Societal Risk and Stock Market Volatility in China: A Causality Analysis

Nuo Xu[1,2] and Xijin Tang[1,2](✉)

[1] Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100190, China
xunuo1991@amss.ac.cn, xjtang@iss.ac.cn
[2] University of Chinese Academy of Sciences, Beijing 100049, China

Abstract. A variety of societal contradictions and conflicts are exposed in China along the process of economic and social transformation. Online societal risk perception is acquired by public searching behavior which has been mapped into respective societal risks based on indicators including national security, economy/finance, public morals, daily life, social stability, government management, and resources/environment. A stable and harmonious society is the basic guarantee for the sound development of the stock market. What we concern about is whether the variations of the societal risk are related to stock market volatility. The correlations between societal risk and stock market volatility are investigated. Although there are no trading data on holidays and weekends, the risk information of no-trading days is also taken into consideration to discuss if there are any impacts on stock market volatility. Three different econometric approaches are developed to explore the relationship between them. The results show that the risk of finance/economy, social stability, and government management could cause the fluctuation of stock market. Moreover, risk information of no-trading days has an impact on the stock's volatility as well. The research demonstrates that capturing online societal risk based on public searching data is feasible and significant.

Keywords: Societal risk perception · Stock market volatility · Granger causality test · Multiple linear regression

1 Introduction

The transformation of development of economy and society in China is at a crucial stage. The contradictions and conflicts among different social strata, such as the rich and the poor, become increasingly salient and exert adverse effects on social stability, leading to a lot of societal risks. It is known that a stable and harmonious society is the basic guarantee for economy development. That is to say, external social circumstances such as nation's security and foreign relations, government policy, anti-corruption, etc. may have impacts on economic

J. Chen et al. (Eds.): KSS 2017, CCIS 780, pp. 175–185, 2017.
https://doi.org/10.1007/978-981-10-6989-5_15

and financial markets. Meanwhile, higher risk may arise in the process of economy development, which poses a threat to social stability. Therefore, in order to guarantee social stability and the healthy growth of national economy, the possible risks existing in economy and finance and the causes of those risks are worth being analyzed. The stock market could be regarded as the barometer of economic development, to a certain extent, reflecting conditions of the national economy. It is of great significance to study the relationship between societal risk and stock market volatility.

In the Web 2.0 era, traditional media has been surpassed by the Internet which is more open and interactive. Internet users are not only content viewers but also content producers. Among many Internet services, search engines have been the most common tools to access information, not only meeting search requirements but also recording foci of netizens. There have been some research achievements exploiting the search data for predicting economic activities, providing a new perspective to understand the conditions of economy. The CPI was well predicted through utilizing Google search query data [1], as well as retail sales [2]. In the field of finance, search query data were successfully used for predicting dynamics of stock market volatility [3], and measuring the retail investors' attention [4,5]. The research on the correlation between the search volume and investors' attention indicated that the increase of search activity was associated with increase of trading activity [6]. Furthermore, there were studies found that an increase in search volume tended to precede stock market falls [7,8]. The search query data are closely related to stock market at information times. In this paper, we try to investigate the relationship between online societal risk acquired by Internet searching behavior and stock market volatility.

Societal risk perception is the subjective evaluation of public concerns to risk events. Traditional research on societal risk perception is studied from social psychology. The psychometric paradigm of risk perception is designed to ask the public to fill out questionnaires about acceptable risks, then assess public's attitude toward critical societal risk [9]. However, the result of this methodology is restricted by the selection of the samples, as well as the authenticity of the answers. Large-scale surveys of societal risk perception are generally expensive and time-consuming to be conducted. The researchers constructed a new framework of societal risk indicators based on word association tests [10]. Online public concerned data are mapped into respective societal risks based on societal risk indicators. Instead of asking respondents to answer questions about societal risk, the query data provided by the public actively reflect what are real concerned, and thus provide researchers a novel access to analyze societal issues.

In this paper, we concern about the question whether the variations of the societal risk have significant effects on stock market volatility. To answer the question, quantitative societal risk levels are presented and the Shanghai Composite Index is chosen to study stock market volatility. Different from previous research [3], the risk level data of holidays and weekends are also taken into consideration to discuss whether there are any impacts on stock price change or not. Three different econometric methods are developed to explore relationships between societal risk levels and Shanghai Composite Index.

This paper is organized as follows: Sect. 2 illustrates the measurement of societal risk and introduction of Shanghai Composite Index. Section 3 introduces three methods on the relationships between societal risk levels and Shanghai Composite Index. Section 4 presents results analysis. Conclusions and future work are given in Sect. 5.

2 Data Review

2.1 Measurement of Societal Risk

Due to the limitations of the traditional psychological methods, online societal risk perception is proposed to acquire public's concerns toward critical societal risk. Baidu is now the biggest Chinese search engine. Baidu hot news search words (HNSW) are based on real-time search behaviors of hundreds of millions of Internet users and released at Baidu News Portal[1], reflecting the current concerns and ongoing social topics. Therefore, HNSW can serve as the corpus to get netizens' attention to the highlighted events and provide a perspective to analyze societal risk.

HNSW and their relevant hot news with URLs at the first page of hot words search results are crawled every hour [11]. Researchers from CAS Institute of Psychology constructed a framework of societal risk indicators including 7 categories which are national security, economy/finance, public morals, daily life, social stability, government management, and resources/environment [10]. Tang has tried to map HNSW into either risk-free event or one event with risk label from these 7 risk categories, then aggregate all risky events over the whole concerns as the on-line societal risk perception [12]. The daily total risk level is obtained by computing the risky proportion of all hot search words in a day. Figure 1 shows daily total risk levels based on HNSW from January of 2012 to December of 2016 with a total number of 1827 data. The generated 5-year (2012-2016) daily societal risk data provide an overview of China's societal risk situation.

Among the 5-year data, the average risk level of 2012 is 0.554, while that of 2013 is 0.586. However, the average risk levels of 2014, 2015 and 2016 are 0.421, 0.330 and 0.27 respectively, with the annual average risk level decreasing year by year. It may show the achievements of our country in building harmonious socialist society in a way. Moreover, the societal risk levels have dropped off during the Spring Festival, as well as the Olympic Games. As the concept "Harmonious society" was proposed in 2004, there were many studies on the measurement of a harmonious society. Such the Beijing's Harmonious Society Index and Green GDP were put forward to measure the harmonious degree of a society [13]. Besides, a study named "Happiness Survey" was carried out by China Central Television (CCTV) by asking people if they were happy and what

[1] http://news.baidu.com/.

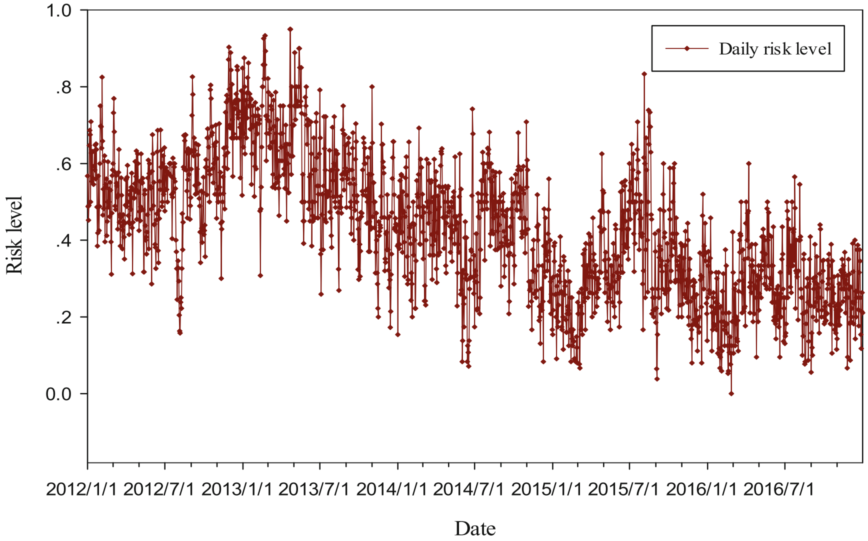

Fig. 1. Daily societal risk level of HNSW

happiness meant. However, these indexes are difficult to be measured and come from different government offices. The daily societal risk levels can serve as an additional way to measure the harmonious degree of a society, which truly and more efficiently reflect the social situations. In other words, the societal risk perception indicators, daily societal risk levels, could be the thermometer of Chinese society to some extent.

Furthermore, the risk level of one category is calculated as one ratio of frequency of words labeled that category over total frequency of hot words. Temporal risk levels of each risk category are acquired, which indicates specific societal risks change over time in details. Researchers had demonstrated that risk levels of each category were predictive of public moods including Happiness, Sadness, Fear, Anger and Disgust when investigating the correlation between societal risk and social moods [14]. In this paper, we study the relation of individual risk level of each category with Shanghai Composite Index.

2.2 Shanghai Composite Index

Shanghai Composite Index is the broadest and most influential comprehensive index across Chinese Security Markets, which can accurately reflect the overall situation of China's stock market. The Shanghai Composite Index is more commonly utilized in many domestic studies [15,16]. In this paper, we download the data from Wind financial terminal, including a total number of 1214 transactions from January 1, 2012 to December 31, 2016. Different from the daily risk level, there are no trading data on holidays and weekends, and thus leads to different sizes of those two sets of data.

The economic and financial issues as one of societal risk categories may cause the oscillations on the risk levels. We wonder if the changes on risk levels of 7 categories could affect financial and economic fluctuations. Therefore, we try to explore the relationship among the levels of 7 categories of societal risks and the Shanghai Composite Index.

3 Correlation Between Societal Risk Levels and Shanghai Composite Index

Baidu hot news search words, online public concerned data, have been mapped into either risk-free events or one event with a label from 7 categories. The risk level of one category is calculated as one ratio of frequency of words labeled that category over the total frequency of hot search words. In this section, we investigate whether the variations of the societal risk have a significant effect on the ups and downs of the stock price. We choose the daily closing price increment to indicate the fluctuation of stock market. Since there are no trading data on holidays and weekends, the data of no-trading days are dropped when analyzing the relation between them. However, risky events happened on no-trading days may have an impact on the future values of Shanghai Composite Index when the next work day is coming. As a result, three approaches with different data processing are developed to analyze the correlation between levels of 7 categories of societal risks and the closing price increment from January 1, 2012 to December 31, 2016.

3.1 Method Based on Granger Causality Test

Due to the lack of trading data on the holidays and weekends, we remove the risk level data of dates that have no transaction data, and we carry out the econometric method of Granger causality test on these two sets of data. The Granger causality test is used to measure the ability to predict the future value of a time series X utilizing prior values of another time series Y [17]. If the time series X is said to Granger-cause Y, the lagged values of X will provide statistically significant information about future values of Y. In order to avoid spurious regression problem, we first test the stationary of each variable (unit root) using the augmented Dickey-Fuller test (ADF test). The original hypothesis is the sequence has at least one unit root. Test results are as shown in Table 1, all the original series are stationary series, that is to say, all variables are suitable for Granger causality test directly.

Then we utilize vector autoregressive model (VAR) to carry out Granger causality test. The daily societal risk level is chosen as independent variable, while the closing price increment is chosen as the dependent variable to detect if societal risk perception can affect the volatility of stock price. As far as the lag orders, we just consider it delaying for 5 days. Table 2 shows the results of Granger causality correlation of the closing price increment and societal risk levels of 7 categories.

Table 1. Results of ADF test

Variable	ADF t-statistics	P value	1% level	5% level	Results
National security	−9.73	0.00	−3.96	−3.41	stationary
Economy/ finance	−10.46	0.00	−3.96	−3.41	stationary
Public morals	−14.34	0.00	−3.96	−3.41	stationary
Daily life	−12.21	0.00	−3.96	−3.41	stationary
Social stability	−9.44	0.00	−3.96	−3.41	stationary
Government management	−8.51	0.00	−3.96	−3.41	stationary
Resources/ environment	−19.08	0.00	−3.96	−3.41	stationary
Closing price increment	−14.95	0.00	−3.96	−3.41	stationary

Table 2. Granger causality correlation of the closing price increment and 7 societal risks

Variable	Lag				
	1 day	2 days	3 days	4 days	5 days
National security	0.0685	0.0926	0.1087	0.1756	0.2599
Economy/finance	0.4523	0.0601	0.1757	**0.0074	**0.0007
Public morals	0.2377	0.4075	0.5714	0.8174	0.6452
Daily life	0.2042	0.4247	0.4813	0.5079	0.6522
Social stability	*0.0372	0.0696	0.0994	0.0669	0.0900
Government management	0.3277	0.2541	0.1861	0.3233	0.4216
Resources/environment	0.8368	0.3535	0.5559	0.6335	0.7661

*Correlation is significant at the 0.05 level
**Correlation is significant at the 0.01 level

According to the results of Granger causality in Table 2, we observe that economy/finance has the Granger causality relation with closing price increment lagging from 4 and 5 days while social stability lagging 1 day.

3.2 Method Based on Decaying Exponential Function

Different from the first approach, in order to capture the risk information of no-trading days, we define a decaying exponential function. Through setting different weight coefficients, the risk information of holidays or weekends is added to the risk level of the day just before holidays or weekends. Therefore, the risk level of the day before holidays or weekends r_m is recomputed based on original risk level and given by Eq. 1.

$$r_m = \sum_{i=0}^{n} \frac{r_{m+i}}{c^{n+1-i}}. \tag{1}$$

Here, n is the length of successive days with no trading data. For normal weekends, $n = 2$. c is a dampening constant that gives a higher weight to the risk level of recent days. In this paper, we choose $c = 2$. The exponential function is chosen so that distant risk level affects r_m without outweighing more recent risk levels.

Based on the defined decaying exponential function above, we obtain a new sequence of risk levels based on the original data. Then ADF test is conducted. We test the unit root of closing price increment and the new sequence of risk levels. The test results in Table 3 show that all of them have no unit root.

Table 3. Results of ADF test

Variable	ADF t-statistics	P value	1% level	5% level	Results
National security	−9.69	0.00	−3.96	−3.41	stationary
Economy/ finance	−10.36	0.00	−3.96	−3.41	stationary
Public morals	−14.45	0.00	−3.96	−3.41	stationary
Daily life	−12.25	0.00	−3.96	−3.41	stationary
Social stability	−9.59	0.00	−3.96	−3.41	stationary
Government management	−8.57	0.00	−3.96	−3.41	stationary
Resources/ environment	−17.44	0.00	−3.96	−3.41	stationary
Closing price increment	−14.95	0.00	−3.96	−3.41	stationary

Again, we employ Granger causality analysis. The results are shown in Table 4. As we can see from Table 4, economy/finance has the Granger causality relation with closing price increment for lagging 2, 4 and 5 days, while social stability has the Granger causality relation for lagging 4 and 5 days.

3.3 Method Based on Multiple Linear Regression 3

To avoid changing the original societal risk data, we apply the multiple linear regressions to analyze the relationship between the temporal data of societal risk and closing price increment, which is chosen as the dependent variable. Different days of the societal risk level of one category are chosen as independent variables. There will generate a corresponding number of sequences to the lag number of days for any risk category. As a result, we construct 7 regression models for each

Table 4. Granger causality correlation of the closing price increment and 7 societal risks

Variable	Lag				
	1 day	2 days	3 days	4 days	5 days
National security	0.2257	0.2848	0.2341	0.2860	0.3357
Economy/finance	0.6153	*0.0190	0.0554	**0.0095	**0.0026
Public morals	0.8974	0.9968	0.9695	0.9109	0.5051
Daily life	0.3423	0.6086	0.6206	0.7193	0.8257
Social stability	0.0778	0.0618	0.0809	*0.0145	*0.0246
Government management	0.5790	0.2877	0.2845	0.4411	0.5636
Resources/environment	0.9005	0.3722	0.5455	0.6995	0.8168

*Correlation is significant at the 0.05 level
**Correlation is significant at the 0.01 level

category of societal risk with a total of 35 risk level sequences. The multiple linear regression function is defined as Eq. 2:

$$y_t = \gamma + \sum_{j=0}^{k} \alpha_j y_{t-j} + \sum_{j=0}^{k} \beta_j x_{t-j} + u_t. \tag{2}$$

Here, y_t represents the closing price increment of the tth day, while y_{t-j} indicates that of lagging j trading days. x_{t-j} is societal risk level of one category which lags j days. α_j and β_j are the regression coefficients. u_t is the error term which reflects the parts of changes in y that cannot be explained by x. To match the lag orders with the two approaches above, we set $k = 5$. The results of statistic values for either the significance of coefficients or significance of multiple linear regression models are respectively listed in Tables 5 and 6.

Table 5. Results of significance of coefficients

Model	Lag				
	1 day	2 days	3 days	4 days	5 days
M1: National security	0.0896	0.3854	0.9964	0.5955	0.2114
M2: Economy/finance	0.2975	*0.0168	0.1648	0.3096	0.8724
M3: Public morals	0.6498	0.7102	0.8918	0.9043	0.5231
M4: Daily life	0.7196	0.6811	0.1005	0.4515	0.8274
M5: Social stability	0.4479	*0.0116	0.0824	0.5251	0.2407
M6: Government management	0.9353	0.4999	0.3789	0.0594	*0.0476
M7: Resources/environment	0.9264	0.8683	0.9208	0.3362	0.8298

*Correlation is significant at the 0.05 level

As shown in Table 6, all the regression models are significant based on F statistics. The sequences of economy/finance lagging 2 days have significant effects on closing price increment, so as those of social stability lagging 2 days, and those of government management lagging 5 days. The result is different from those of previous two approaches.

Table 6. Results of significance of multiple linear regression models

Model	F-statistics	P value	D.W
M1	5.5301	0.0000	2.0026
M2	5.8288	0.0000	1.9993
M3	5.2323	0.0000	2.0008
M4	5.4664	0.0000	1.9992
M5	6.3818	0.0000	1.9977
M6	5.6184	0.0000	1.9984
M7	5.2745	0.0000	2.0000

4 Results Analysis

Based on the above results, three different approaches demonstrate distinguishing findings on the correlation between societal risk and stock volatility. As to the first approach, the original data sequence of societal risk level is used to carry out the research, while the risk data on no-trading days are ignored without regarding their effects on stock volatility. The experimental results show that economy/finance and social stability have significant effects on closing price increment. In terms of the second approach, the decaying exponential function is adopted to capture the risk information of no-trading days by processing the original data. The results are different from the first with different delay days on the risk of economy/finance and social stability. The subcategories of economy/finance include financial problems and economic problems. As is known, the macroeconomic policy adjustment, the economic slowdown, the monetary policy, RMB exchange rate etc. will cause the fluctuation of stock prices. Besides, stable social situations is a vital prerequisite for development of finance and economy. Events that threaten social stability such as public security, crime, and major infectious diseases etc. which are the subcategories of social stability have a certain impact on stock market.

Compared to the two previous modeling approaches, not only do we consider the effects of risk information on no-trading days in the third approach, but also avoid processing the raw data. We construct regression models by generating a corresponding number of sequences to the lag number of days for any risk category. Moreover, the result has a novel finding that the risk government management has an impact on closing price increment, which shows that the raw data carry more information. The risk levels of holidays and weekends influence the closing price increment through different processing of risk information on no-trading days. In other words, the search queries data on holidays and weekends may have some effects on stock market. The subcategories of government management such as government policy and anti-corruption etc. also influence the volatility of the stock market. For instance, the economy and stock market of Lvliang city had crashed since the anti-corruption movement in Shanxi. However, the data processing of the third approach is a more complex work due to the construction of a total of 35 risk level sequences for each category.

The experimental results of the three methods show that social risk perception has a certain prediction effect on the Shanghai Composite Index. Given all that, the collective risk perception, reflect the public opinions' impacts on the stock market volatility, as well as social and economic conditions to some extent.

5 Conclusions

Societal risk perception is the subjective evaluation of public concerns to risk events. It reflects the public attitudes to social issues as well as government decision-making, which are the key indicators for effective social management and policy making. With the rapid growth of Internet data, an increasing number of researchers make use of user-generated contents to study social issues.

In this paper, we map on-line community concerns into respective societal risk events based on Baidu search engine by using the public searching behavior. Then we conduct the research on exploring the relationship between societal risk perception and stock market volatility. The main contributions and innovations are summarized as follows.

(1) Societal risk perception is quantitatively described. We present a reliable, scalable and timely assessment of the societal risk instead of large-scale survey which is expensive and time consuming to be conducted.
(2) The study on correlation between societal risk and stock price change is first carried out. Public searching queries are mapped into specific societal risk events. Then we analyze what kinds of risk events with risk labels from 7 risk categories have impacts on stock volatility. The results show that the risk of finance/economy, social stability and government management may cause the fluctuation of stock market.
(3) Different from previous research, the risk information of holiday is also taken into consideration to discuss the impact on stock price change based on three different methods. The results indicate that risk information of no-trading days can affect the stock volatility.

Lots of improvements are needed. The intrinsic mechanism of the relationship between societal risk and stock price fluctuation requires further research. Meanwhile, quantitative societal risk data which are more time-sensitive will be essentially used to study more social issues.

Acknowledgments. This research is supported by National Key Research and Development Program of China (2016YFB1000902) and National Natural Science Foundation of China (61473284 & 71371107).

References

1. Zhang, C., Lv, B.F., Peng, G., Liu, Y., Yuan, Q.Y.: A study on correlation between web search data and CPI. In: Gaol, F. (ed.) Recent Progress in Data Engineering and Internet Technology. LNEE, vol. 157, pp. 269–274. Springer, Heidelberg (2012). doi:10.1007/978-3-642-28798-5_36
2. Choi, H., Varian, H.: Predicting the present with Google trends. Econ. Rec. **88**, 2–9 (2012)
3. Thomas, D., Jank, S.: Can internet search queries help to predict stock market volatility? Eur. Financ. Manage. **22**, 171–192 (2016)
4. Bank, M., Larch, M., Peter, G.: Google search volume and its influence on liquidity and returns of German stocks. Fin. Markets Portfolio Mgmt. **25**, 239–264 (2011)
5. Da, Z., Engelberg, J., Gao, P.: In search of attention. J. Financ. **66**, 1461–1499 (2011)
6. Takeda, F., Wakao, T.: Google search intensity and its relationship with returns and trading volume of Japanese stocks. Pac. Basin Financ. J. **27**, 1–18 (2014)
7. Curme, C., Preis, T., Stanley, H.E., Moat, H.S.: Quantifying the semantics of search behavior before stock market moves. Proc. Nat. Acad. Sci. **111**, 11600–11605 (2014)
8. Zhao, Y., Ye, Q., Li, Z.: The relationship between online attention and share prices. In: WHICEB 2014 Proceedings, pp. 462–469 (2014)
9. Xie, X.F., Xu, L.C.: The study of public risk perception. Psychol. Sci. **6**, 723–724 (2002). (in Chinese)
10. Zheng, R., Shi, K., Li, S.: The influence factors and mechanism of societal risk perception. In: Zhou, J. (ed.) Complex 2009. LNICSSITE, vol. 5, pp. 2266–2275. Springer, Heidelberg (2009). doi:10.1007/978-3-642-02469-6_104
11. Hu, Y., Tang, X.: Using support vector machine for classification of Baidu hot word. In: Wang, M. (ed.) KSEM 2013. LNCS, vol. 8041, pp. 580–590. Springer, Heidelberg (2013). doi:10.1007/978-3-642-39787-5_49
12. Tang, X.J.: Exploring on-line societal risk perception for harmonious society measurement. J. Syst. Sci. Syst. Eng. **22**, 469–486 (2013). doi:10.1007/s11518-013-5238-1
13. Tang, X.J.: Applying search words and BBS posts to societal risk perception and harmonious society measurement. In: IEEE International Conference on Systems, Man, and Cybernetics, pp. 2191–2196. IEEE Press (2013). doi:10.1109/SMC.2013.375
14. Dong, Y.H., Chen, H., Tang, X.J., Qian, W.Y., Zhou, A.Y.: Collective emotional reaction to societal risks in China. In: IEEE International Conference on Systems, Man, and Cybernetics, pp. 557–562. IEEE Press (2015). doi:10.1109/SMC.2015.108
15. Zhao, B., He, Y., Yuan, C., Hang, Y.: Stock market prediction exploiting microblog sentiment analysis. In: IEEE International Joint Conference on Neural Networks, pp. 4482–4488. IEEE Press (2016). doi:10.1109/IJCNN.2016.7727786
16. Chu, X., Wu, C., Qiu, J.: A nonlinear Granger causality test between stock returns and investor sentiment for Chinese stock market: a wavelet-based approach. Appl. Econ. **48**, 1915–1924 (2016). doi:10.1080/00036846.2015.1109048
17. Gao, T.M.: Econometric Analysis and Modeling. Tsinghua University Press, Beijing (2006). (in Chinese)