Investor Sentiment Index and Option Price Volatility Based on MIDAS Model: Evidence from China

Haiyan Xiao[1]*, Yingxin* Hao[2]*, and Sirong* Wu[3*]

[1]School of Mathematics and Economics, Hubei University of Education, 430205 Wuhan, China
[2]School of Statistics and Mathematics, Zhongnan University of Economics and Law, 430073 Wuhan China
[3]Shenzhen Shunheng Rongfeng Supply Chain Technology Co., Ltd, China

Abstract. The paper selects the transaction data of the option market and network data from June 1, 2015 to February 2, 2018. The principal component analysis is adopted to construct investor sentiment index. The OLS and MIDAS model are employed to study the influence of investor sentiment index on the option implied volatility in Shanghai Stock Exchange 50ETF. The empirical results show that the MIDAS model with more high-frequency information has stronger interpretation ability than the same frequency model. Investor sentiment index based on traditional indicators has a negative effect on the option implied volatility while the excessive attention of investors on the Internet would exert positive pressure on the option market. The conclusion can well explain the inherent mechanism of investor sentiment affecting option implied volatility. Therefore, it is of great practical significance to study the influence of investor sentiment on the option price volatility in China

1 Introduction

Investor sentiment is a kind of emotion formed by investors in view of their perception of various information. It contains the systematic deviation of investors' expectation of the future so that it affects the sustainability of the security market. As an active representative in China's financial market, the implied volatility of the option market is formed by the trading behavior of the market, and represents the expectation of investors for the future price volatility. It plays a very important guiding role in reflecting the pricing and optimization of options, the allocation of option assets and other regulatory policies. Under such a circumstance, it is of great importance to study the interaction between investor sentiment and implied volatility, especially to the risk prevention and sustainable development of the security market.

Bing Han [1] studied the impact of investor sentiment changes on the S&P500 option price. Zhang and Lin [2] found that there was a positive nonlinear spillover effect between investor sentiment and stock market return rate. Xiong et al. [3] found that in the market

---

* Corresponding author: wusirong_sharon@163.com

environment with and without trading restrictions, investor sentiment had opposite effects on the market risk management ability of CSI 300 stock index futures, but different influence depth on the price discovery ability. It is found that there are few studies on the relationship between investor sentiment and implied volatility, and in the above empirical studies, all the empirical models adopted have a common feature: the data frequency of the explanatory variable and the explained variable in the model is the same, which cannot be satisfied in practice. The traditional processing method is to convert the high-frequency data to the low-frequency data by calculating the average value of the data, or to convert the low-frequency data to the high-frequency data by interpolation method. However, it can lead to the loss of information or data distortion, which may eventually result in the reduction of research credibility.

For a more comprehensive study of the impact of investor sentiment on option market volatility, there are three contributions to the paper. Firstly, unlike prevailing studies which employed transaction data related to the security market to construct investor sentiment index, we also introduced Baidu Search Index as a novel proxy sentiment index. Secondly, as for keywords extracting, which is different from previous research [4], the keywords were only related to several specific asset names. We collected information from posts and comments from China's large investment social media Oriental Fortune by using web scraping technology, then obtained the keywords by performing text analysis. This procedure could measure the investor sentiment in a more comprehensively way. Last but not least, the mixed data model MIDAS was used to analyze the impact of investor sentiment on the implied volatility of option. The paper constructed daily, weekly and monthly investor sentiment indexes, and the effects of mixed sentiment on the implied volatility of options can be studied on a more microscopic time scale.

2 Index system

2.1 Option implied volatility

Implied volatility refers to the volatility implied by option prices observed in the market. Fleming J et al. [5] studied the volatility index of the options derived from the S&P100 index, which has been proved that can better reflect investors' expectations of the option market. China iVIX index, issued by the Shanghai Stock Exchange, is used to measure the volatility expectations of the Shanghai 50ETF over the next 30 days. It was first announced on June 1,2015 and stopped publishing on February 13, 2018. Due to the clear negative correlation between China iVIX index and the market index, it can be referred as the "panic index" of investor [6]. Therefore, the paper employ the data of China iVIX index closing price to describe the implied volatility of SSE 50ETF option market.

2.2 Investor sentiment index

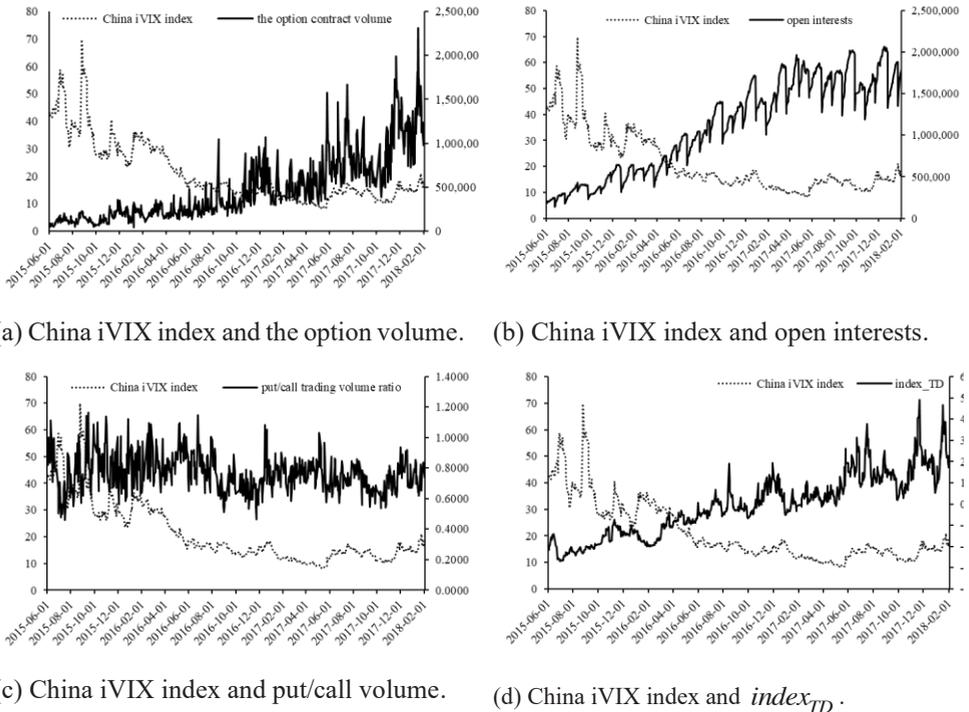#### 2.2.1 Construct investor sentiment index based on traditional index

Traditional methods usually employ indirect indicators which include the transaction data related to the stock market to construct investor sentiment index. Since the investment decisions of these stock market transaction data are made by investors based on their psychological characteristics, there always seems to be some inherent correlation that can indirectly reflect investor sentiment. Closed-end fund can be adopted as the investor sentiment index [7,8]. Compared with a single indicator, a composite sentiment index can quantify the sentiment of option investors in a more comprehensive and multi-angle

perspective. Therefore, some researchers adopted the composite indicators. Baker and Wurgler [9] defined sentiment based on the closed-end fund discount, NYSE share turnover, the number and average first-day returns on IPOs, the equity share in new issues and the dividend premium. Later, Kim et al. [10] also adopted this index.

In order to construct a composite index, three representative indicators are selected as the original variables in the Chinese options market: the option contract volume, the number of open interests and the put/call trading volume ratio. The principal component analysis method used by Baker and Wurgler [9] in constructing the sentiment will be adopted. The option contract volume, the number of open interests and the put/call trading volume ratio are marked as x_1, x_2, x_3. Standardize the data before using this method. In order to eliminate the influence of variance, we extract the principal components from the correlation coefficient matrix. The composite index is constructed by the first principal component which explains 73.69% of the (standardized) sample variance. Hence, we obtain a daily sentiment index based on traditional indicators. As in Eq. 1:

$$index_{TD} = 0.6073x_1 + 0.5911x_2 + 0.5313x_3. \tag{1}$$

We now use the Fig. 1 to illustrate the rationality of the investor sentiment composite index. The time-series diagrams of all variables during the sample period are shown in Fig. 1. We can conclude that there exists a highly correlation between the three variables and China iVIX index, which makes support to the construction of the traditional investor sentiment index. It is found that the traditional investor sentiment index is negative correlated with the option implied volatility from Fig. 1(d).

(a) China iVIX index and the option volume.   (b) China iVIX index and open interests.

(c) China iVIX index and put/call volume.   (d) China iVIX index and $index_{TD}$.

Fig. 1. Proxy variables, $index_{TD}$ and the China iVIX index.

2.2.2 Construct investor sentiment index based on Baidu index

The rapid development of Internet has provided new support for investor psychology and behavior researches. Investors are more willing to search information or express their opinions and attitudes on the online social platforms. Therefore, adopting the social media data to measure investor sentiment index can provide a better explanation for the impact of investors in the stock market.

Some researchers used microblogging forums tweets including Twitter and Weibo to show the relationship between public sentiment and the stock market. Kim et al. [11] collected text messages on the Yahoo! and found investor sentiment had predictive ability for stock returns. Wang et al. [12] was the first attempt to use Baidu Search Index for research. Therefore, using public opinion information and network information as the source of emotional information has become a common method, which has been verified that it can reflect emotions of the investors directly [4].

Scraped the text data of the postings and comments of Shanghai Stock Exchange 50ETF (short for SSE 50ETF) Option Bar from Oriental Fortune website, which include 2,326 comments from June 1, 2015 to February 2, 2018. The text information of the posts and comments crawled from the Oriental Fortune Website can extract the most frequently spoken words by shareholders, which are exactly the investors most concerned about and are most closely connected with the option market.

Keywords were selected based on the principle of frequency greater than 100 and included in Baidu Index.These 13 proxy variables including option, 50ETF, call option, market, trading, market trend, Shanghai Stock Exchange, index, buy, account opening, account, threshold and short are assigned labels as z_1, L, z_{13}. Same as constructing traditional investor sentiment index, adopting the principal component analysis method. The first three principal components f_1, f_2, f_3 can total explain 71.8% of the (standardized) sample variance, as in Eq. 2:

$$f_1 = 0.2913z_1 + 0.3673z_2 + L + 0.0008z_{13}$$
$$f_2 = -0.2278z_1 - 0.2643z_2 + L + 0.4355z_{13} \qquad (2)$$
$$f_3 = -0.0130z_1 + 0.0157z_2 + L - 0.2286z_{13}.$$

Thus, a new composite index can be constructed by the weighted sum of the first three principal components, where the weights are chosen as the variance contribution rate, as in Eq. 3:

$$index_{BD} = 44.67\%f_1 + 18.15\%f_2 + 8.98\%f_3. \qquad (3)$$

Fig. 2 shows a highly similar fluctuations and trends between the $index_{BD}$ and China iVIX index. Unlike the $index_{TD}$, the correlation between the $index_{BD}$ and China iVIX index is positive. Therefore, it is rational to say that the $index_{BD}$ represents an overall search enthusiasm on SSE 50ETF options.
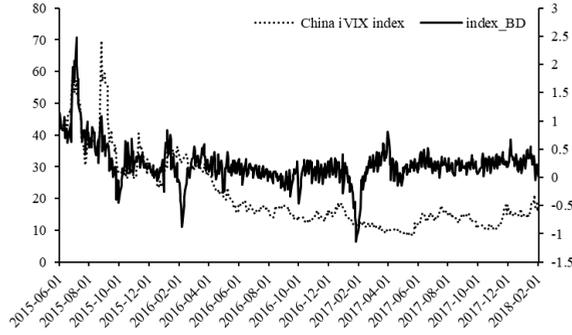
Fig. 2. China iVIX index and $index_{BD}$.

3 Methodology

Ghysels et al. [13] proposed the MIDAS model based on the distributed lag model, which can effectively study the actual reflection of high-frequency explanatory variables on low-frequency interpreted variables and avoid information loss or data distortion. The expression of the univariate MIDAS model is as Eq. 4:

$$y_t = \alpha + \beta W\left(L^{\frac{1}{m}};\theta\right)x_t^{(m)} + \varepsilon_t, \quad t = 1, \text{L}, T, \tag{4}$$

where m is the frequency multiple difference between the high-frequency data and the low-frequency data. $x_t^{(m)}$ means the m samples are taken from time $t-1$ to time t . Parameter β represents the overall influence of explanatory variables $x_t^{(m)}$ on y_t . ε_t is the error term, and T represents the entire period of y_t . $W\left(L^{\frac{1}{m}};\theta\right)$ is a polynomial weight function dependent on the parameters K and θ , the specific expression is shown in Eq. 5.

$$W\left(L^{\frac{1}{m}},\theta\right) = \sum_{k=0}^{K} B(k;\theta)L^{\frac{k}{m}}, \tag{5}$$

where $L^{\frac{1}{m}}$ is the lag operator of $x_t^{(m)}$, and $L^{\frac{k}{m}}x_t^{(m)} = x_{t-k/m}^{(m)}, k = 0,1,\text{L}, K$. K is the maximum lag order of the high-frequency variable.

The MIDAS model involves the selection of polynomial weighting function and how to select parameter vector θ and lag order K in polynomial weighting function. Liu et al. [14] pointed out that Almon polynomial functions are mostly used in the prediction and analysis of Financial market volatility, while exponential Almon polynomial functions and Beta polynomial functions are mostly used in the analysis and prediction of macroeconomic. In the paper, Almon polynomial function is selected as the weight function in Eq. 6.

$$B\left(k;\theta\right)=\frac{\theta_1 k+\theta_2 k^2+\text{L}+\theta_p k^p}{\sum_{k=1}^{K}\left(\theta_1 k+\theta_2 k^2+\text{L}+\theta_p k^p\right)}. \tag{6}$$

For the choice of parameter vectors θ, we employ $p=2$ following Clements and Galvao [15]. For the selection of the lag order K, Gao and Yang [16] compared the adjust goodness of fit and significance level of parameter estimation with different lag orders. In the paper we select the optimal lag order according to the significance level of parameter estimation. Therefore, the model established in the paper is Eq. 7.

$$V_t=\alpha+\beta_1 B\left(L^{\frac{1}{m}};\theta_1\right)index_{TD_t}^{(m)}+\beta_2 B\left(L^{\frac{1}{m}};\theta_2\right)index_{BD_t}^{(m)}+\varepsilon_t, \tag{7}$$

where V_t is the closing price of China iVIX index. $index_{TD}$ and $index_{BD}$ are investor sentiment index.

4 Empirical analysis

The paper mainly studies two cases: (1) whether the investor sentiment index has a significant impact on option implied volatility and (2) whether the MIDAS model has a strong interpretation ability. To ensure that there is no spurious regression, we have to testify the stationarity of time series data before the empirical analysis. The result shows that all variables have passed the ADF test. In order to measure the pros and cons of the interpretation ability of the MIDAS model, Clements and Galvao [15] compared and analyzed the MIDAS model with the benchmark model which is OLS method under the same frequency data. Along this line, the parameter results of both models are given in Table 1 and Table 2 respectively.

Table 1. Estimation results of OLS.

Parameter	Monthly		Weekly	
	value	t-Statistic	value	t-Statistic
α	17.0064 ***	15.2034	17.4272 ***	24.6974
β_1	-5.4056 ***	-8.5374	-5.4299 ***	-13.4035
β_2	15.9687 ***	5.0480	15.3601 ***	8.7348
$Adj-R^2$	0.8133		0.7103	

Note: ***, significance at 1% level; **, significance at 5% level; *, significance at 10% level.

The estimation results of the benchmark model are given in Table1. Table 2 shows the estimation results of the MIDAS models with different frequencies. For both MIDAS model and OLS model, the parameter estimation value of $index_{TD}$ is significantly negative, while $index_{BD}$ is significantly positive. These results are consistent with the conclusions drawn from Fig. 1(d) and Fig. 2.

By comparing the $Adj-R^2$ of different models, we can judge the fit of the model. When the explained variable is the monthly implied volatility of options, the fitting effect of the MIDAS model using daily data for the explanatory variable is better than that of the OLS homo-frequency model. The same results are obtained when the weekly data are used as
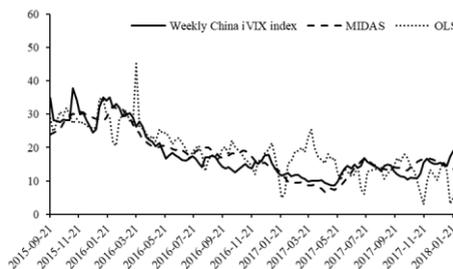
explanatory variables. This indicated that the mixed frequency model has stronger explanatory power than homo-frequency model. However, the fitting effect of Monthly/Weekly MIDAS model is lower than that of monthly OLS model with the homo-frequency. This shows that the interpretation ability of the MIDAS model is not uniformly better than the traditional regression model, but depends on factors such as the difference in frequency multiples used. Fig. 3 shows the fitting results obtained by using Monthly/Daily and Weekly/Daily data. It is intuitive to see that using high-frequency data can better simulate lowfrequency financial data.

Table 2. Estimation results of MIDAS.

Parameter	Monthly / Weekly		Monthly / Daily		Weekly / Daily	
	value	t-Statistic	value	t-Statistic	value	t-Statistic
α	6.1487	1.6529	14.7704***	11.3183	11.8794***	8.5274
β_1	-67.3223**	-2.5978	-1.4671**	-2.4705	-1.4110***	-2.7740
θ_{11}	72.8006**	2.3852	0.2536**	2.1594	0.2688*	1.9146
θ_{12}	-16.9833**	-2.2230	-0.0093**	-2.1072	-0.0088	-1.1418
K_1	3		25		17	
β_2	16.4099*	1.84498	1.2318*	1.6718	1.5493***	3.2233
θ_{21}	-14.0669*	-1.7044	-0.1595	-1.3916	-0.3237**	-2.3390
θ_{22}	2.9652*	1.8301	0.0054	1.5165	0.0233***	2.9410
K_2	4		34		16	
m	4		22		5	
$Adj - R^2$	0.7896		0.8216		0.8418	
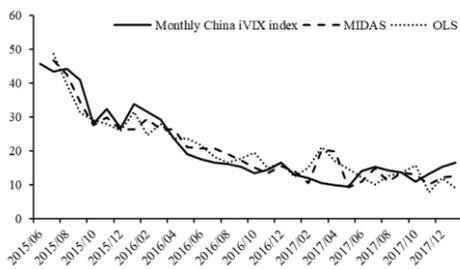
(a) fitting effect on monthly data. (b) fitting effect on weekly data.

Fig. 3. OLS and MIDAS fitting effect on China iVIX index.

It can be seen from the above empirical analysis results that Baidu Index can reflect the level of investor attention to a certain extent. If investors are interested in an option product, they will increase the search volume of relevant information, conversely, the increase in the amount of search information will also affect investors' investment expectations and behaviors, thereby affecting asset price fluctuation. Therefore, the excessive attention of investors on the Internet will exert positive pressure on the option market, which will increase the implied volatility of the option market.

In order to better verify the validity and robustness of MIDAS model, the closing price of the China iVIX index is replaced by the trading volume of SSE 50ETF. The regression results are consistent with the above results, indicating that the research results of the paper are robust.

5 Conclusions

Using the principal component analysis method, the paper constructs two investor sentiment indexes from different angles and frequencies. Two models, MIDAS model and OLS model, are established to analyze the influence of investor sentiment on the implied volatility of options. The empirical results show that the investment behavior of investors has a significant negative impact on the implied volatility of options, and investors' excessive attention to the Internet can increase the implied volatility of the options market. In the process of research, we found that the mixed frequency model has better explanatory power than the homo-frequency model.

The findings of the paper can show the internal mechanism of the influence of investor sentiment on the option implied volatility, and provide references for other relevant researchers. Meantime, as for relevant government departments, in order to ensure a stable, healthy and sustainable development of the security market, it is recommended that the government improve the information disclosure system; monitor investor sentiment changes on official platforms; reasonably guide investors to invest rationally; prevent malicious transmission of false information on social media, and prevent disruption of order in the security market.

References

1. B. Han, Rev. Financ. Stud. **21,** 387 (2008)
2. G. Zhang, Y. Lin, J. Appl. Stat. Manag. **Online,** 1 (2020)
3. X. Xiong, K. Xu, D. Shen, Syst. Eng. - Theory Pract. **40**, 2252 (2020)
4. X. Meng, X. Meng, Y. Hu, Macroecon. **1**, 144 (2016)
5. J. Fleming, B. Ostdiek, R. E. Whaley, J Futures Markets **16**, 353 (1996)
6. R. E. Whaley, J. Portfolio Manage. **26**, 12 (2000)
7. C. M. C. Lee, A. Shleifer, R. H. Thaler, J Finance **46**, 75 (1991)
8. G. W. Brown, Financ. Anal. J. **55**, 82 (1999)
9. M. Baker, J. Wurgler, J Finance **61**, 1645 (2006)
10. J. S. Kim, D. Ryu, S. W. Seo, J. Bank Financ. **42**, 166 (2014)
11. S. H. Kim, D. Kim, J. Econ. Behav. Organ. **107,** 708 (2014)
12. P. Wang, W. Zhang, X. Li, D. Shen, J. Econ. Interact. Coord. **14,** 377 (2019)
13. E. Ghysels, P. Santa-Clara, R. Valkanov, Chapel Hill: University of North Carolina, 2004
14. J. Liu, H. Liu, Z. Yin, Econ. Sci. **5**, 23 (2010)
15. M. P. Clements, B. Galvao, J. Appl. Econ. (Chichester Engl) **24**, 1187 (2009)
16. B. Gao, C. Yang, Int. Rev. Econ. Finance **49**, 69 (2017)