ANALYZING CONTENTS OF MARKET SENTIMENT BASED ON INVESTERS' EMOTION

Sanggi Lee [a] and Joonhyuk Song [b,*]

Abstract. The study investigates the stock market using emotion index calculated from SMD based on investors' emotion. In the VAR anlaysis, we find that the correlation between the KOSPI200 return and emotion score sum is highest in 2- or 3- day lag. This study concludes that explanatory power of the SMD emotion index is limited in explaining the Korean stock market yet.

1. Introduction

Fama (1969) argued that the securities market is highly competitive and the newly recognized information is reflected very quickly in the market price, so it is difficult to obtain excess profits based on the Efficient Market theory. However, Shiller (2003) explained the unreasonable price movement in the stock market by recognizing market instability based on the Behavioral Finance arguments.

Following the lead of Shiller, many researchers have analyzed the anomalies and irrationality of investors to quantify the stock market psychology. In the literature, stock prices and irrationality in the market were found to be caused by the psychology of the crowd, stock price bubble, and price-tracking transactions. In this study, we examine the influence of market psychology on the stock market by adopting the SMD (Social Media Data) emotion index as a substitute variable of the abnormal movements of the stock price due to the investor's mood, psychology, and emotion.

As the Internet technology and infrastructure of the 1990s have dramatically developed, a great number of studies have reported that not only market trading information but also public opinion on SMD investment in virtual space can be affecting stock price volatility. As a result of quantifying informal public opinion by

using SMD emotion index using big data analysis technique, it is regarded that the basis for market psychology research has been established.

In the case of showing the significance of investment performance using inefficiency of market and asymmetry of information, the KOSCOM HINT system[1] analyzes the yields of 2,153 stocks by trader for 1 year from January to December 2016. The percentage of the theme-related[2] transaction stocks (1,070) in the total investment stocks (2,153) account for 54.9%. The average return on the theme stocks was 1.86%, and the average return on stocks without themes was lower than the average return on the stocks with -2.81%. This is interpreted as a meaningful return difference reflecting market psychology formed through the formation process of virtual space such as portal, private investment club, SNS, blog, etc.

We analyze the total transaction history by retention period and find that the longer the retention period is, the lower the overall rate of return, similar to previous studies. This is known as disposition effect which describes the investors' tendency to sell stocks whose prices rise and hold stocks whose prices decrease. The spread of information and the speed of transmission due to the development of the internet is accelerating its transaction, so it is estimated to be the excess return using liquidity inflow according to the information acquisition speed of the noise trader.

In this study, we aim to investigate the emotions of investors based on the SMD emotional sentiment, which is a substitute variable of market psychology and analyze the effect on stock market based on the relation between SMD and stock return.

2. Preceding Literature

Tetlock (2007) showed that the DJIA (Dow Jones Industrial Average) index temporarily fell, but generally recovered from the fall within a week at the negative evaluation from the WSJ (Wall Street Journal) news articles based on the correlation analysis between the DJIA and the WSJ news columns. He claimed that the effect is lost in a relatively short period of time. Barber and Odean (2008) argued that if stock issuers are mentioned in Dow Jones Newswires, then individual investors increase buy orders and institutional investors shorts their stocks, resulting

---

[1]It is an investment simulation system operated by KOSCOM. It supports profitable competition for 6 months and 1 year with the same amount of investment by investors in securities broadcasting for professional investors.

[2]It is grouped according to the theme information formed by the portal and securities investment community such as Naver, Daum etc., and classified into groups with specific themes classified such as presidential theme, new technology theme, and bio theme and groups without any specific themes

in the increase in trading volume in the next day. This is because, unlike institutional investors, the main information acquisition path of individual investors is influenced by SMD data such as news media. Hong et al. (2004) showed that the impact of sociability is more active in states with higher stock market participation rates than other states. This is a meaningful demonstration that social activity and stock market are related to each other.

There are also studies showing the direct relationship between the exposure in the cyberspace and the stock return. Mitchell et al. (1994) reported a direct relationship among Dow Jones & Company's daily news volume, trading volume, and market returns. Bollen et al. (2011) and Caligo et al. (2016) reported that public moods and sentiments like twitter will affect DJIA forecasts[3] and investors' views[4].

Among the domestic studies, Kim and Lee (2013) stated that the correlation between the stock price of a company and the number of exposure of a specific company on SNS showed a positive correlation, and Kim et al. (2014) found that users' messages from the stock market discussion room provided by PAXNET which is stock information provider, Naver and Daum are useful for stock price prediction, and also found that the richer the amount of SMD, the better the prediction of the stock price. Yun (2015), Jeong et al. (2015), and Moon et al. (2016) showed that it is possible to predict stock price movement directions and volatilities based on big data emotion analysis. Gam and Shin (2010) argued that in order to further improve the performance of existing value-added strategies, we need to consider market psychology and financial characteristics as additional factors.

3. Explanation of Data and Basic Statistics

3.1. Explanation of data The stock market trading data is the daily market price including high price, low price, close price, volume and daily yield data of the KOSPI during the 27-month period from January 2015 to March 2017 obtained from FN Guide. Market sentiment data using SMD is composed of KOSPI200 representative stocks consisting of about 200 representative stocks in the Korean market from January 2015 to March 2017. We use SMD emotion data provided by KOSCOM for

---

[3]In a survey on whether Twitter sentiment was correlated with DJIA in Twitter mood predictions of the stock market, Twitter sentiment showed 86.7 % accuracy in predicting DJIA stock price index in 3 days.

[4]Twitter, acquired in local news in the Philippines, showed that the closing index of the Philippine Stock Exchange (PSE) could be explained by the news.

the collected number of emotions, positive emotion score, negative emotion score, emotional score and emotional level.

In order to analyze the time series characteristics of SMD emotional data, macroeconomic variables and causal relations, data such as daily KOSPI200 index, industry classification by category and monthly economic index (leading, trailing, accompanying) were acquired and utilized from the FN Guide. In order to confirm the validity of the stock trading profit according to the theme of virtual space, HINT system transaction history of KOSCOM simulation investment system was secured from January, 2016 to December, 2016.

3.2. The computation method of SMD emotion index

3.2.1. *The analysis concept for SMD emotion index* The analysis flow of SMD emotion index can be found from the KOSCOM emotion index conception on panel (A) in Figure 1. From SNS data (Twitter), WEB data (internet cafes, blogs, intellectual and news portals, professional news channels, stock companies, research centers), the emotion analysis could be delivered by using stock emotion dictionary which is specially connected for stock industry covering atypical text data collections and emotion words from saved and refined shaped model, Emotion analysis is produced from the emotional scores starting from the wording level 1 to 5 along with positive·negative emotional score facts out of stock emotion dictionary. This emotional scores is computed from the calculation of positive negative emotional scores. With this calculation of emotional scores, the emotional level is executed from level 1 to 7 through emotional level execution model. The emotion data is based on the collection data from $(T - 1)$ 16:00 to (T) regular market 8:00. With this data, each markets emotional sum including KOSPI200, collected number of emotions, emotional score and emotional level could be computed.

3.2.2. *The collection site and keyword for SMD* The keywords are gathered from the collection site after sorting and collecting based on the standard criteria for each sectors collection keyword. Overall, the keyword are gathered using 1,500 collection keyword criteria out of KOSPI200, and collection sites such as SNS data(Twitter), WEB data(cafes, blogs, knowledgeiN and news portals, professional news channels, stock companies, research centers) and other related financial sites.

3.2.3. *Emotion analysis* The method of emotion analysis is processed as panel (B) in Figure 1. Like SMD emotion analysis flow, SMD is sorted out through refined

shaped analysis model out of collected sentences, then major patterns and words are selected. The emotion scores are allocated to these selected patterns and words after mapping stock section dictionary patterns. The emotion analysis patterns are getting emotional scores based on 5 levels of strong negative (-2), negative (-1), even (0), positive (1), and strong positive (2). The dictionaries specializing for stock industry to analyze emotion are organized by stock emotion dictionarys 5.000 sections, filtering dictionarys 4,700 sections and ending words dictionarys 1,800 sections. The emotion dictionary is organized by sections and patterns with emotional scores with 5 levels of positive·negative factors.

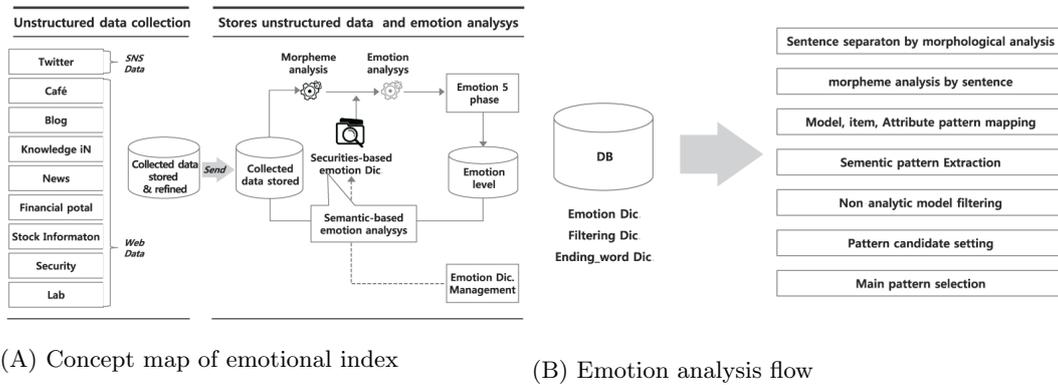(A) Concept map of emotional index

(B) Emotion analysis flow

Figure 1. KOSCOM's SMD emotion analysis

SMD emotion scores are computed after calculating each sectors emotion scores of positive and negative from the emotion level 1 to 5 out of stock emotion dictionary based on the collected patterns and words. Also, the emotion score sum of KOSPI are computed as well. If the emotion scores are not able to be computed, then the scores are left as an empty blank. Each sectors emotion scores are revalued after weekends and holidays. To reduce over collecting situation during those periods of weekends and holidays compared to normal working days, the weighted correlation factors which the collection data from regular market time (Tuesday, Wendesday, Thursday and Friday) divided by the collection data from the weekends and holidays (from Friday 16:00 before Monday 08:00) within 30 days are applied. During normal working days, the emotion scores are not revalued.

$$\text{Emotion score}_{it} = (\text{Positive score}_{it} + \text{Negative score}_{it}) \times \text{Weight}_{it}$$

$$\text{Weight}_{it} = \begin{cases} \frac{\sum_{j=1}^{-30} \text{Count weekday}_{ijt}}{\sum_{j=1}^{-30} \text{Count weekend}_{ijt}}, & \text{Weekday after weekend \& holiday} \\ 1, & \text{Weekday} \end{cases}$$

The emotion level is decided by SMD emotion level which is represented on the scale of 1 to 7 showing the degree of T days emotion scores compared to its previous trend. The emotion level 1 means the lowest grade and 7 means the highest grade.

After the revaluation based on the weighted factors between weekends and normal working days as mentioned above, the computation of emotional level is sorted as level 4 (even) if there is no emotion collected factors within 30 days. Otherwise, the T days emotion level is computed by the grade between 1 and 7 showing the depth indication if its scores are not exactly same compare to its previous trend. Moreover, if the emotion level is same as its previous trend, it is sorted as upper whisker(mean value $+1.5\times$IQR $(3/4 - 1/4)$) and lower whisker (Mean value $-1.5\times$IQR) to eliminate outlier.

$$\text{Emotion level}_{it} = \text{Quantile(Emotion score}_{it}, 7)$$

$$\text{Emotion level}_{it} = \begin{cases} 7, & \text{if Emotion score} > \text{Q3} +1.5 \times \text{IQR} \\ 1, & \text{if Emotion score} < \text{Q1} -1.5 \times \text{IQR} \end{cases}$$

The computation of emotional level is revalued to level 7 if the result showing above the highest and level 1 showing below the lowest after average arrangement by applying weekends/working days weighted factor and eliminating outlier based on the standard criteria of emotion scores.

3.3. SMD data basic statistics

3.3.1. *Basic statistics and time series haracteristics* To examine the stationarity of the data, we conduct an Augmented Dickey-Fuller test and the results are shown in Table 1. The collected number of emotions (No. of emotions), the positive emotion score (Positive score), the negative emotional score (Negative score), the emotional score (Emotion score), and the emotional level (Emotion level) are all stationary.

Clarify all the abbreviations if you are to submit this as your Ph.D. dissertation. PLEASE write the narration in a way that the readers can understand... If you do not clean up the below, there is not much I can help you. I am taking the current situation with a great concern. You also need to rewrite the Korean abstract to meet the 50 words limit.

Also, both the KOSPI daily emotion score representative index (KSE Esd) and the KOSPI daily emotion level representative index (KSE Eld) are also stationary.

Emotion data	Test Stat.	1% C.V.	5% C.V.
No. of emotions	−17.22	−3.43	−2.86
Positive score	−17.2	−3.43	−2.86
Negative score	−18.3	−3.43	−2.86
Emotion score	−17.63	−3.43	−2.86
Emotion level	−18.48	−3.43	−2.86
KSE Esd	−12.7	−3.43	−2.86
KSE Eld	−9.37	−3.43	−2.86

Table 1. Augmented Dickey-Fuller test

The basic statistics of SMD emotion data can be characterized as follows. First, as shown at the table 2, the emotion scores are higher than 0, indicating more positive emotion scores than negative emotion scores. Second, as the table 3 by market sectors, The collected number of emotions (No. of emotions) showed the same trend as the emotion score. The emotion level showed a value of around 4 points regardless of the emotion score by market sectors and in total. Hence, the following analysis was investigated mainly based on the emotion score.

Class.	Observation	Average	Std.dev.	Min	Max
No. of emotions	53,141	141.48	894.74	0	28,561
Positive emotion score	53,141	104.45	665.40	0	19,241
Negative emotion score	53,141	−39.80	267.10	−13,373	0
Emotion score	53,141	64.65	435.59	−1,187	13,187
Emotion level	53,141	3.92	2.031	1	7

Table 2. Basic statistics of emotion data in total

Class.	Observed cases	Average	S.Dev.	Min	Max
No. of emotions	56,943	76.9	327.08	0	58,408
Positive score	56,943	56.21	232.89	0	40,807
Negative score	56,943	−23.3	125.95	−21,030	0
Emotion score	56,943	32.9	143.31	−4,933	19,777
Emotion level	56,943	3.89	1.99	1	7

Table 3. Basic statistics of emotion data by market

Among the average value of the collected number of emotions, the emotion score and the emotion level based on the primary industry classification, as shown at the Table 5, the manufacturing industry has the largest number of 33,638 emotions collected, and the education service industry has the least number of 339 emotions

collected. There is no industry showing negative sentiment on the basis of the emotion score. Wholesale and retail industry are the highest at 74.38, and transportation is the lowest at 7.11. The differences in the collected number of emotions and emotion scores among the industries are 10 times and up to max 30 times indicating meaningful significance. The collected number of emotions and emotion scores are higher for those industrial groups that received the attention of the market.

ID	Industry class.	No. of emotions	E.Score	E.Level
H	Transportation industry	1,170	7.11	3.93
D	Electricity, gas, steam, water industry	443	8.79	4.06
N	Project facilities management and project support service industry	349	11.31	3.90
P	Education service industry	339	12.07	3.97
R	Art, sport and leisure-related service industry	509	12.62	4.00
K	Financial & insurance industry	3,079	15.77	3.92
C	Manufacturing industry	33,638	29.80	3.91
J	Publication, image, broadcast communication, and information service industry	2,092	30.42	3.91
F	Construction industry	1,621	37.87	3.96
M	Specialized and science & technology industry	5,165	43.54	3.93
S	Associations & organizations, repair and other individual service industry	377	59.79	3.80
G	Retail & wholesale industry	3,592	74.38	3.93

Table 5. Basic statistics of emotion data by industry

4. Research Methods and Models

4.1. SMD daily emotional representative Index Model

4.1.1. *SMD daily emotional representative index calculation method* Daily representative emotion index is calculated based on the emotion score and the emotion level after weighting them with market capitalization.The market capitalization weighted value of the daily representative emotion index is calculated after dividing the daily market value of stock by the sum of daily market value of all stocks and summed up to yield the representative emotion score and level.

$$Esd_t = \sum_{i=1}^{n} \frac{\mathrm{Sv}_{jt}}{\sum_{j=1}^{n} \mathrm{Sv}_{jt}} \cdot \text{Emotion score}_{it}$$

$$Eld_t = \sum_{i=1}^{n} \frac{\text{Sv}_{jt}}{\sum_{j=1}^{n} \text{Sv}_{jt}} \cdot \text{Emotion level}_{it}$$

Esd : Daily representative emotion score index for items

Eld : Daily representative emotion level index for items

Sv : Market value of daily items

n : Number of items with daily emotion index

i, j : Item with daily emotion index

4.2. SMD Emotion Index VAR Analysis Tetlock (2007) analyzed the interaction between the strength of negative sentiments of the Wall Street Journal(WSJ) column and the Dow Jones Return. Tetlock applied the method of VAR to count the emotion words of WSJ column according to emotional classification criteria[5]. And he also analyzed the influence of Dow Jones Return on the negative and weak categories, which were the pessimism media factors with the highest volatility and explanatory power. The results explained higher negative sentiment leads to a decline in market prices: however, it restores to the previous price within a week. In addition to that, high or low negative emotion predicts high trading volume. Finally, low market returns lead to high negative sentiment.

Based on the Tetlock methodology, this study examines whether KOSPI emotion score sum, which is the difference between market's positive and negative sentiment, is predictive of the KOSPI200 return and trading volume. Like Tetlock, the trend is controlled or eliminated by adding exogenous variables and dummy variables. The KOSPI emotion score sum is used instead of daily representative emotion score index, similar to simple counting method applied in Tetlock.

4.2.1. *Data composition* From January 2015 to March 2017, we applied KOSPI200 daily return, trading volume[6] and KOSPI emotional score sum[7] as substitutional factors of emotion index.

All variables use the lag value up to the last 5 days. The dummy variables are

---

[5]The General Inquirer's Harvard IV-4 criteria categorized for the psychological dictionary

[6]Tetlock used the 60-day moving average after converting the trading volume to the natural log. But, in this study, we did not use it because the VAR coefficient of trading volume is not meaningful. The VAR coefficient of the KOSPI200 daily returns and the market-based KOSPI emotion score sum is slightly increased when we use the 60-day moving average after converting the trading volume to the natural log. Trading volume scaled down by 10,0000,000.

[7]Emotion score is scaled down by 100,000.

included to control the January effect and the Monday effect emanating from the excessive collection of data on Monday. In addition to that, the exogenous variable is used as a control variable ($Exog$) for the KOSPI200 historical volatility which is calculated as Residual^2 value of the KOSPI200 return deducted by the moving average value of past 60 days.

$$AvgK200R_t = \frac{\sum_{t=1}^{n} K200R_t}{n}$$
$$\epsilon_t = K200R_t - AvgK200R_t$$
$$Exog_{t-1} = \epsilon_t^2 - \frac{\sum_{t=0}^{60} \epsilon_t^2}{60}$$

$AvgK200R_t$: Average of KOSPI200 return from t to past 60 days

$K200R_t$: Return of KOSPI200

4.2.2. *VAR model* We construct a tri-variate VAR (Vector Autoregressive) model. First, we construct three dependent variables models with the KOSPI emotion score sum, trading volume and KOSPI200 returns. The independent variables of each model are the lag values[8] of the past 5 days of all dependent variables, dummy variable eliminating January effect and KOSPI emotional score sum after holiday and lagged volatility as an exogenous variable.

$$
\begin{aligned}
KOSPI200_t = {} & \alpha_1 + \beta_1(L) \cdot L5(KOSPI200_t) + \gamma_1(L) \cdot L5(Essd_t) \\
& + \delta_1(L) \cdot L5(Vlm_t) + \lambda_{11} \cdot Exog_{t-1} + \lambda_{12} \cdot Dummy_{1Month,Monday} \\
& + \varepsilon_{1t}
\end{aligned}
$$
$$
\begin{aligned}
Essm_t = {} & \alpha_2 + \beta_2(L) \cdot L5(KOSPI200_t) + \gamma_2(L) \cdot L5(Essd_t) + \delta_2(L) \cdot L5(Vlm_t) \\
& + \lambda_{21} \cdot Exog_{t-1} + \lambda_{22} \cdot Dummy_{1Month,Monday} + \varepsilon_{2t}
\end{aligned}
$$
$$
\begin{aligned}
Vlm_t = {} & \alpha_3 + \beta_3(L) \cdot L5(KOSPI200_t) + \gamma_3(L) \cdot L5(Essd_t) + \delta_3(L) \cdot L5(Vlm_t) \\
& + \lambda_{31} \cdot Exog_{t-1} + \lambda_{32} \cdot Dummy_{1Month,Monday} + \varepsilon_{3t}
\end{aligned}
$$

$KOSPI200_t$: Rate of return on daily KOSPI200

$Essd_t$: Emotional score daily sum for items,

$$\text{i.e.} \sum_{i=1}^{n} \text{Emotion score}_{it}$$

Vlm_t : Average of daily trading volume

---

[8] The lag orders of the VAR analysis are determined based on the AIC.

$$Exog_{t-1} \;:\; \text{Moving average of the past 60 days of KOSPI200 return}$$
$$\text{Residual}^2$$
$$Dummy_{1Month,Monday} \;:\; \text{Removal of January effect and emotion tendency}$$
$$\text{of Monday}$$
$$\alpha \;:\; \text{Constant term}$$
$$\beta, \gamma, \delta, \lambda \;:\; \text{Factor sensitivity}$$
$$\varepsilon_{it} \;:\; \text{Error term}$$

5. Empirical Results

5.1. Empirical analysis of SMD daily emotion index

5.1.1. *SMD daily emotional representative index analysis* Daily representative index by the emotion score could not be found meaningful relationship with KOSPI200 daily returns. Examining Figure 2, the KOSPI monthly emotion index moved in tandem with KOSPI200 until the 2015, but this comovement seemed to fade and the gap became wider after then.
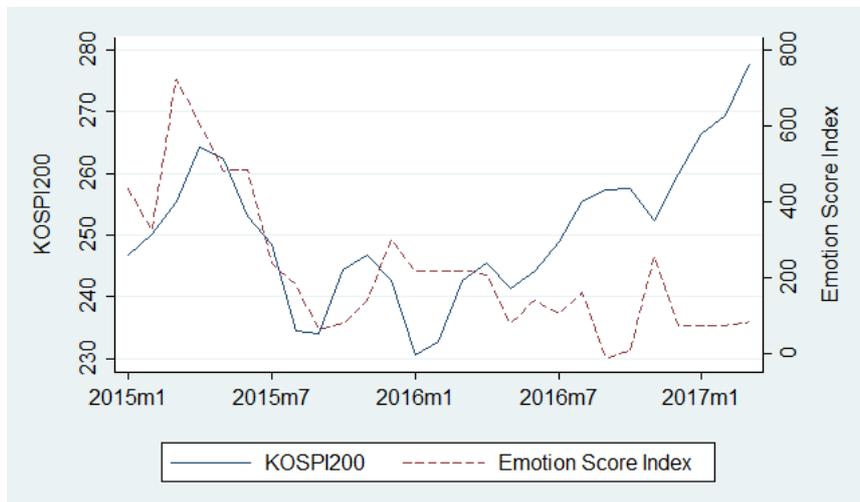

Figure 2. Emotion score index & KOSPI200

5.2. SMD emotion index VAR analysis Based on the Tetlock (2007) VAR analysis methodology, we summarize our results as follow. The correlation between

the KOSPI emotion score sum and the KOSPI200 returns is affected by the past lag negatively. The negative correlation for the lagged variables between trading volume and the KOSPI200 returns are also found. However, the KOSPI emotion score sum and transaction volume do not show any consistent relationship due to the opposite signs in the VAR coefficients. The VAR estimates are presented in Table 7.

The coefficient of L5 daily return in the KOSPI200 equation shows statistical significance at a standard confidence level. The KOSPI emotion score sum (Essd) showed a statistical significance for L2, and emotion indexes of the past two days and KOSPI200 day return moved to the opposite direction. This implies that the KOSPI emotion score sum is meaningful in explaining KOSPI200 daily returns.

L3 coefficient of KOSPI200 daily return in the Essm equation indicates a statistical significance, which implies that the KOSPI emotion score sum moves negatively against the past 3 KOSPI200 daily return. The trading volume has a negative effect on KOSPI200 daily return at L5, while the trading volume shows the positive relationship with the KOSPI200 emotion score sum from the Vlm equation. From the VAR analysis, we conclude that KOSPI daily emotion sum shows a statistically significant explanatory power in explaining daily KOSPI200 returns.

Independent var.	Lag	Dependent var.		
		KOSPI200	Essd	Vlm
5*KOSPI200	L1	−0.011	0.000	−0.006
	L2	−0.008	0.000	0.014
	L3	−0.054	−0.003***	−0.001
	L4	−0.018	−0.001	0.006
	L5	−0.111***	−0.001	−0.018*
5*Essd	L1	1.155	0.335***	−0.215
	L2	−3.140*	0.174***	−0.101
	L3	1.593	0.115**	−0.238
	L4	1.157	0.077*	1.153***
	L5	0.185	0.096**	−0.275
5*Vlm	L1	0.202	0.003	0.507***
	L2	−0.448**	−0.004	0.125***
	L3	0.141	0.001	0.019
	L4	0.011	0.005	0.063
	L5	−0.157	−0.010**	0.059
Exog	L1	−0.324	0.014*	−0.012
Dummy		Included	Included	Included
Constant		0.087	0.024***	0.112***

Note: *, **, and *** indicate statistical significance at 10%, 5%, 1%,respectively.

Table 7. VAR Estimates

5.2.1. *Forecasting error variance decomposition* To understand the shock propagation mechanism, we conduct a forecasting error variance decomposition and the results are shown in Table 9. For the case of KOSPI200 returns, Essm and Vlm shocks have no effect initially, but increase up to 1% after 10 days. For the case of Essm and Vlm, KOSPI200 returns shock explains 2.1% and 1.4% respectively after 10 days. Looking into these results, its own shock is the dominant factor in explaining forecasting error variance.

Impulse	KOSPI	Essd	Vlm	KOSPI	Essd	Vlm	KOSPI	Essd	Vlm
Response Time	KOSPI200			Essd			Vlm		
1	1.000	0.000	0.000	0.010	0.990	0.000	0.008	0.007	0.986
2	0.997	0.001	0.002	0.011	0.988	0.001	0.010	0.006	0.984
3	0.986	0.007	0.008	0.012	0.987	0.001	0.009	0.005	0.986
5	0.984	0.008	0.008	0.020	0.978	0.002	0.009	0.011	0.980
10	0.982	0.009	0.010	0.021	0.975	0.004	0.014	0.016	0.971

Table 9. Forecasting error variance decompostion

6. Conclusion

This study shows the explanatory power of the stock market using emotion index calculated by SMD based on investors' emotion. Moreover, this study offers the method of indexing SMD emotion data and its processing, and the characteristic and meanings of emotional indexes are presented.

In the empirical analysis, this study suggests evidence of emotion investment by confirming the relationship between psychology in virtual space and stock trading returns. Specifically, the return of stocks with various themes in virtual space is higher than stocks without themes. In addition to that, a new method of identifying the emotional trends and fluctuations of the market is presented by calculating representative sentiment score and level. Emotion indexes show differences by industry. The industry with the highest emotion score is the manufacturing industry and the lowest is the education service industry.

In the VAR analysis, the correlation between KOSPI emotion score sum and KOSPI200 returns is statistically significant negative for the lag in the past 2 or 3 days. The effect between the KOSPI200 returns and negative emotion scores applied by Tetlock does not show any statistical significance in the Korean stock market.

Although the SMD emotion index shows some potential in Korean stock markets, the explanatory power of the SMD emotion index in the overall stock market is limited. This could be explained by the relatively short period of data collected in the sample. This limitation could be relieved as we accumulate more data over time. We hope this study would provide a first step towards future research in discovering the relationship between investors' emotion and market returns.

ACKNOWLEDGMENT

REFERENCES

1. B.M. Barber & T. Odean: All That Glitters: The Effect of Attention and News on the Buying Behavior of Individual and Institutional Investors. Review of Financial Studies **21** (2008), no. 2, 785-818.
2. J.H. Bollen & X. Zeng: Twitter Mood Predicts the Stock Market. *Journal of Computer Science* **2** (2011), no. 1, 1-8.
3. A.R. Caligo, A.M. Sison, T. Bartolome & III Tanguilig: Prediction Model of the Stock Market Index Using Twitter Sentiment Analysis. International Journal of Information Technology and Computer Science **8** (2016), no. 10, 11-21.
4. E.F. Fama: Efficient Capital Markets: A Review of Theory and Empirical Work. Journal of Finance **25** (1969), no. 2, 28-30.
5. Hyeonggyu Gam & Yongjae Shin: The effect of market psychology and corporate financial characteristics on the performance of investment strategy. Industrial Economics Research **23** (2010), no. 1, 429-450.
6. H. Hong, K.D. Jeffrey & S.C. Jeremy: Social Interaction and Stock Market Participation. Journal of Finance **59** (2004), no. 1, 137-163.
7. Ji Seon Jeong, Dong Sung Kim & Jong Woo Kim: Influence analysis of Internet buzz to corporate performance: Individual stock price prediction using sentiment analysis of online news. Journal of Intelligence and Information Systems **21** (2015), no. 4, 37-51.
8. Taehwan Kim & Sangyong Lee: Does SNS Exposure Affect Stock Prices? Focusing on Twitter and Blogs in Korea. A Symposium Paper for The Korean Society for Management Information (2013), 491-499.
9. Young Min Kim, Jae Jeong & Suk Jun Lee: A Study on the Stock Market Prediction Based on Sentiment Analysis of Social Media. Entrue Hournal of Information Technology **13** (2014), no. 3, 59-69.

10. M.L. Mitchel & M.J. Harold: The Impact of Public Information on the Stock Market. Journal of Finance **49** (1994), no. 3, 923-950.

11. Juncheol Moon, Seongsu Kang & Junho Kim: A Study on the Performance and Investment Behavior Classified by the Type of Investors International Accounting Research **65** (2016) 155-178.

12. R.J. Shiller: From Efficient Markets Theory to Behavioral Finance. The Journal of Economic Perspectives **17** (2003), 83-104.

13. P.C. Tetlock: Giving Content to Investor Sentiment: The Role of Media in the Stock Market. Journal of Finance **62** (2007), no. 3, 1139-1168.

14. Hongwon Yun: Correlation Analysis of Social Sentiment and Stock Prices. Journal of the Korea Institute of Information and Communication Engineering **19** (2015), no. 7, 1593-1598.

[a]DEPARTMENT OF ECONOMICS, HANKUK UNIVERSITY OF FOREIGN STUDIES, SEOUL 02450 REPUBLIC OF KOREA
Email address: welsuc@naver.com

[b]DEPARTMENT OF ECONOMICS, HANKUK UNIVERSITY OF FOREIGN STUDIES, SEOUL 02450, REPUBLIC OF KOREA
Email address: jhsong@hufs.ac.kr