Investors Attention and the Effects on Stock market: An Empirical Study Based on Stock forum

Wen Long, Lijing Guan, Lingxiao Cui*
Research Center on Fictitious Economy & Data Science, Chinese Academy of Sciences
School of Economics & Management, University Chinese Academy of Sciences
Key Laboratory of Big Data Mining and Knowledge Management, Chinese Academy of Sciences
Beijing 100190, China
*Email: clxdsjyx@163.com

***Abstract*—This paper studies the relationship between China stock market performances and investors' attention, where the attention is measured by the page view of China's largest stock forum. The results show that trading volume, intraday volatility and liquidity have positive effects on investors' attention, while the effects of stock return to the attention is not significant. Finally, we demonstrate that investor attention is a significant alpha factor by using supervised learning algorithm.**

Keywords-Investors attention; trading volume; intraday volatility; liquidity

I. INTRODUCTION

Traditional capital asset pricing models assume the investors are rational and make their portfolio selection according to the efficient frontier. However, this requires that investors are in the state of information symmetry. Obviously, this is not a realistic assumption especially in Chinese stock market, where more than 80% investors are individual investors. Brad M. and Barber and T. O. (2008), Andrew et al (2009), Mondria (2010) indicate that investor attention is inclined to generate buying pressure for stocks as the reason of the difficulty that investors can't obtain all the information they need to make rational decision. The mechanism of how investors' limited attention could affect the stock market has been attracted lots of attention.

However, the measurement of investors' attention is difficult, and scholars use various kinds of data as proxies of investor attention. Chemmanur and Yan (2009) treat advertising expenditure as the proxy of investor attention, and they find that the more advertising expenditure, the more attention will be paid by investors. Peress (2008) uses company's report frequency by The Wall Street Journal as the proxy of investor attention, which indicates that a higher investor attention will lead to higher return and larger volume. Another type of proxies is constructed by stock market data. Using market volume as the proxy of investor attention, Gervais et al (2001) demonstrate that stocks with abnormal volume usually attract more investor attention, which may cause stock price rising in nearly a month. Seashole and Wu (2007) indicate the limit up of a stock can be a good proxy of investor attention in Chinese stock market, because when a stock reaches its limit up, investors would pay more attention to it. With the popularity of the Internet, the information on the internet becomes more and more representative. Nowadays, data from search engines and internet forums are widely used as the proxy of investor attention. Thomas and Stephan (2012) adopt the Google search data of the keyword "Dow" as a proxy of investor attention to stock market, and their research shows that investor attention is the Granger cause of realized volatility of the Dow Jones.

II. INDICES CONSTRUCTION

A. Investor attention indice

In Chinese stock market, the proportion of individual investors is considerably high. Stock forum is an important platform, where individual investors search and exchange information. In this work, our data are all from Eastmoney stock forum. Eastmoney is a financial portal founded in 2004, and its stock forum's visits and posts have reached a certain scale since the year of 2011. It has become the largest and the most influential financial portal according to several authoritative surveys. We extract the page view daily data from the posts on this stock forum, which can be a proxy of investor attention. In detail, we collect the time series of page view of 24 stocks, which are the constituent stocks of CSI 300 energy index on Eastmoney stock forum from November 1th, 2011 to September 30th, 2015. In order to capture the variation information of stocks with a low investor attention, we normalize each page view time series to get the attention indices. Then, the energy sector's investor attention can be constructed by summing up all the 24 stocks' "normal" attention indices. We also define the abnormal attention index:

$$AbnPV_t = PV_t - Med\left(PV_{t-1},...,PV_{t-14}\right) \quad (1)$$

Where, Med is the median of last two weeks' data. Abnormal attention index excludes the effects of increasing trends.

B. Market indices

To investigate the relation between energy sector attention and market performances, we consider five different

2375-9259/16 $31.00 © 2016 IEEE
DOI 10.1109/ICDMW.2016.112

market indices: energy sector log return R_t, turnover volume $Volume_t$, Intraday volatility $Volatility_t$, and market liquidity $LiquiInd_t$. The explicit definition of intraday volatility and liquidity are as follows:

$$Volatility_t = Ln\left(P_t^{high} / P_t^{low}\right) \quad (2)$$

$$LiquiInd_t = Volume_t / AbsR_t \quad (3)$$

Where $P^{High,Low}$ indicates the intraday highest and lowest price respectively, $AbsR_t$ is the absolute value of price return. The smaller influence of the large trading volume indicates the good condition in liquidity (Amihud, 2002).

III. EMPIRICAL ANALYSIS AND RESULTS DISCUSSION

A. The description for investor attention of energy sector

Fig.1 shows the changes of investor attention in energy sector from the beginning of 2012 to the beginning of 2015. It can be seen that investor attention has experienced a rapid growth from May to June in 2013, and the average value surges from about 100 points to more than 600 points. Meanwhile, the dropping speed is increasing for the market index during this period. The CSI 300 index has declined from 2300 points in May 31, 2013 to 1849 points in June 25, 2013. Also, with the advent of bull market during July to August in 2014, the investors' attention in energy sector increased to nearly 1000 points. We can see from Fig.1 that almost all the abnormal changes in stock market can cause the increasing of investor attention. However, whether the changes of investor attention can impact the performance or predict the behavior of stock market? In the following part, we will explore the interactive relationship between stock market and investor attention.

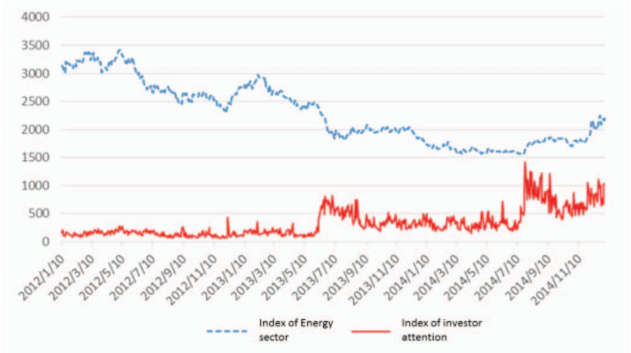

Figure.1 The investor attention and the index of energy sector

B. Stationary testing

In order to avoid the problem of false regression, we apply the ADF root test on investor attention, the abnormal attention index and five market indicators mentioned above. All the results rejected the existence of the unit roots in 1% significant level, which verified the stable of all the time series.

C. The correlation of investor attention and market indices

According to the traditional three factor model of Fama and French (1992), stock return can be explained by three factors which are market risk, company size and company Price-to-Book Ratio. We take energy sectors' return R_t as dependent variable and the company size factor SMB_t, market risk factor RP_t and company Price-to-Book Ratio factor HML_t as the control variables. We construct two multiple regression models by taking investor attention index and the abnormal attention index as endogenous variables separately.

$$R_t = \alpha + \beta_0 RP_t + \beta_1 SMB_t + \beta_2 HML_t + \beta_3 PV_t + \varepsilon_t \quad (4)$$

$$R_t = \alpha + \beta_0 RP_t + \beta_1 SMB_t + \beta_2 HML_t + \beta_3 AbnPV_t + \varepsilon_t \quad (5)$$

The regression results are as follows:

Table 1 The multiple regression results

Independent variables	Model 4	Model 5
Intercept	-0.00167*	-0.00141***
	(-1.951)	(-3.1917)
RP	1.1799***	1.1833***
	(30.9417)	(31.3612)
SMB	-0.1160	-0.1211
	(-1.5267)	(-1.6160)
HML	0.3106***	0.3132***
	(4.5256)	(4.6290)
AbnPV		7.8643e-06***
		(2.8492)
PV	2.295e-06	
	(1.1397)	

*, **and *** indicate significance at the10%, 5%, and 1% levels, respectively. T-statistics are in parentheses.

The results in Tab.1 shows that the return of energy sector is significantly correlated with market risk factor and company Price-to-Book Ratio factor, the abnormal investor attention is also correlated with energy sector's return at the 1% significant level. But, there is no obvious correlation between the normal investor attention and stock return. The results demonstrate that the abnormal investor attention at day t can bring a positive return. We have also examined whether the investor attention could affect the future return at day $t+1$, but the result isn't significant, which also reflects the sensitivity of the forum page view as the proxy variable of investor attention, that is to say the abnormal investor attention could be effectively reflected by stock return at that very day.

We further investigate the correlation of investor attention with trading volume, intraday volatility and market liquidity. We construct the following three models taking investor attention as independent variable and the absolute return of the market portfolio $AbsRP_t$ as control variables:

$$MarInd_t = \alpha + \beta_0 AbsRP_t + \beta_1 PV_{t-k} + \varepsilon_t \quad (6)$$

$$MarInd_t = \alpha + \beta_0 AbsRP_t + \beta_1 \frac{1}{n}\sum_{k=1}^{5} PV_{i-k} + \varepsilon_t \quad (7)$$

$$MatInd_t = \alpha + \beta_0 AbsRP_t + \beta_1 \frac{1}{n}\sum_{k=6}^{10} PV_{t-k} + \varepsilon_t \quad (8)$$

$MarInd_t$ denotes the trading volume $Volume_t$, intraday volatility $Volatility_t$ and market liquidity $LiquiInd_t$ separately. The first model examines how investor attention in $t-k$ day ($k=0,1...4$) impact the three market indices. And, the second model and the third model test whether the average investor attention in the past one-week and the past two weeks can influence the three market indices. The regression results in Tab.2 verify that investor attention in $t-k$ day ($k=0,1...4$) and in the past one-week can bring a significant positive impact on 1% significant level, and the coefficient is decreasing along with the increasing of the lags of period. The result also suits for the intraday volatility, that means the increasing investor attention can bring a stronger volatility in the following one or two week. The impact to liquidity also indicates that the higher of investor attention, the stronger of the sector liquidity, and the influence can last one to two weeks.

To investigate the interrelationship of investor attention with trading volume and intraday volatility, we use VAR model to test the interactive relationship. The data of the two models are $(PV, Volume)$ and $(PV, Volatility)$ separately. We adopt AIC rule to determine the lagged ranks, which are 2-order and 1-order. It is necessary to test the models' stability before the VAR is estimated. If the model isn't stable, the result will not be reliable. The results show that the characteristic polynomial roots of the above models are out of the unit circles, which means the results are stable. The estimated results of the two models are shown in Tables 3 and 4.

Table 2. The regression results of investor attention to the three market indices

Independent variables	Dependent variables		
	Volume	Intraday volatility	Liquidity
PV_t	0.00712***	7.148e-6***	4.5265***
	(20.72)	(4.3120)	(4.800)
PV_{t-1}	0.00673***	4.905e-6***	5.0719***
	(18.66)	(2.913)	(5.355)
PV_{t-2}	0.00638***	5.431e-6***	4.572***
	(16.62)	(3.152)	(4.6887)
PV_{t-3}	0.00630***	6.997e-6***	4.258***
	(16.00)	(4.023)	(4.2899)
PV_{t-4}	0.00613***	5.998e-6***	4.736***
	(14.83)	(3.344)	(4.6518)
$\frac{1}{n}\sum_{k=1}^{5} PV_{t-k}$	0.00799***	8.399e-6***	6.468***
	(17.18)	(3.998)	(5.452)
$\frac{1}{n}\sum_{k=6}^{10} PV_{t-k}$	0.00679***	8.406e-6***	6.035***
	(12.35)	(3.676)	(4.653)

*, **and *** indicate significance at the 10%, 5%, and 1% levels, respectively. T-statistics are in parentheses.

Table 3. The VAR test of investor attention and trading volume for energy sector

	Volume			PV		
	Coefficient	T-statistics	P-value	Coefficient	T-statistics	P-value
Intercept	6.83796e5***	3.28177	0.00114	39.2185***	2.6555	0.0083
PV_{t-1}	2267.33223***	2.73096	0.00668	0.6955***	11.8196	5.9020e-27
$Volume_{t-1}$	0.48809***	8.39018	1.625e-15	-1.141e-06	-0.2768	0.7820
PV_{t-2}	-791.72031	-0.93091	0.35260	0.1305**	2.1662	0.0310
$Volume_{t-2}$	0.23793***	4.08525	5.582e-05	7.395e-06*	1.7914	0.0741

*, **and *** indicate significance at the10%, 5%, and 1% levels, respectively.

Table 4. The VAR test of investor attention and intraday volatility for energy sector

	Volatility			PV		
	Coefficient	T-statistics	P-value	Coefficient	T-statistics	P-value
Intercept	0.8147***	14.3027	3.2141e-36	-363.9634	-0.5349	0.593
PV_{t-1}	9.06e-6***	3.2632	0.0012	0.8302***	25.0389	2.236e-77
$Volatility_{t-1}$	0.1963***	3.4898	0.0005	420.7044	0.62596	0.531785

*, **and *** indicate significance at the10%, 5%, and 1% levels, respectively.

We can see from the results that the autoregressive effects are significant for investor attention, trading volume and intraday volatility. Investor attention in $t-1$ day can obviously affect the trading volume in t day, but there is no significant impact for trading volume to investor attention. The similar phenomenon also has been tested in American stock market. Bordino et al (2012) use the Google search index as the proxy variable of investor attention, and find that investor attention has a positive effect on the following days' trading volume, but, on the contrary, the influence of trading volume to investor attention is not obvious. The results also show that the higher of the investor attention,

the stronger of the following days' volatility, but there is no significant effect for the investor attention to volatility, which is not consistent with the results of European market. Vozlyublennaia (2014) finds that the volatility in $t-1$ day has a positive effect on investor attention in t day in European market. Therefore, the investors in Chinese stock market are more irrational and risky so that they are not sensitive for the intraday volatility.

IV. Application of Investor Attention in Portfolio Selection

In previous sections, it has been shown that investor attention can affect future trading volume and intraday volatility, which indicates that the irrational investor behavior makes stock market more inefficient. However investor attention cannot predict the future stock return, only abnormal attention can affect the current stock return, which makes this index profitless. There is still promising that the investor attention influences stock return in a nonlinear way which cannot be captured by linear regression analysis. In this section, we use modern data-mining algorithm to investigate the profitability of this indices.

We use the investor attention indices and other market features to find the portfolio with a higher return. First we construct training data from the past market data and label the highest few stocks as a different class. Then supervised learning algorithm try to capture the pattern in the data set. If the algorithm can achieve a higher accuracy when considering the attention indices, it can show that this index is profitable. The market features we consider are opening price, highest price, lowest price and volume weighted average price. To make the range of different stock data be the same, each stock's data are scaled by its own closing price. We use weekly market data to construct the training data and the log return is the criteria for making the label. Considering that we only want to extract the pattern of stocks with higher return, it is reasonable to set the number of classification to be two: higher abnormal return and other. We then define different quantile (5%,10%,20%) to capture different level of abnormal return. When one stock's return is above the defined quantile, the label of this sample data can be set to "high return". The time range of training data is the same with the investor attention indices, which is from 2011 to 2015. When taking investor attention indices into account, we investigate two features of attention indices. One is the percentage of the attention, the other is the change rate of the attention indices, which we also use log return of attention indices instead. Random forest, SVM with RBF kernel and KNN are applied to learning the pattern in the Chinese stock market. In finance practice, people usually care more about the accuracy or success rate, therefore we don't include the recall rate or other performance evaluation index. The accuracy rates are all obtained from 5-fold cross validation.

Table 5. The accuracy of different models

		0.1	0.2	0.3
Market data	KNN	58.3%	59.1%	59.2%
	SVM (RBF)	56.6%	58.7%	57.9%
	SVM (Linear)	56.9%	58.7%	57.7%
	Logistic	47.9%	50.2%	55.3%
	Random Forest	57.3%	57.5%	55.4%
Market data + investor attention	KNN	60.3%	61.2%	59.2%
	SVM (RBF)	59.1%	59.3%	58.5%
	SVM (Linear)	59.3%	60.4%	59.8%
	Logistic	55.6%	57.2%	52.1%
	Random Forest	60.2%	62.1%	60.4%

From the results, it can be seen that with the investor attention indices, nearly all models have approved their accuracy. Nearly 60% accuracy shows that investor attention can affects stock market in certain pattern which can be tracked by data mining algorithm at least in the time range from 2011 to 2015. Together with other strategies, investor attention can be a promising alpha factor.

V. Conclusion and Discussion

The proportion of individual investors in China's stock market has been in a high level, especially after the start of the bull market in 2014. The number of the new account openings in Chinese A-share market is increasing in recent years and individual investors have become an important part to the development of Chinese market. With the popularity of Internet, Internet forum is the main channel for individual investors to vent their emotions and pick up information. Therefore, in this paper, we use the page view data in Easternmoney stock forum to construct the investor attention index of energy industry and test the interactive relationship between investor attention and stock market. The results show that there is a significant correlation between investor attention and stock market indices. Firstly, investor attention is positively correlated with stock return in energy sector; also, investor attention can affect trading volume, intraday volatility and liquidity significantly, and the effect can last one to two weeks. Further, we use VAR to test the interrelationship. We find that investor attention is not sensitive to trading volume and intraday volatility; the abnormal trading volume and intraday volatility also can't attract individual investors' attention effectively. The phenomenon is opposite to European market, but the conclusions reflect Chinese individual investors are irrational and risk-taking. Finally we test that investor attention can be a significant alpha factor by using supervised learning algorithm.

Acknowledgment

This research was supported by National Natural Science Foundation of China (No.71101146), China Postdoctoral

Science Foundation Funded Project (No.119103S265) and University of Chinese Academy of Sciences.

REFERENCE

[1] Amihud Y. Illiquidity and Stock Returns:Cross-Section and Time-series Effects[J]. Journal of Financial Markets,2002,5(1):31-56.

[2] Andrew B. Abel, Janice C. Eberly, Stavros Panageas. Optimal Inattention to the Stock Market with Information Costs and Transactions Costs[J]. Nber Working Paper Series. National Bureau of Economic Research. Cambridge, MA.2009.

[3] Brad M. Barber, T. O. All That G litters: The Effect of Attention and News on the Buying Behavior of Individual and Institutional Investors[J]. The Review of Financial Studies, 2008, 21(2): 785-818.

[4] Bordino, I., Battiston, S., Caldarelli, G., et al. Web Search Queries Can Predict Stock Market Volumes[J]. PLoS ONE, 2012,7(7).

[5] Chemmanur T J, Y. A. Advertising, attention, and stock returns[J]. Working paper. 2009.

[6] Fama, E. F., French, K. R. The Cross-Section of Expected Stock Returns[J]. The Journal of Finance, 1992, 47(2), 427-465.

[7] Gervais S, Kaniel R, Mingelgrin D H. The high‐volume return premium[J]. The Journal of Finance, 2001, 56(3): 877-919.

[8] Mondria, J. Portfolio choice, attention allocation, and price comovement[J]. Journal of Economic Theory, 2010, 145: 1837-1864.

[9] Peress J. Media coverage and investors' attention to earnings announcements[J]. New York, 2008, 33: 1-57.

[10] Seasholes M S, Wu G. Predictable behavior, profits, and attention[J]. Journal of Empirical Finance, 2007, 14(5): 590-610.

[11] Thomas Dimp, Stephan Jank. Can Internet search queries help to predict stock market volatility? [J]. Working paper. 2013.

[12] Vozlyublennaia, N. Investor attention, index performance, and return predictability[J]. Journal of Banking & Finance, 2014, 41:17-35.