creative commons

COMMONS DEED

저작자표시-비영리-변경금지 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.

다음과 같은 조건을 따라야 합니다:

저작자표시. 귀하는 원저작자를 표시하여야 합니다.

비영리. 귀하는 이 저작물을 영리 목적으로 이용할 수 없습니다.

변경금지. 귀하는 이 저작물을 개작, 변형 또는 가공할 수 없습니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.

이것은 이용허락규약(Legal Code)을 이해하기 쉽게 요약한 것입니다.

Disclaimer

dCollection

조선대학교
CHOSUN UNIVERSITY

2022 Aug

Master's dissertation

A Study on the Relationship between Investor Herding and Social Media Sentiment in the Korean Stock Markets during COVID-19

Chosun University

Division of Business Administration

Jinjoo Yoon

조선대학교
CHOSUN UNIVERSITY

A Study on the Relationship between Investor Herding and Social Media Sentiment in the Korean Stock Markets during COVID-19

COVID-19 기간 동안 국내 주식시장의 투자자 군집행동과
소셜 미디어 심리의 관계에 대한 연구

2022년 8월 26일

조선대학교　대학원

경영학과

윤　진　주

조선대학교
CHOSUN UNIVERSITY

A Study on the Relationship between Investor Herding and Social Media Sentiment in the Korean Stock Markets during COVID-19

지도교수 오 갑 진

이 논문을 경영학석사학위 신청 논문으로 제출함

2022년 4월

조선대학교 대학원

경영학과

윤 진 주

조선대학교
CHOSUN UNIVERSITY

윤진주의 석사학위논문을 인준함

위원장 조선대학교 교수 이 계 원 (인)

위 원 조선대학교 교수 이 현 철 (인)

위 원 조선대학교 교수 오 갑 진 (인)

2022년 5월

조선대학교 대학원

조선대학교
CHOSUN UNIVERSITY

Contents

조선대학교
CHOSUN UNIVERSITY

List of Tables

조선대학교
CHOSUN UNIVERSITY

List of Figures

조선대학교
CHOSUN UNIVERSITY

Abstract

COVID-19 기간동안 국내 주식시장의 투자자 군집행동과 소셜 미디어 심리의 관계에 대한 연구

윤 진 주
지도교수 : 오갑진
경영학과
조선대학교 대학원

본 논문에서는 머신러닝을 활용하여 소셜 미디어에서 형성된 투자자들의 심리지수를 직접적으로 측정하여 분석에 사용하였다. 매수의견점수를 나타내는 Bullish 변수와 소셜 미디어의 비정상적인 정보 창출 활동을 나타내는 AICA 변수가 투자자들의 허딩에 미치는 영향을 실증 분석하였다. 그리고 COVID-19 기간 동안 소셜 미디어로부터 생성된 변수들이 투자자의 군집행동에 미치는 영향력을 실증 분석하였다. 실증 분석결과. AICA 변수는 개인투자자 군집행동과 통계적으로 유의한 양의 상관성을 가지는 것을 확인하였고 매수의견 점수를 나타내는 bullish변수와는 모든 투자자의 군집현상과 통계적으로 유의한 음의 관계를 가지는 것을 확인하였다. COVID-19 이후 Bullish변수가 개인투자자의 군집현상에 미치는 영향이 더 강해진 것을 확인하였다. 따라서 본 연구에서는 소셜 미디어에서 형성된 정보가 투자자들의 군집행동을 설명할 수 있는 새로운 요인이라는 것을 실증분석을 통해 밝혔다.

조선대학교
CHOSUN UNIVERSITY

I. Introduction

Herding means imitating the behavior of others or following decisions (Barber et al. 2009). When investors impact the market while herding, they form bubbles, make an abnormal impact, or create greater momentum. Examples of herding are bank runs and system collapses (Diamond et al. 1983). Therefore, it is important to know which factors cause the herding. Herding has been studied by dividing it into several investor types. Looking at the previous studies on the institutional investors' herding, money managers may mimic the actions of other money managers in order to preserve reputation and compensation (Scharfstein et al. 1990; Trueman 1994; Graham 1999). Institutional investors tend to follow each other in buying and selling the same securities and their own lag trades and that they tend to follow momentum strategies (Sias 2004). The institution has been herding like this in the past and now.

The study on the herding of individual investors is very sparse. In the past, individual investors were considered as noise trader. Therefore, they do not herd. But recently, the use of social media has increased rapidly, individual investors still have a noise characteristics, but there are various sources to obtain information, and there are many phenomena that herd through various information channels. Therefore, if we can confirm the effect of activity and opinion of social media which can play an important role in the market through the study, we will observe whether the herding affects the market. In particular, we conducted research on the data of Naver Financial stock discussion room, which has a very large activity in Korea, is working well. Gamestop is a representative example of individual investors who are able to herding on social media and focus on specific stocks to disturb the market. In other words, it is an example that shows that individuals are united and affect the price of the market. After COVID-19, the trading volume of individual investors increased. The number of posts posted by Naver Financial stock discussion room, which has a

조선대학교
CHOSUN UNIVERSITY

large activity of social media, also increased. The correlation coefficient of these two is 0.78, which is large. So, as opinion sharing becomes active in social media, we want to see if opinion is formed and this affects the market by creating a herding of individual investors.

It is difficult to use the sentiment dictionary built in general words because there is no specialized sentiment dictionary in the financial market in Korea and words are used interchangeably according to the context. Also, there is a limit to classify the sentiment of text using common sentiment dictionary because the writers who create post represent differently according to the source of text, and the words according to the source of text are also changed. In addition, financial markets are sensitive to economic conditions, and the sentiments of the words used in the text may vary depending on the time when the text is created. Therefore, there is a limit to use them for analysis using the existing sentiment dictionary. This study aims to classify the posts' sentiment into positive opinions and negative opinion by using the classification model of machine learning without using the sentiment dictionary. The more positive sentiment the post has, the more buying opinion we can see and the more negative sentiment the post has, the more selling opinion we can see. SVM, Support Vector Machine, on of the machine learning methods, is used for classifying posts into buying and selling opinions according to the sentiment. The purpose of this study is to identify the degree of opinion score, which is the daily investment sentiment index of individual companies, by integrating the posts of each company classified as buying and selling opinions by the classification model of machine learning. The process of collecting and processing the posts is carried out by Python. The reference point for distinguishing the upper and lower returns is the return rate equivalent to the extreme 1% of the daily return.

In this study, we use a Bullish measure presented in the paper of Antweiler and Frank (2004), as an investment sentiment index for the posts. We follow Lakonishok

조선대학교
CHOSUN UNIVERSITY

et al. (1992) herding measure to estimate daily trading herding index by each stock and each type of investor covering retail, institutional and foreign investor. We modify the model of Da et al. (2011; 2015) to compute the Abnormal Search Volume(ASVI) index. They download each listed firm's historical search volume index(SVI) from Google Trends because Google Search Volume Index can be the proxy for the information demand of informal individual investors. However, Google Trends has a disadvantage of including the amount of searched for people who do not invest in stocks. Therefore, in this paper we redefine information creation activity (ICA, hereafter) as a variable that indicated information creation and supply by utilizing the number of daily posts in Naver Financial stock discussion room where investors form their opinions directly by modifying the model of abnormal search volume index provided by Da et al. (2011; 2015).

Regression analysis is conducted to investigate the abnormal effect of social media information creation activity, AICA on the herding behavior of individual investors. Bullish is negatively related to herding when sentiment is not good. The main variables have stronger impact in investor's herding behavior after COVID-19. Finally the individual investors' transaction volumes are divided into three groups and regression analysis is conducted for each group.

The sample firms that have transactions are chosen from January 2018 to December 2021, so 2237companies are selected as samples. In this study, the posts of Naver Financial stock discussion room for four years from 2018 to 2021 are collected by web crawling method using Python. The stock price data about the sample firms were collected from DataGuide provided by FnGuide. Empirical results are as follow.

Regression analysis is conducted to investigate the effect of social media information creation activity, ICA, on the herding behavior of individual investors. The abnormal information creation activity(AICA, hereafter) variable, which means information creation activity of social media, is created by retail investors, so it is assumed that it

조선대학교
CHOSUN UNIVERSITY

has a positive correlation. In other words, it can be confirmed that the information creation supply of social media leads to actual transactions. Also, we could observe that AICA have a negative effect on institutional and foreign investors' herding, because they are not the main subjects of writing in social media. Institutional investors may obtain more advanced and complicated information than individual investor, who can obtain financial information using online surfing.

Regression analysis is conducted to check whether the Bullish is negatively related to herding when sentiment is not good. And it is applied to all types of investors. The Bullish variable has statistically significant negative relationship with the three investor types of herding. That is, when sentiment and bullish is not good, all three types of investor can confirm their herding behavior. Another egression analysis is conducted to check whether the main variables have stronger impact in investor's herding behavior after COVID-19. The impact of the Bullish variable on retail investor herding has become stronger after COVID-19. The effect of the bullish variable on retail investor herding is stronger after COVID- 19. In other words, this indicates that individual investors are herding when they sell and do not herding when they buy. Foreigner herding shows a pattern of opposition before and after COVID-19, which should be confirmed through further verification. Finally, the individual investors' transaction volumes are divided into three groups and regression analysis is conducted for each group. In case of bullish, it is confirmed that there is a statistically significant negative relationship regardless of the trading activity of individual investors. AICA, which indicates information creation activity, shows that there is no correlation at all in high group with large trading volume of individual investors. n other words, it is confirmed that it does not affect the herds at all. The reason for this result is that the herding is to be explained, and this result seems to have come out because the value of the herding itself is small. In addition, most of the transactions consist of individual investors, which can be interpreted as having a

조선대학교 CHOSUN UNIVERSITY

noise characteristic, so the herding seems to have decreased and these results are likely to come out.

The contribution of this paper as follows. We propose investor sentiment constructed based on informal text specially adapted to the context of the dataset without relying on preexisting dictionaries. The study on the herding of individual investors is very sparse. We use variables using social media to analyze the Herding of individual investors. And the effect of social media activity and opinion on investors' herding is confirmed, suggesting that these can be used as a variable to explain the herding. The effect of variables formed from social media before and after COVID-19 on investors' herding is confirmed.

The remainder of this study is organized as follows. Section 2 reviews the past literature. Section 3 offers the data and methodology. Section 4 presents the results and analysis. The final section consists of the conclusions.

조선대학교
CHOSUN UNIVERSITY

II. Literature Review

With the rapid increase in the amount of text data that can directly represent the writer's psychology, the improvement of computer performance and the development of analysis technology, the amount of text data, which is atypical data, can be handled at a lower calculation cost. The approach to build investment psychological indicators through text data has increased. This approach can be divided into two parts. One is to analyze the opinions of texts, and then to establish a list of emotional words such as positive and negative words, and then to measure the text's sentiments and to examine the relationship with the stock market. In the early days, the text data is measured using Harvard IV-4 Dictionaries, a common emotion dictionary that can be used in all fields, and the relationship with stock returns is analyzed (Tetlock 2007; Kothari et al. 2009). However, there is a limit that the classification of the texts using such public emotion dictionary did not give sufficient accuracy (Li 2010), and Loughran and McDonald (2011) established and presented a specialized emotion dictionary in the stock market.

The results of the analysis of the prediction of stock returns by classifying the sentiments by the dictionary of Loughran and McDonald (2011) and the Harvard IV-4 Dictionaries, which is a common emotion dictionary, are as follows: First, the predictions of the sentiments were high when the sentiments were classified into the stock market specific emotion dictionary (Henry et al. 2016), and the current sentiments of Loughrna and McDonald are estimated to be high. It is actively used in research in English-speaking countries. In the case of analyzing the emotion of the text at the present time using the existing emotional dictionary, it is difficult to divide the text containing the words mainly used on the Internet because the words that can be used as other emotions can not be accurately classified according to the words or times that do not exist in the past (Kraus et al. 2017; Ke et al. 2019). There is a

조선대학교
CHOSUN UNIVERSITY

limit to analyzing the investment sentiment of text data based on the previous. Next, there is machine learning as a way to classify the text's sentiments even if, there is no existing emotional dictionary. Machine learning is not only dependent on the presence of emotional dictionary, but also is becoming more popular as computer technology development makes it easier to handle large amounts of unstructured data (Ke et al. 2019). Accordingly, studies are increasing to measure investment sentiment and analyze the relationship with stock returns by classifying the text data through the classification model of machine learning. Antweiler and Frank (2004), an early study that analyzed online posts, classified the posts' sentiments and measured the scores of the purchase opinions and the degree of disagreements to examine the effects of stock returns and volatility. As a result, the higher the buyout opinions of the posts, the lower the return rate of the next day. It is more accurate, and the stock market movement is related to the active degree of the individual investor's sentiment and discussion board, which are the main users of the post. Unlike foreign countries where there is a specialized emotional dictionary in the stock market, Korea does not have a specialized emotional dictionary in the stock market, so there is no active study to analyze the stock market through text data. The domestic research using text data for stock market analysis measured investment sentiment using Naver Financial stock discussion room posting, and the number of posts by investment opinion is analyzed by writers as investment sentiment, and the relationship with herding is analyzed.

In economics and finance with the term herding or herd behavior, we mean the process where economic agents are imitating each other actions or base their decisions upon the actions of others (Spyrou 2013). In herding literature, there are so many herding measurement. The cross-sectional absolute deviation (CSAD) and the cross-sectional standard deviation of returns (CSSD) regards the overall herding existing in a specific market and do not allow researchers to distinguish between

조선대학교 CHOSUN UNIVERSITY

different types of investors as important (Christie et al. 1995; Chang et al. 2000; Demirer et al. 2006; Chiang et al. 2010; Lao et al. 2011; Chong et al. 2017). So, we follow the LSV measure (Lakonishok et al. 1992) to estimate herding behavior of each investor type. The LSV is able to categorize market participants into different types of investors (Wermers 1999; Shyu et al. 2010; Choi et al. 2015). Li et al. (2017) apply a trading volume based herding measure and bring to a conclusion that individual investors are more sensitive to public information than institutional investors, and exhibit more significant herding to public information.

The most of the literature about herding exploring different types of investors explores institutional investor herding behavior. Few studies have concentrate on retail investor herding. Barber et al. (2009) find that retail investor herding behavior can be systematic and especially that buyout behavior could be driven by past performance and the abnormal trading volume observed from individual invetor trading. Therefore, they document systemic trading among individual investor as noise traders.

In the Korea stock market where individual investors' trading volume is numerically superior, so there is a crucial need to understand individual investor herding in depth. This would also be a substantial contribution to the literature. Da et al. (2011) first contend that abnormal search volume index (ASVI) is a better proxy of investors searching for information. Hsieh et al. (2020) discover that ASVI as a relevant proxy of the information demand allows us to examine the relationship between information demand and individual investor herding behavior.

De et al. (1990) propose that noise traders who are overreaction or undereaction to information cause stock prices to momentarily diverge from their appropriate price. There are two categories of investor sentiment indicators, direct emotional and indirect indicators. The former is constructed based on the survey of bullish/bearish views of investors for the future market. For example, Shiller et al. (1996) carry out a semi annual survey of institutional investors' viewpoints on the US and Japanese

조선대학교
CHOSUN UNIVERSITY

stock markets to obtain sentiment indicators of stock market. The latter is computed from market observation, and includes excess abnormal return (Barber 2008), trading volume (Gervais et al. 2001; Barber 2008). It is relatively difficult to directly measure investor sentiment and literature on it is scarce. Da et al. (2015) examine whether investors can predict a company's earnings and profits by using Google search engine to search for the company's manufactured goods. They discover that the higher the ASVI, the more precise the prediction for the positive profits surprise, systematized unexpected earnings, and abnormal earnings during the earnings announcement period. Bank et al. (2011) also indicate that there is a positive association between searching in online and trading movements, which implies that search engines can lower information asymmetry and enhance the individual investors' readiness to invest.

There are also previous studies on the herding of Korean market and investors. The Korean market is examined by Choea et al. (1999) who employ order and trade data between 1996 and 1997 and find crucial evidence of evidence of herding by foreign investors before the crisis. Kim and Wei (2002) examine the behavior of investors in Korea and find that institutional investors herd significantly less than retail investors and that resident institutional and retail investors herding considerably less than non-resident investors. Choi and Yoon (2020) research herding behavior and the relationship between herding behavior and investor sentiment. Korean investor sentiment is proxy through the VKOSPI that is Korea's proxy implied volatility index based on the KOSPI200 option. So we directly, the sentiment is measured and it uses in the analysis.

We modify the model of Da et al. (2019) to compute the Abnormal Information Creation Activity (ASVI) index. They store each listed firm's historical search volume index (SVI) from Google Trends because Google Search Volume Index can be the proxy of the information demand of informal retail investors. However, Google Trends has a disadvantage of including the amount

조선대학교
CHOSUN UNIVERSITY

of searched for people who do not invest in stocks.Therefore, in this paper we redefined information creation activity (ICA, hereafter) as a variable that indicated information creation and supply by utilizing the number of daily posts in Naver Financial stock discussion room where investors form their opinions directly by modifying the model of Abnormal Information Creation Activity index.

The study on the herding of individual investors is very sparse, and the relationship between information creation of social media and the segment formed here is directly measured and the herding is studied. We set up the following hypotheses and conducted regression analysis to verify them.

Hypothesis 1. AICA variable, which means the information creation activity of social media is positively related to retail investor herding behavior.

We expect there is a positive significant coefficient generated by the AICA variable when dependant variable, Herd, is retail investor herding behavior. Da et al. (2011) discover that utilizing the Google search engine is the most convenient method for individual investors to obtain numerous information. Individual investors are relatively inferior to another investors, so they will use the stock discussion room to obtain information and share opinions, and the main user of the stock discussion room may be retail investors (Da et al. 2011; 2015). Therefore, in this study, regression analysis is conducted to investigate the effect of social media information creation activity, AICA, on the herding behavior of individual investors

Hypothesis 2. Bullish variable is negatively related to all types of investors' herding.

We conduct the regression which includes investment sentiment indicator, Bullish.

조선대학교
CHOSUN UNIVERSITY

According to the behavioral finance, when the direction of utility is the loss than profit, it more sensitive, it reacts (Kahneman 1979). In other words, there is greater pain when it falls, decreases or is bad. Based on this theory, regression is conducted to check whether the bullish is negatively related to herding when bullish is not good. And we check that it is applied to all types of investors.

Hypothesis 3. relationship between variables and retail investor's herding is likely to becomes stronger after COVID-19.

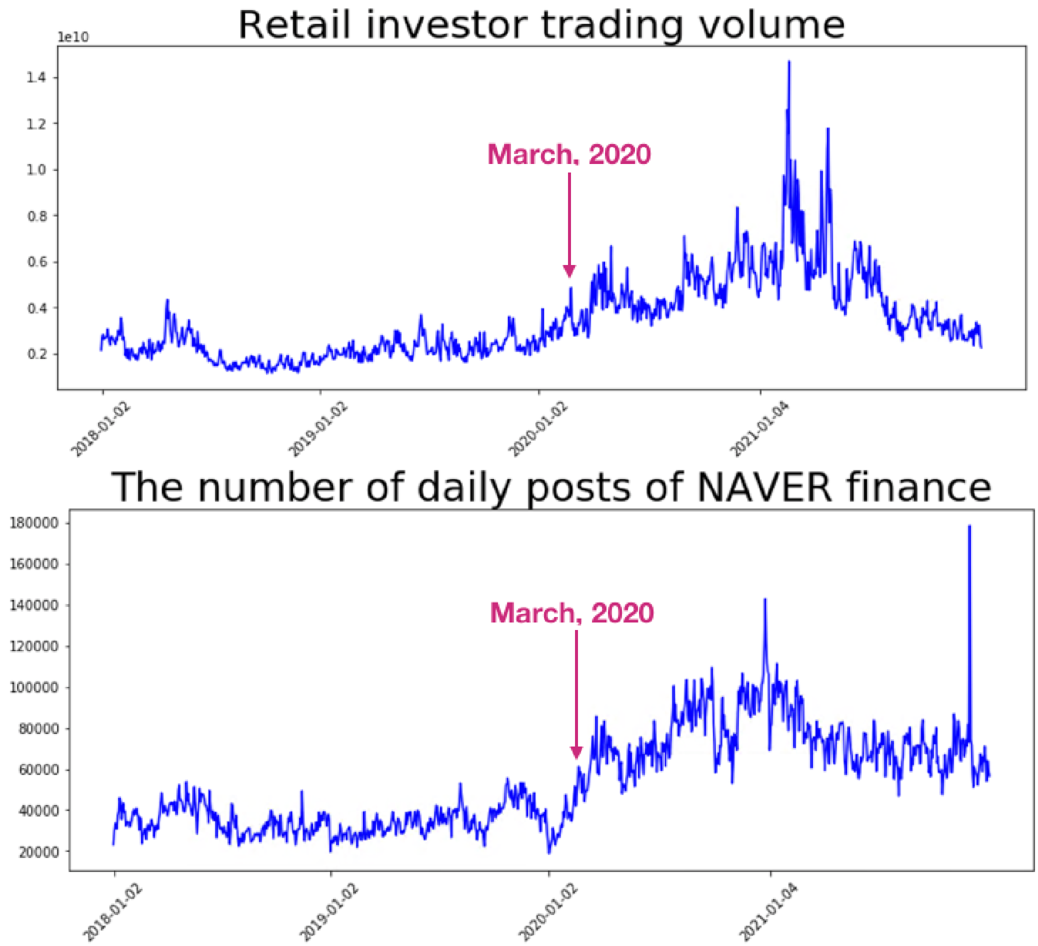
We estimate the regression by including a dummy variable equal to 1, if the value after March 22, 2020, when social distance was implemented in Korea, and 0 otherwise. After COVID-19, retail investors increased and social media activity also boosts in Figure 1. So regression is conducted to check whether the variables have stronger impact in retail investor herding behavior after COVID-19.

Finally, the individual investors' transaction volumes are divided into three groups and regression analysis is conducted for each group. Since the main users who write posts are individual investors, we looked at the group's herding divided according to the proportion of individual investors.

조선대학교
CHOSUN UNIVERSITY

[Figure 1] Trading volume of retail investor
and the number of daily posts of Naver Financial

This figure shows that trading volume of retail investor and the number of daily posts of Naver Financial during the sample period.



조선대학교
CHOSUN UNIVERSITY

III. Data and Methodology

3.1 Data

Naver, the largest portal site in Korea, provides information on the Korean financial market through Naver Financial (http://finance.naver.com), which provides stock discussion room services that allow stock investors to share their opinions and information. Naver Financial stock discussion room is most actively used by users and has the largest number of posts among the stock relate social media in Korea.

This study utilizes web-crawling method to download the posts of each company from January 2018 to December 2021 in the stock discussion room provided by Naver Financial. Downloaded posts are stored daily for each company and the posts on the website stored are as shown in Figure 2. Each posts consist of a part of writers' ID and IP, the date and time of writing and the number of good, bad and lookup. If the part of writer's ID and IP provided by the site are the same, it is assumed to be the same writer and used for analysis. In this study, the sample data is selected from 2237 listed companies that are trades in KOSPI and KOSDAQ market from January 2018 to December 2021.

Figure3 shows the distribution of the average number of posts per day (panel A) and per time zone (panel B) used on the stock discussion room. Most of posts can be seen on weekdays, when stock trading takes place (am 9:00) before closing (pm 3:30).Therefore, this study used only the posts posted during the stock market opening during trading days when measuring investment sentiment.

The financial market data including adjusted price, daily transaction volume, investor type and so on is acquire from FnGuide. By using the data obtained from FnGuide, the market capitalization is calculated using the adjusted price and the number of common stock shares. The return is calculated from the adjusted price of individual companies. STD is standard deviation of daily return for individual stock during the quarter.

조선대학교 CHOSUN UNIVERSITY

조선대학교
CHOSUN UNIVERSITY

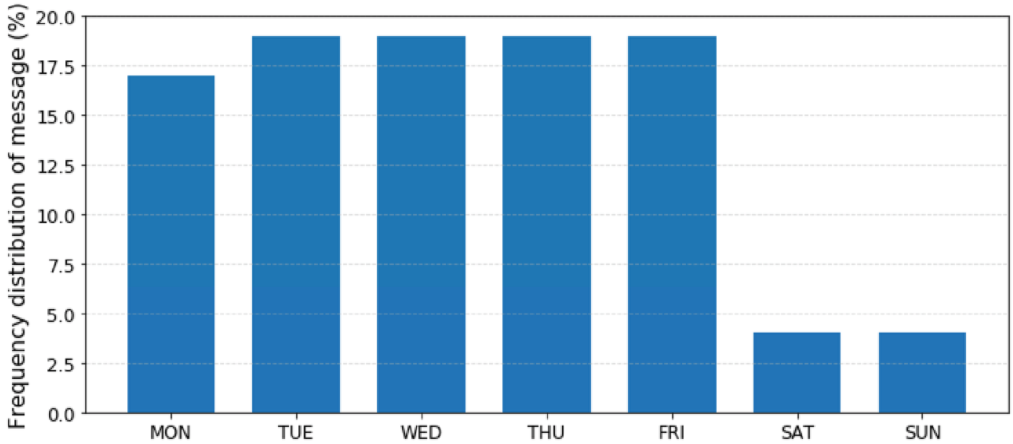
[Figure 2] The posts of Naver Financial stock discussion room

The Naver Financial stock discussion room is a bulletin board that allows stock investors to exchange information and opinion for each company. These pictures are screen captured by posts of the stock debate room. In this study, we used message and date of the posts.

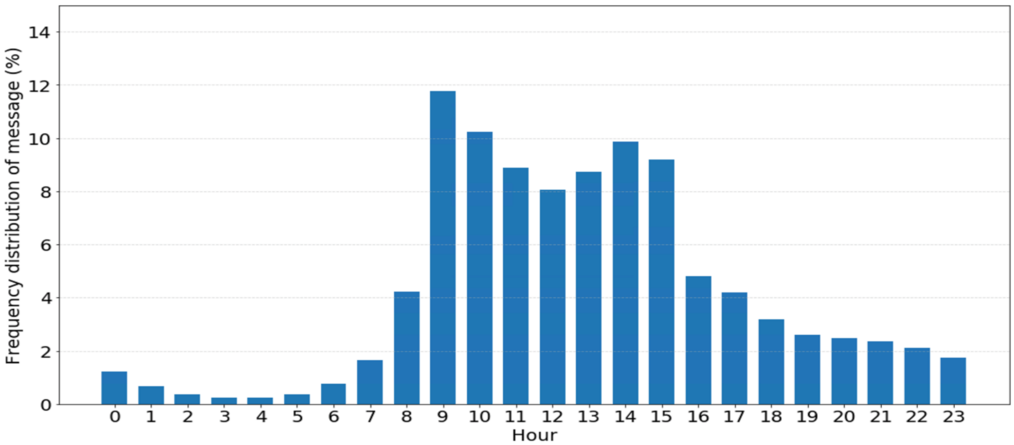
조선대학교
CHOSUN UNIVERSITY

[Figure 3] Distribution of posts by day and time

Panel A. Distribution of posts by day



Panel B. Distribution of posts by time



조선대학교
CHOSUN UNIVERSITY

3.2 Methodology

3.2.1 Opinion labeling process

This papar's purpose is to analyze the relationship between investment sentiment formed on Naver Financial stock discussion room and each type of investors' herding by measuring the degree of buying and selling opinions in the posts. In this study, text data are collected from the Naver Financial stock discussion room to under stand the investment sentiment formed social media.

The more positive sentiment the post has, the more buying opinion we can see and the more negative sentiment the post has, the more selling opinion, we can see. SVM, Support Vector Machine, on of the machine learning methods, is used for classifying posts into buying and selling opinions according to the sentiment. The purpose of this study is to identify the degree of opinion score, which is the daily investment sentiment index of individual companies, by integrating the posts of each company classified as buying and selling opinions by the classification model of machine learning. The process of collecting and processing the posts is carried out by Python.

The general method of classifying text data is to assign the sentiment score of each word using the sentiment dictionary when there is an sentiment dictionary in which the sentiments of the word are classified, and then sentiment scores are given by summing it up, and to classify text data according to the characteristics through the machine learning method when there is enough text data without the sentiment dictionary.

It is difficult to use the sentiment dictionary built in general words because there is no specialized sentiment dictionary in the financial market in Korea and words are used interchangeably according to the context. Also, there is a limit to classify the sentiment of text using common sentiment dictionary because the writers who create

조선대학교
CHOSUN UNIVERSITY

post represent differently according to the source of text, and the words according to the source of text are also changed. In addition, financial markets are sensitive to economic conditions, and the sentiments of the words used in the text may vary depending on the time when the text is created. Therefore, there is a limit to use them for analysis using the existing sentiment dictionary. This study aims to classify the posts' sentiment into positive opinions and negative opinion by using the classification model (SVM) of machine learning without using the sentiment dictionary. Learning data for applying classification model is constructed by using the daily return calculated by adjusted by price of individual firms. After sorting the daily returns, the posts of the companies with the highest returns are regarded as buying opinions and on the contrary the posts of the companies with the lowest returns are assumed as selling opinions. The reference point for distinguishing the upper and lower returns is the return rate equivalent to the extreme 1% of the daily return.

3.2.2 Bullish measure

In this study, we use Bullish measure presented in the paper of Antweiler and Frank (2004), as an investment sentiment index for the posts. The Bullish measurement is as follows :

$$Bullish_{i,t} = \ln\frac{1 + M_{i,t}^{BUY}}{1 + M_{i,t}^{SELL}} \tag{1}$$

$Bullish_{i,t}$ the measure of the buyout opinion score, is calculated by the daily (t) number of buying opinion posts ($M_{i,t}^{BUY}$) of the firm (i) and the daily (t) number of selling opinion post ($M_{i,t}^{SELL}$) of the firm(i). This measurement has a large value as the number of posts of the buyout opinion increases. The daily number of posts used

조선대학교
CHOSUN UNIVERSITY

to calculate $Bullish_{i,t}$ is calculated by including only the number of posts posted on the trading day and time because the writers who write on the stock discussion room generally tend to write when the stock market is opened according to Figure 3.

3.2.3 Herding Measure

We follow Lakonishok et al. (2004) herding measure to estimate daily (t) trading herding index by each stock (i) and each type of investor covering retail, institutional and foreign investor. Herding measure is as follows :

$$H_{i,j,t} = B_{i,j,t} / (B_{i,j,t} + S_{i,j,t}) \tag{2}$$

$$Herd_{i,j,t} = (|H_{i,j,t} - p_{j,t}| - AF_{i,j,t}) \times 100 \tag{3}$$

where $B_{i,j,t}$ is the number of buying trades of stock (i) for investor type (j) in a given day (t) and $S_{i,j,t}$ is the number of selling trades of stock (i) for investor type (j) in a given day (t). And where $H_{i,j,t}$ is the number of buying trades to total number of trades of stock(i) for investor type (j) in a given day (t). $p_{j,t}$ is the average value of $H_{i,j,t}$ for all stocks for type (j) of investor in a given day (t). $AF_{i,j,t}$ is the adjustment factor for herding measure that adjusts the scale different for trading volume. $AF_{i,j,t}$ accounts for the fact that under the null hypothesis of no herding, which is when the probability of a investor being a net buyer of any stock (i) is $p_{j,t}$, the absolute value of $H_{i,j,t} - p_{j,t}$ is greater than or equal to zero. So $AF_{i,j,t}$ is defined the expected value of $|H_{i,j,t} - p_{j,t}|$ under the null hypothesis of no herding. Since $B_{i,j,t}$ follows a binomial distribution with probability $p_{j,t}$ of success. So $AF_{i,j,t}$ is easily calculated given $p_{j,t}$ and the number of investors active in stock(i) in that time (t). For any stock (i), when the number of investors active, $AF_{i,j,t}$ decreases.

조선대학교
CHOSUN UNIVERSITY

3.2.4 AICA Measure

We modify the model of Da et al.[9, 10] to compute the Abnormal Information Creation Activity (AICA) index. They download each listed firm's historical search volume index (SVI) from Google Trends because Google Search Volume Index can be the proxy for the information demand of informal individual investors. However, Google Trends has an disadvantage of including the amount of searched for people who do not invest in stocks. Therefore, in this paper we redefine information creation activity (ICA) as a variable that indicated information creation and supply by utilizing the number of daily posts in Naver Financial stock discussion room where investors form their opinions directly by modifying the model of abnormal search volume index provided by Da et al. (2011; 2015). We adjust ICA variable to control for time trends. The definition of AICA is the log of ICA during the day minus the log of median ICA during the previous eight days. The AICA measure is as follows :

$$AICA_{i,t} = \log(ICA_{i,t}) - \log[Med(ICA_{i,t-1}, \ldots, ICA_{i,t-8})] \tag{4}$$

Where the $\log(ICA_{i,t})$ is the logarithm of ICA for each stock(i) in that time(t) and $\log[Med(ICA_{i,t-1}, \ldots, ICA_{i,t-8})]$ the logarithm of median value ICA for each stock(i) during the previous eight days.

3.2.5 Regression model

We set up the hypotheses and conduct regression analysis to verify them. Following studies by Romer (1979) and Li et al (2010) .who argue that investor herding may be affected by beliefs about other types of investors when the consensus occurs. So we add a control variable when the dependant variable of $Herd_{i,j,t}$ is retail investor herding behavior, $Herd_{i,k,t}$ and $Herd_{i,l,t}$ variable representing the control variable are institutional and foreign investor herding behavior and vice versa.

조선대학교 CHOSUN UNIVERSITY

We used these variables as follow : $Herd_{i,j,t}$ is the one type (j) of investor herding among the three types. $Herd_{i,k,t}$ and $Herd_{i,l,t}$ are the others of investor herding type except for the type (j) of investor herding. $AICA_{i,t}$ is abnormal ICA for stock (i) in time (t). $\ln SIZE_{i,t}$ is the value taking the natural log on the market capitalization. $RET_{i,t-1}$ is the past returns of stock (i) during the previous 1 day. $PER_{i,t}$ is the P/E ratio for stock (i) $\ln VOL_{i,t}$ is the value taking the natural log on trading volume for stock (i) in time (j). $STD_{i,t}$ is the standard deviation of daily returns for stock (i) during the previous 3 months. $\epsilon_{i,j,t}$ is disturbance term. The year fixed effect and industry fixed effect are applied to all regression models. The domestic market is expected to have an effect on each industry, and this effect is removed through industry fixed effect and analyze. The year fixed effect is also performed for the same reason.

To confirm our Hypothesis 1, "AICA variable which means the information creation activity of social media is positively related to retail investor herding behavior", we expect there is a positive significant coefficient generated by β_1 in equation 5 when dependant variable $Herd_{i,j,t}$ is retail investor herding behavior (Da et al. 2011; 2015). Individual investors are relatively inferior to another investors, so they will use the stock discussion room to obtain information and share opinions, and the main user of the stock discussion room may be retail investors. Therefore, in this study, regression analysis is conducted to investigate the effect of social media information creation activity, AICA, on the herding behavior of individual investors.

$$\begin{aligned} Herd_{i,j,t} = \alpha_0 + \ & \beta_1 \times AICA_{i,t} + \beta_2 \times Herd_{i,k,t} + \beta_3 \times Herd_{i,l,t} + \beta_4 \times \ln SIZE_{i,t} \\ & + \beta_5 \times RET_{i,t-1} + \beta_6 \times PER_{i,t} + \beta_7 \times \ln VOL_{i,t} + \beta_8 \times STD_{i,t} \\ & + Year\ dummy + Industry\ \ dummy + \epsilon_{i,j,t} \end{aligned} \tag{5}$$

조선대학교
CHOSUN UNIVERSITY

To confirm our Hypothesis 2, "Bullish variable is negatively related to all types of investors' herding", we expect there to be a positive significant coefficient generated by β_1 in equation 6 when dependant variable $Herd_{i,j,t}$ in all types of investors' cases. According to the behavioral finance, when the direction of utility is the loss than profit, it more sensitives, it reacts (Kahneman 1979). In other words, there is greater pain when it falls, decreases or is bad. Based on this theory, regression is conducted to check whether the bullish is negatively related to herding when bullish is not good. And we check that it is applied to all types of investors using Eq. (6)

$$\begin{aligned} Herd_{i,j,t} = \alpha_0 &+ \beta_1 \times Bullish_{i,t} + \beta_2 \times AICA_{i,t} + \beta_3 \times Herd_{i,k,t} + \beta_4 \times Herd_{i,l,t} \\ &+ \beta_5 \times \ln SIZE_{i,t} + \beta_6 \times RET_{i,t-1} + \beta_7 \times PER_{i,t} + \beta_8 \times \ln VOL_{i,t} \\ &+ \beta_9 \times STD_{i,t} + Year\,dummy + Industry\;\; dummy + \epsilon_{i,j,t} \end{aligned} \tag{6}$$

To confirm our Hypothesis 3, "relationship between variables and retail investor's herding is likely to becomes stronger after COVID-19", we estimate the regression as in Eq.(7) by including a dummy variable equal to 1, if the value after March 22, 2020, when social distance is implemented in Korea, and 0 otherwise. After COVID-19, retail investors increase and social media activity also increase. Therefore, regression analysis is conducted to check whether the variables have stronger impact in investor's herding behavior after COVID-19.

$$\begin{aligned} Herd_{i,j,t} = \alpha_0 &+ \beta_1 \times Bullish_{i,t} + \beta_2 \times Bullish_{i,t} \times COVID_t + \beta_3 \times AICA_{i,t} + \beta_4 \times AICA_{i,t} \times COVID_t \\ &+ \beta_5 \times Herd_{i,k,t} + \beta_6 \times Herd_{i,l,t} + \beta_7 \times \ln SIZE_{i,t} + \beta_8 \times RET_{i,t-1} + \beta_9 \times PER_{i,t} \\ &+ \beta_{10} \times \ln VOL_{i,t} + a_{11} \times STD_{i,t} + Year\,dummy + Industry\,dummy + \epsilon_{i,j,t} \end{aligned} \tag{7}$$

조선대학교
CHOSUN UNIVERSITY

IV. Empirical results

4.1 Summary statistics

For empirical analysis, the sample firms that have transactions are chosen from January 2018 to December 2021, so 2237companies are selected as samples. In this study, the posts of Naver Financial stock discussion room for four years from 2018 to 2021 are collected by web crawling method using Python. The stock price data about the sample firms are collected from DataGuide provided by FnGuide.

Table 1 is summary statistics of the variables used in regression analysis and it represents summary statistics of whole sample companies calculated with the average value of daily data. In the panel A of Table 1, when we look at the level of the herding, we can see that foreign and institutional investors' herding levels are higher than retail investor's herding level. It means foreign and institutional investors are more herding than retail investor. The kurtosis of the Bullish variable shows heavy tailed because it has a larger value than normal distribution. This means that the influence of each company is not the same, so it is meaningful to use the data in this study. In the panel B of Table 1 represents the correlation between the variables used in regression analysis. The correlation between Bullish and AICA that are main variables in this study is very low, which shows that there is no problem using this two variables for regression analysis at the same time.

조선대학교
CHOSUN UNIVERSITY

〈Table 1〉 Summary statistics

Panel A. Summary statistics of regression variables

Table 1 reports the summary statistics of regression variables used in regression analysis.AICA is abnormal information creation activity. Bullish is an investment sentiment indicator that represents buyout opinion score. Herd Retail, Herd Fore and Herd Insti represent retail, foreign and institutional herding behavior measure, respectively. lnSIZE is the value taking the natural log on the market capitalization. PER is the P/E ratio. lnVOL is the value taking the natural log on the trading volume. STD is the standard deviation of daily returns the previous 3 months. Panel A represents summary statistics of regression variables and Panel B represents correlation between regression variables.

Variable	Mean	Std	Med	Min	Max	Skew	Kurt	N
AICA	0.058	0.392	0.000	−1.602	3.275	0.820	2.587	801,443
Bullish	0.061	0.939	0.000	−9.325	8.893	0.356	11.326	801,443
Herding_Retail	5.356	5.954	3.251	0.000	50.711	2.099	5.342	801,443
Herding_Foreign	15.190	11.858	13.453	0.000	65.791	0.795	−0.006	801,443
Herding_Insti	23.497	17.818	28.880	0.000	71.670	0.027	−1.283	801,443
lnSIZE	26.523	1.441	26.258	21.695	33.929	0.981	1.303	801,443
RET	0.124	3.650	0.000	−30.004	30.056	1.700	15.371	801,443
PER	79.191	727.767	21.280	0.020	129000	58.381	5469.17	801,443
lnVOL	12.231	1.632	12.146	3.219	20.742	0.237	0.698	801,443
STD	0.031	0.015	0.027	0.003	0.153	1.452	2.926	801,443

조선대학교
CHOSUN UNIVERSITY

Panel B. Correlation between regression variables

Variable	AICA	Bullish	Herding Retail	Herding Foreigner	Herding Institution	lnSIZE	RET	PER	lnVOL	STD
AICA	1									
Bullish	0.116	1								
Herding Retail	0.025	−0.103	1							
Herding Foreigner	−0.008	−0.042	0.199	1						
Herding Institution	0.031	0.034	−0.114	0.106	1					
lnSIZE	−0.071	−0.076	0.352	−0.156	−0.457	1				
RET	0.118	0.114	−0.037	−0.011	0.006	−0.002	1			
PER	0.006	0.003	−0.015	−0.003	0.003	−0.003	0.005	1		
lnVOL	0.180	0.184	−0.211	−0.222	−0.011	0.113	0.123	0.021	1	
STD	−0.031	0.113	−0.278	−0.151	0.116	−0.201	0.061	0.031	0.415	1

조선대학교
CHOSUN UNIVERSITY

4.2 Herding behavior regression estimates

To confirm our Hypothesis 1, "AICA variable which means the information creation activity of social media is positively related to retail investor herding behavior", regression analysis is conducted to estimate the effect of AICA on the herding behavior of retail investors using Eq. (5). Table 2 shows that the Hypothesis 1 is correct. AICA variable has statistically significant positive relationship with retail investor herding.In other words, it can be confirmed that the information creation supply of social media leads to actual transactions. Also, we could observe that AICA have a negative effect on institutional and foreign investors' herding, because they are not the main subjects of writing in social media. Institutional investors may obtain more advanced and complicated information than individual investor, who can obtain financial information using online surfing (Da et al. 2011; 2015).

To confirm our Hypothesis 2, "Bullish variable is negatively related to all types of investors' herding", we apply Eq. (6), which includes investment sentiment indicator. According to the behavioral finance, when the direction of utility is the loss than profit, it more sensitives, it reacts (Kahneman 1979). In other words, there is greater pain when it falls, decreases or is bad. Based on this theory, regression is conducted to check whether the bullish is negatively related to herding when bullish is not good. And we check that it is applied to all types of investors using Eq. (6). The table 3 shows that Hypothesis 2 is correct. Bullish variable has statistically significant negative relationship with the all types of investors. That is, when sentiment or bullish is not good investor herding occurs.

To confirm our Hypothesis 3, "relationship between variables and retail

조선대학교 CHOSUN UNIVERSITY

investor's herding is likely to becomes stronger after COVID-19", we estimate the regression as in Eq. (7) by including a dummy variable equal to 1, if the value after March 22, 2020, when social distance was implemented in Korea, and 0 otherwise. After COVID-19, retail investors increased and social media activity also boosts. So regression is conducted to check whether the variables have stronger impact in retail investor herding behavior after COVID-19. Table 3 shows the effect of major variables on herding before and after COVID-19. The effect of the bullish variable on retail investor herding is stronger after COVID-19. In other words, this indicates that individual investors are herding when they sell and do not herding when they buy. Foreigner herding shows a pattern of opposition before and after COVID-19, which should be confirmed through further verification.

Finally the individual investors' transaction volumes are divided into three groups and regression analysis is conducted for each group. In Table 5, in case of bullish, it is confirmed that there is a statistically significant negative relationship regardless of the trading activity of individual investors. AICA, which indicates information creation activity, shows that there is no correlation at all in high group with large trading volume of individual investors. In other words, it is confirmed that it did not affect the herding at all. The reason for this result is that the herding is to be explained, and this result seems to have come out because the value of the herding itself is small. In addition, most of the transactions consist of individual investors, which can be interpreted as having a noise characteristic, so the herding seems to have decreased and these results are likely to come out.

조선대학교
CHOSUN UNIVERSITY

〈Table 2〉 Relationship between investor herding behavior and AICA

Table 2 reports the impact on AICA on the retail,foreign and institutional investor herding measure. This table presents parameter in Eq.(5). Herd Retail, Herd Fore and Herd Insti represent retail, foreign and institutional herding behavior measure, respectively. AICA is abnormal ICA. lnSIZE is the value taking the natural log on the market capitalization. PER is the P/E ratio. lnVOL is the value taking the natural log on the trading volume. STD is the standard deviation of daily returns the previous 3 months. The industry and the year effect is considered. The statistical significance of 1%, 5% and 10% is indicated by *, ** and ***, respectively.

조선대학교
CHOSUN UNIVERSITY

	Herding_Retail	Herding_Institution	Herding_Foreign
Inter.	−25.8953*** (−178.961)	142.0393*** (332.390)	60.1607*** (191.250)
AICA	1.2666*** (84.200)	−0.5733*** (−12.284)	−0.5281*** (−16.037)
Herding_Retail		0.2147*** (62.358)	0.4603*** (193.407)
Herding_Institution	0.0225*** (62.358)		0.0220*** (27.891)
Herding_Foreigner	0.0969*** (193.407)	0.0441*** (27.891)	
PER	−0.0000*** (−3.468)	−0.0000 (−0.238)	0.0000*** (4.811)
lnSIZE	1.7216*** (359.783)	−5.9666*** (−415.566)	−1.8919*** (−172.482)
lnVOL	−0.6947*** (−166.677)	0.6061*** (46.339)	−0.7587*** (−82.436)
RET	−0.0218*** (−0.0218)	0.0087*** (1.800)	0.0605*** (−17.650)
STD	−31.1645*** (−68.766)	33.9496*** (24.202)	−66.9539*** (−67.767)
Year fixed effect	Yes	Yes	Yes
Industry fixed effect	Yes	Yes	Yes
adR^2	0.274	0.226	0.130
No.observations	801,443	801,443	801,443

조선대학교
CHOSUN UNIVERSITY

〈Table 3〉 Relationship between investor herding behavior and Bullish

Table 3 reports the impact on Bullish on the retail, foreign and institutional investor herding measure. This table presents parameter in Eq. (6). Herd Retail, Herd Fore and Herd Insti represent retail, foreign and institutional herding behavior measure, respectively. Bullish is an investment sentiment indicator that represents buyout opinion score.AICA is abnormal ICA. lnSIZE is the value taking the natural log on the market capitalization. PER is the P/E ratio. lnVOL is the value taking the natural log on the trading volume. STD is the standard deviation of daily returns the previous 3 months. The industry and the year effect is considered. The statistical significance of 1%, 5% and 10% is indicated by *, ** and ***, respectively.

조선대학교
CHOSUN UNIVERSITY

	Herding_Retail	Herding_Institution	Herding_Foreigner
Inter.	−25.8556***	142.0463***	60.1659***
	(−178.734)	(332.407)	(191.266)
Bullish	−0.1459***	−0.0683***	−0.0564***
	(−24.534)	(−3.716)	(−4.345)
AICA	1.2935***	−0.5603***	−0.5174***
	(85.784)	(−11.970)	(−15.667)
Herding_Retail		0.2143***	0.4600***
		(62.232)	(193.206)
Herding_Institution	0.0224***		0.0220***
	(62.232)		(27.872)
Herding_Foreigner	0.0967***	0.0441***	
	(193.206)	(27.872)	
PER	−0.0000***	−0.0000	0.0000***
	(−3.551)	(−0.250)	(4.794)
lnSIZE	1.7133***	−5.9698***	−1.8945***
	(359.926)	(−415.046)	(−172.451)
lnVOL	−0.6808***	0.6124***	−0.7536***
	(−161.883)	(46.432)	(−81.197)
RET	−0.0186***	0.0103**	0.0618***
	(−11.764)	(2.105)	(17.953)
STD	−30.9363***	34.0498***	−66.8367***
	(−68.291)	(24.261)	(−67.655)
Year fixed effect	Yes	Yes	Yes
Industry fixed effect	Yes	Yes	Yes
adR^2	0.274	0.226	0.130
No.observations	801,443	801,443	801,443

조선대학교
CHOSUN UNIVERSITY

〈Table 4〉 The effect of variables on the herding of investor before and after COVID-19

Table 4 reports the impact on Bullish on the retail, foreign and institutional investor herding measure before and after COVID-19. This table presents parameter in Eq. (7) by including a dummy variable equal to 1, if the value after March 22, 2020, when social distance is implemented in Korea, and 0 otherwise. Herd Retail, Herd Fore and Herd Insti represent retail, foreign and institutional herding behavior measure, respectively. Bullish is an investment sentiment indicator that represents buyout opinion score. AICA is abnormal ICA. lnSIZE is the value taking the natural log on the market capitalization. PER is the P/E ratio. lnVOL is the value taking the natural log on the trading volume. STD is the standard deviation of daily returns the previous 3 months. The industry and the year effect is considered. We only show the results of the main variables in the table. The statistical significance of 1%, 5% and 10% is indicated by *, ** and ***, respectively.

조선대학교
CHOSUN UNIVERSITY

	AICA	AICA X COVID	Bullish	Bullish X COVID
Panel A. dependant variable is retail investor herding				
Herding_Retail	1.2935***(85.784)		−0.1459***(−24.534)	
	1.1582***(54.929)	0.2578***(8.913)	−0.1252***(−20.125)	
	1.282***(85.014)		−0.0443***(−5.254)	−0.1764***(−14.552)
	1.1229***(52.964)	0.3122***(10.721)	−0.0347***(−4.090)	−0.1920***(−15.724)
Panel B. dependant variable is institutional investor herding				
Herding_Institution	−0.5603***(−11.970)		−0.0683***(−3.716)	
	−0.1984***(−3.039)	−0.7025***(−7.861)	−0.0979***(−5.092)	
	−0.5596***(−11.950)		−0.0481*(−1.847)	−0.0951**(−2.537)
	−0.2095***(−3.191)	−0.6852(−7.613)	−0.0692***(−2.642)	−0.0609(−1.613)
Panel C. dependant variable is foreign investor herding				
Herding_Foreigner	−0.05174***(−15.667)		−0.5604***(−4.345)	
	−0.6841***(−14.853)	0.3377***(5.355)	−0.0835***(−6.150)	
	−0.5019***(−15.189)		−0.2144***(−11.662)	0.2736***(10.349)
	−0.6366***(−13.747)	0.2637***(4.153)	−0.2062***(−11.157)	0.2604***(9.781)

조선대학교
CHOSUN UNIVERSITY

〈Table 5〉 Regression analysis by individual investor group

This table reports the result of regression analysis on the investment sentiment index obtained from the posts, Bullish and the investor's herding. Bullish, a variable obtained from the posts, uses observations during the trading day during the sample period from January 2018 to December 2021. The statistical significance of 1%, 5% and 10% is indicated by *, ** and ***, respectively. In all regression analysis, the industry fixed effect and the year fixed effect is considered

	Herding_Retail		
Trading volume of retail	LOW	MIDDLE	HIGH
Inter.	−12.8171***	−0.2351	−0.0387
	(−41.463)	(−1.331)	(−0.032)
Bullish	−0.1945***	−0.1352***	−0.0301***
	(−11.034)	(−18.664)	(−10.088)
AICA	3.1114***	0.5616***	0.0092
	(76.861)	(32.576)	(1.097)
Herding_Institution	0.1121***	0.035***	0.0057***
	(108.735)	(96.043)	(26.911)
Herding_Foreigner	0.1149***	0.1376***	0.0696***
	(91.410)	(264.995)	(225.153)
PER	−0.0000	0.000	−0.0000*
	(−1.334)	(0.725)	(−1.789)
lnSIZE	1.1998***	0.1801***	0.0615***
	(104.165)	(25.828)	(13.327)
lnVOL	−0.6917***	−0.2263***	−0.0613***
	(−60.688)	(−44.635)	(−24.812)
RET	0.0106**	−0.0202***	−0.0092***
	(2.062)	(−10.740)	(−11.961)
STD	−97.4623***	−13.368***	−3.2586***
	(−64.120)	(−36.639)	(−13.906)
Year fixed effect	Yes	Yes	Yes
Industry fixed effect	Yes	Yes	Yes
adR^2	0.196	0.307	0.212
No.observations	267,159	267,159	266,811

조선대학교
CHOSUN UNIVERSITY

V. Conclusion

Herding means imitating the behavior of others or following decisions. It is important to know which factors cause the herding. If we can confirm the effect of activity and opinion of social media which can play an important role in the market through the study, we will observe whether the herding affects the market. In particular, we conducted research on the data of Naver Financial stock discussion room, which has a very large activity in Korea, is working well. After COVID-19, the trading volume of individual investors increased. The number of posts posted by Naver Financial stock discussion room, which has a large activity of social media, also increased. So, as opinion sharing becomes active in social media, we want to see if opinion is formed, and this affects the market by creating a herding of individual investors.

In this study, we use Bullish measure presented in the paper of Antweiler and Frank (2004), as an investment sentiment index for the posts. We follow Lakonishok et al. (1992) herding measure to estimate daily trading herding index by each stock and each type of investor covering retail, institutional and foreign investor. We modify the model of Da et al.(2011; 2015) to compute AICA index. Therefore, in this paper we redefine ICA as a variable that indicated information creation and supply by utilizing the number of daily posts in Naver Financial stock discussion room where investors form their opinions directly by modifying the model of abnormal search volume index provided by Da et al..

Regression analysis is conducted to investigate AICA on the herding behavior of individual investors. We conduct Bullish is negatively related to herding when sentiment is not good. And the main variables have stronger impact in investor's herding behavior after COVID-19. Finally the individual investors' transaction volumes are divided into three groups and regression analysis is conducted for each group.

The results of regression analysis are as follows.

AICA variable has statistically significant positive relationship with retail investor herding. The AICA variable, which means information creation activity of social media, is created by retail investors, so it is assumed that it has a positive correlation. In other words, it can be confirmed that the information creation supply of social media leads to actual transactions. Also, we could observe that AICA have a negative effect on institutional and foreign investors' herding, because they are not the main subjects

조선대학교
CHOSUN UNIVERSITY

of writing in social media. We discover the Bullish is negatively related to herding when sentiment is not good. The Bullish variable has statistically significant negative relationship with the three investor types of herding. That is, when sentiment and bullish is not good, all three types of investor can confirm their herding behavior. Regression analysis is conducted to check whether the main variables have stronger impact in investor's herding behavior after COVID-19. The impact of the Bullish variable on retail investor herding has become stronger after COVID-19. The effect of the bullish variable on retail investor herding is stronger after COVID-19. In other words, this indicates that individual investors are herding when they sell and do not herding when they buy. Foreigner herding shows a pattern of opposition before and after COVID-19, which should be confirmed through further verification. Finally, the individual investors' transaction volumes are divided into three groups and regression analysis is conducted for each group. In case of bullish, it is confirmed that there is a statistically significant negative relationship regardless of the trading activity of individual investors. AICA shows that there is no correlation at all in high group with large trading volume of individual investors. In other words, it is confirmed that it did not affect the herding at all. The reason for this result is that the herding is to be explained, and this result seems to have come out because the value of the herding itself is small. In addition, most of the transactions consist of individual investors, which can be interpreted as having a noise characteristic, so the herding seems to have decreased and these results are likely to come out.

The limitations of this study are as follows. In this paper, we examine the herding by investor without distinguishing between herding's buy-side and sell-side. The analysis of herding further shows that the effects of Bullish and Aica variables on buying and selling herding can be analyzed more effectively. We regard that the purchase quantity and the selling quantity are all formed by other investors.
But in reality, an investor can do a lot of transactions. Therefore, if we use orderbook data, we can solve these problems and more accurately measure herding. We can further segment the effects on the main variables used in the analysis.

조선대학교
CHOSUN UNIVERSITY

References

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.

Bank, M., Larch, M., & Peter, G. (2011). Google search volume and its influence on liquidity and returns of German stocks. *Financial markets and portfolio management*, *25*(3), 239-264.

Barber, B. M., & Odean, T. (2008). All that glitters: The effect of attention and news on the buying behavior of individual and institutional investors. *The review of financial studies*, *21*(2), 785-818.

Barber, B. M., Odean, T., & Zhu, N. (2009). Systematic noise. *Journal of Financial Markets*, *12*(4), 547-569.

Chang, E. C., Cheng, J. W., & Khorana, A. (2000). An examination of herd behavior in equity markets: An international perspective. *Journal of Banking & Finance*, *24*(10), 1651-1679.

Chiang, T. C., & Zheng, D. (2010). An empirical analysis of herd behavior in global stock markets. *Journal of Banking & Finance*, *34*(8), 1911-1921.

Choe, H., Kho, B. C., & Stulz, R. M. (1999). Do foreign investors destabilize stock markets? The Korean experience in 1997. *Journal of Financial economics*, *54*(2), 227-264.

Choi, K. H., & Yoon, S. M. (2020). Investor sentiment and herding behavior in the Korean stock market. *International Journal of Financial Studies*, *8*(2), 34.

Choi, N., & Skiba, H. (2015). Institutional herding in international markets. *Journal of Banking & Finance*, *55*, 246-259.

Chong, T. T. L., Liu, X., & Zhu, C. (2017). What explains herd behavior in the Chinese stock market?. *Journal of Behavioral Finance*, *18*(4), 448-456.

Christie, W. G., & Huang, R. D. (1995). Following the pied piper: Do individual returns herd around the market?. *Financial Analysts Journal*, *51*(4), 31-37.

Da, Z., Engelberg, J.,& Gao, P. (2011). In search of attention. *The journal of finance*, *66*(5), 1461-1499.

Da, Z., Engelberg, J., & Gao, P. (2015). The sum of all FEARS investor sentiment and asset prices. *The Review of Financial Studies*, *28*(1), 1-32.

De Long, J. B., Shleifer, A., Summers, L. H., & Waldmann, R. J. (1990). Positive feedback investment strategies and destabilizing rational speculation. *the Journal*

조선대학교
CHOSUN UNIVERSITY

of Finance, *45*(2), 379–395.

Demirer, R., & Kutan, A. M. (2006). Does herding behavior exist in Chinese stock markets?. *Journal of international Financial markets, institutions and money*, *16*(2), 123–142.

Diamond, D. W., & Dybvig, P. H. (1983). Bank runs, deposit insurance, and liquidity. *Journal of political economy*, *91*(3), 401–419.

Henry, E., & Leone, A. J. (2016). Measuring qualitative information in capital markets research: Comparison of alternative methodologies to measure disclosure tone. *The Accounting Review*, *91*(1), 153–178.

Hsieh, S. F., Chan, C. Y., & Wang, M. C. (2020). Retail investor attention and herding behavior. *Journal of Empirical Finance*, *59*, 109–132.

Gervais, S., Kaniel, R., & Mingelgrin, D. H. (2001). The high-volume return premium. *The Journal of Finance*, *56*(3), 877–919

Graham, J. R. (1999). Herding among investment newsletters: Theory and evidence. *The Journal of Finance*, *54*(1), 237–268.

Kahneman, D. (1979). Prospect theory: An analysis of decisions under risk. *Econometrica*, *47*, 278.

Ke, Z. T., Kelly, B. T., & Xiu, D. (2019). *Predicting returns with text data* (No. w26186). National Bureau of Economic Research.

Kim, W., & Wei, S. J. (2002). Foreign portfolio investors before and during a crisis. *Journal of international economics*, *56*(1), 77–96.

Kothari, S. P., Li, X., & Short, J. E. (2009). The effect of disclosures by management, analysts, and business press on cost of capital, return volatility, and analyst forecasts: A study using content analysis. *The Accounting Review*, *84*(5), 1639–1670.

Kraus, M., & Feuerriegel, S. (2017). Decision support from financial disclosures with deep neural networks and transfer learning. *Decision Support Systems*, *104*, 38–48.

Lakonishok, J., Shleifer, A., & Vishny, R. W. (1992). The impact of institutional trading on stock prices. *Journal of financial economics*, *32*(1), 23–43.

Lao, P., & Singh, H. (2011). Herding behaviour in the Chinese and Indian stock markets. *Journal of Asian economics*, *22*(6), 495–506.

Li, F. (2010). Textual analysis of corporate disclosures: A survey of the literature. *Journal of accounting literature*, *29*(1), 143–165.

Li, W., Rhee, G., & Wang, S. S. (2017). Differences in herding: Individual vs.

조선대학교
CHOSUN UNIVERSITY

institutional investors. *Pacific-Basin Finance Journal*, *45*, 174-185.

Loughran, T., & McDonald, B. (2011). When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks. *The Journal of finance*, *66*(1), 35-65.

Scharfstein, D. S., & Stein, J. C. (1990). Herd behavior and investment. *The American economic review*, 465-479.

Sias, R. W. (2004). Institutional herding. *The Review of Financial Studies*, *17*(1), 165-206.

Spyrou, S. (2013). Herding in financial markets: a review of the literature. *Review of Behavioral Finance*.

Tetlock, P. C. (2007). Giving content to investor sentiment: The role of media in the stock market. *The Journal of finance*, *62*(3), 1139-1168.

Trueman, B. (1994). Analyst forecasts and herding behavior. *The review of financial studies*, *7*(1), 97-124.

Shiller, R. J., Kon-Ya, F., & Tsutsui, Y. (1996). Why did the Nikkei crash? Expanding the scope of expectations data collection. *The Review of Economics and Statistics*, 156-164.

Shyu, J., & Sun, H. M. (2010). Do institutional investors herd in emerging markets? Evidence from the Taiwan stock market. *Asian Journal of Finance & Accounting*, *2*(2), 1.

Wermers, R. (1999). Mutual fund herding and the impact on stock prices. *the Journal of Finance*, *54*(2), 581-622.