor sentiment on stock returns at a significance level of 0.01. The coefficient for the contemporaneous investor sentiment (SI) is positive (0.01594) and highly significant, which confirms the direct and immediate positive relationship between investor sentiment and stock returns identified in the previous model. The Wald chi-squared statistic reinforces the significance of the model, and the model's log likelihood suggests a good fit. This strengthens the reliability of the FGLS results and their consistency with the LSDV analysis.

The regression results also confirm the negative influence of investor sentiment when lagged by one to three days (SI_1 to SI_3). SI_1, with a coefficient of −0.00312, exhibits the strongest negative relationship, indicating that the previous day's sentiment exerts a corrective influence on stock returns. The subsequent lags, SI_2 and SI_3, show a declining negative impact, which is aligned with the expectation that the effect of sentiment on stock returns diminishes over time. The coefficients for SI_4 and SI_5 are also negative and remain statistically significant at the 0.017 level, although with smaller magnitudes, which indicates a continuing but reducing negative effect of past sentiment on current stock returns. From SI_6 onwards to SI_10, the coefficients are not statistically significant at conventional levels (p-values are above 0.1), suggesting that the impact of investor sentiment from six days prior or more does not have a statistically discernible effect on current stock returns.

Overall, the FGLS regression confirms the nuanced temporal dynamics observed between investor sentiment and stock returns. It supports the hypothesis that while investor sentiment has an immediate and significant impact on stock returns, this influence is counteracted in the following days, with its detectability in stock returns dissipation after three days.

V. Conclusion

This study leverages the bullish index from Guotai'an Databank's public opinion archives as a proxy for individual investor sentiment at the stock level, utilizing panel data from 144 constituent stocks of the CSI 300 Index spanning from January 16, 2019, to December 1, 2021. Employing FGLS and LSDV techniques, this paper addresses issues of homoskedasticity and heteroskedasticity, as well as contemporaneous correlation and endogeneity within the panel data.

From the detailed analysis, this research draws two principal conclusions:

1) Effective Proxy for Investor Sentiment: The findings validate the bullish index as an effective measure for gauging individual investor sentiment within China's securities market at the individual stock level. This result provides robust evidence of how investor sentiment functions in practical market settings and showcases the potential and effectiveness of using online text data for sentiment analysis;
2) Impact of Individual Investor Sentiment on Stock Returns: The empirical results demonstrate that current individual investor sentiment significantly enhances contemporaneous stock returns, while sentiment lagged by one to three days has a pronounced negative impact on stock returns. These findings highlight the market's rapid response to changes in investor sentiment and how such sentiment influences stock prices in the short term. Recognizing this pattern is crucial for market participants when considering the timing and depth of sentiment's impact in investment decision-making processes.

This research not only enriches the theoretical framework concerning the relationship between individual investor sentiment and market performance within behavioral finance but also offers practical guidance for regulatory bodies and investors on utilizing investor sentiment information to predict market dynamics and formulate investment strategies. Future studies might explore additional types of sentiment indicators and expand the research to stock markets in other countries or regions to verify the generalizability and robustness of these findings. Further investigation into the interplay between sentiment indicators and other fundamental market factors would also enhance understanding of how investor sentiment operates within complex market environments.

Conflict of Interest

The author declares no conflict of interest.

References

Aboody, D., Lehavy, R., Even-Tov, O., *et al.* 2018. Overnight returns and firm-specific investor sentiment. ***Journal of Financial and Quantitative Analysis***, 53(2): 485–505.

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., Stein, J. C. 2004. Market liquidity as a sentiment indicator. ***Journal of Financial Markets***, 7(3): 271–299.

Baker, M., Wurgler, J. 2006. Investor sentiment and the cross-section of stock returns. Journal of Finance, 61(4): 1645–1680.

Barberis, N., Shleifer, A., Vishny, R. 1998. A model of investor sentiment. ***Journal of Financial Economics***, 49(3): 307–343.

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.

Bollen, J., Mao, H., Zeng, X. 2011. Twitter mood predicts the stock market. ***Journal of Computational Science***, 2(1): 1–8.

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

Da, Z. J., Engelberg, J., Gao, P. 2011. In search of attention. ***The Journal of Finance***, 66(5): 1461–1499.

De Long, J. B., Shleifer, A., Summers, L. H., *et al.* 1990. Noise trader risk in financial markets. ***Journal of Political Economy***, 98(4): 703–738.

Fama, E. F., French, K. R. 1993. Common risk factors in the returns on stocks and bonds. ***Journal of Financial Economics***, 33(1): 3–56.

Fernandes, C. M. D., Goncalves, P. M. M. G., Vieira, E. F. S. 2013. Does sentiment matter for stock market returns? Evidence from a small European market. ***Journal of Behavioral Finance***, 14(4): 253–267.

Fisher, K. L., Statman, M. 2003. Consumer confidence and stock returns. ***The Journal of Portfolio Management***, 30(1): 115–127.

Fisher, K. L., Statman, M. 2000. Investor sentiment and stock returns. ***Financial Analysts Journal***, (4): 16–23.

Frazzini, A., Lamont, O. A. 2007. The earnings announcement premium and trading volume. Available at SSRN 986940.

Gao, Z. H., Ren, R., Zhang, B. 2020. Media coverage and the cross-section of stock returns. ***Journal of Financial and Quantitative Analysis***, 55(2): 549–580.

Hausman, J. A. 1978. Specification tests in econometrics. ***Econometrica: Journal of the Econometric Society***, 1251–1271.

Kim, S., Kim, D. 2014. Investor sentiment from internet message postings and the predictability of stock returns. ***Journal of Economic Behavior & Organization***, 107: 691–729.

Lee, C. M. C., Shleifer, A., Thaler, R. H. 1991. Investor sentiment and the closed-end fund puzzle. ***The Journal of Finance***, 46(1): 75–109.

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.

Neal, R., Wheatley, S. 1998. Do measures of investor sentiment predict returns? ***Journal of Financial and Quantitative Analysis***, 33(4): 523–547.

Oskamp, S. 2000. A sustainable future for humanity? How can psychology help? ***American Psychologist***, 55(5): 496.

Ranco, G., Aleksovski, D., Caldarelli, G., *et al*. 2015. The effects of Twitter sentiment on stock price returns. ***PloS One***, 10(9): e0138441.

Renault, T. 2017. Intraday online investor sentiment and return patterns in the US stock market. ***Journal of Banking & Finance***, 84: 25–40.

Sabherwal, S., Sarkar, S. K., Zhang, Y., *et al*. 2008. Online talk: does it matter? ***Managerial Finance***, 34(6): 423–436.

Slovic, P., Lichtenstein, S. 1971. Comparison of Bayesian and regression approaches to the study of information processing in judgment. ***Organizational Behavior and Human Performance***, 6(6): 649–744.

Slovic, P., Fischhoff, B., Lichtenstein, S. 1982. Why study risk perception? ***Risk Analysis***, 2(2): 83–93.

Sun, L., Najand, M., Shen, J. 2016. Stock return predictability and investor sentiment: a high-frequency perspective. ***Journal of Banking & Finance***, 73: 147–164.

Tumarkin, R., Whitelaw, R. F. 2001. News or noise? Internet postings and stock prices. ***Financial Analysts Journal***, 57(3): 41–51.

Tversky, A., Kahneman, D. 1971. Belief in the law of small numbers. ***Psychological Bulletin***, 76(2): 105.

Tversky, A., Kahneman, D. 1974. Judgment under uncertainty: Heuristics and biases. ***Science***.