Analysis of the Impact of Investor Sentiment on Stock Price Crash Risk

Haonan Yu *

School of Statistics, Capital University of Economics and Business, Beijing, China

* Corresponding Author Email: 32022120069@cueb.edu.cn

Abstract. The stability of the stock market is very important to the country. Some researchers have found that sentiment has an impact on stock price crash risk. However, research mainly focuses on media sentiment and company sentiment, and research on investor sentiment is not mature enough. This article collates the literature to explore the relationship between investor sentiment and stock price crash risk. This article provides a systematic classification of indicators that measure stock price crash risk. Then, when introducing the methods of measuring investor sentiment, the indicators for measuring investor sentiment are classified, and the scope of application of the three mainstream methods is summarized. This article introduces three models commonly used in the academic community for this problem and conducts a comparative analysis of the three mainstream models. Research has found a positive relationship between investor sentiment and stock price crash risk. As crisis sentiment rises, a large number of investors will begin to implement stop-loss strategies and even sell at extremely low prices to minimize losses. As a result, the likelihood of a stock market price crash increases significantly. In addition, the selection of data sets affects the measurement of investor sentiment to a great extent, and the development and utilization of alternative data is a future research direction. Beyond this, there is reverse causation between investor sentiment and stock price crash risk. Therefore, it is necessary to consider reducing or avoiding the impact of endogeneity problems when selecting a model.

Keywords: Alternative data; sentiment; investor sentiment; stock price crash risk.

1. Introduction

Currently, China is facing problems such as insufficient demand for domestic economic development and turmoil in the international financial market, which have brought challenges to China's financial stability. In this context, the security and stability of the stock market, which is an important part of the financial market, is very important.

When stock market prices fluctuate severely, there will be a risk of stock price collapse, which will have a huge impact on the financial market and even the entire economic system. Since the establishment of the Shanghai and Shenzhen stock exchanges in the 1990s, China's stock market has often experienced sharp declines. After the financial crisis in 2008, China's stock market has been in a sluggish state. Although the stock market ushered in a bull market in 2014, it brought some improvement. But not long after, in the second half of 2015, more than a thousand stocks fell to the limit, and in early 2018, there was a global stock price plunge [1]. These events not only affect investors' personal wealth but also negatively impact social and economic development at the macro level.

With the rapid development of technology and finance, the combined application of finance and technology has become a new development direction. Applying a series of new technologies such as big data, cloud computing, and blockchain to study stock price crash risk can help scholars conduct empirical research, and then help scholars discover potential factors that affect stock price crash risk [2]. This has certain implications for all parties in the stock market.

Currently, research on stock price crash risk is mainly conducted at the market level and company level. At the market level, a large number of studies have been carried out from the perspective of behavioral finance, taking into account the impact of investor cognitive biases and market heterogeneity. At the same time, some studies take the information asymmetry theory as the core. At the company level, research mainly starts from internal factors, focusing on the company's financial status, management decisions, and corporate culture as the starting point to explore the impact of stock price crash risk. Research considers external factors and mainly focuses on the impact of the macroeconomic environment and social and political policies on stock price crash risk [2, 3].

A large number of scholars have explored and verified the influencing factors of stock price crash risk through empirical research. Existing research considers the impact of financial reporting, corporate disclosure, management incentives, management characteristics, capital markets, corporate governance, informal supervision mechanisms, and other factors on stock price crash risk, and these studies mostly use traditional financial data [4]. In recent years, many studies have considered the impact of media tone, employee quality, information exchange among investors, online interactions between company management and investors, corporate digital transformation, environmental penalties, and other factors on stock price crash risk. These studies use more alternative data, and the research results enrich the factors that affect the risk of stock price crash [5-10].

An increasing number of studies have proven that market sentiment can affect stock returns and stock volatility, and can even cause asset prices to deviate from their intrinsic value in the long term [11]. Existing research focuses more on media sentiment and company sentiment. Investor sentiment is difficult to collect and measure, resulting in a gap in research in this direction. As an important part of the stock market, investors explain the causes of stock price crash risks from the perspective of investor sentiment, which is of certain significance for stabilizing the operation of the stock market and reducing market risks. Therefore, this article collates the literature to explore the relationship between investor sentiment and stock price crash risk.

This article explains the theoretical basis of stock price crash risk and systematically classifies the indicators for measuring stock price crash risk. The importance of studying investor sentiment is then pointed out. When introducing the methods of measuring investor sentiment, the indicators for measuring investor sentiment are classified and the scope of application of the three mainstream methods is summarized. Finally, the three models commonly used in the academic community for this problem are pointed out. Based on the model operation results presented by different scholars in different articles, the three mainstream models are comparatively analyzed. This article aims to make a certain contribution to the development of the field of stock price crash risk.

2. Stock Price Crash Risk

2.1. Theoretical Basis

Underlying the risk of stock price crashes is the tendency of managers to withhold bad news for long periods, to store it away. When bad news accumulates to a certain threshold, it will be suddenly disclosed to the market, causing a sharp drop in stock prices [3].

2.2. Measurement Method

According to existing literature, there are roughly the following five indicators for measuring stock price crash risk. Negative return skewness coefficient NCSKEW, return fluctuation ratio DUVOL, dummy variable Crash, number of crashes Count, and implied volatility IV_SKEW [1, 3]. In the academic world, there is currently no theory that can point out which method is superior. In most empirical studies, scholars use NCSKEW and DUVOL indicators simultaneously, and the remaining indicators mentioned above are mostly used for robustness tests. Scholars made corresponding adjustments to the NCSKEW and DUVOL indicators so that they can be applied in different situations, and proved the robustness of the results through the similarity of the signs and significance of the coefficients [3].

3. Investor Sentiment

3.1. Importance

Based on the "information hiding hypothesis", scholars use the information disclosure effect and emotional contagion effect to explain the formation mechanism of stock price crash risk: The former assumes that investors react quickly and consistently to information in the market and discusses how the quantity and quality of information disclosure affect stock prices. The latter starts from the heterogeneity of investors and analyzes the impact of investor differences and limited rationality on the risk of a stock price crash. China's stock market is different from other countries in that it has obvious retail characteristics. Investors' psychological expectations and decision-making behaviors have an important impact on the operation of the stock market. On the one hand, due to the single channel for individual investors to obtain information, it is difficult to accurately predict the future fluctuation trend of stock prices in the face of information asymmetry. On the other hand, individual investors have limited professional knowledge and are prone to cognitive biases such as herding and herd mentality. Extreme changes in investor sentiment may trigger violent fluctuations in stock prices, thereby exacerbating the risk of stock price collapse [11]. It can be seen that studying investor sentiment will promote academic research on stock price crash risk.

3.2. Index

Investor sentiment can be reflected through communication among investors, investor attention, and market-wide sentiment. In existing research, some scholars apply dictionary-based sentiment lexicon methods and corpus-based machine learning methods for text analysis. The metrics mentioned below are commonly used to quantify investor communication, investor attention, and market-wide sentiment. NEG_SENT is defined as the proportion of negative words focusing on the L&M sentiment category [12]. The negative word list includes 2349 words such as failure, loss, bankruptcy, and decline. UN_SENT is defined as the proportion of uncertainty words that focus on the L&M sentiment category [12]. The uncertainty word list includes 291 words such as dependence, volatility, exposure, risk, and speculation. The NEG_SENT and UN_SENT indicators take into account the communication between investors very well. The FEARS index is obtained by summarizing Google search volume keywords using a negative tone [12]. The FEARS index is a good reflection of investor attention. The VSTOXX index expresses investors' fear and economic uncertainty caused by the implied volatility of the Euro STOXX 50 Index [12]. The VSTOXX index reflects market-wide sentiment very well.

3.3. Research Methods

3.3.1. Dictionary-based sentiment lexicon

NEG_SENT, UN_SENT, and FEARS indicators are dictionary-based sentiment lexicon methods. The sentiment lexicon method requires the definition of a dedicated word set [12]. The process is subjective, dependent on the definer's ideas, and can be tedious. In addition, whether the vocabulary included in the sentiment vocabulary is comprehensive and whether there are omissions may affect the quantitative process of sentiment analysis. In addition, whether the emotional word set established for a specific text is applicable to other types of texts and how to measure its applicability are also issues that need to be considered. These problems reflect that the scientific nature and universal applicability of emotional lexicon methods still need to be improved. However, the sentiment lexicon method is still widely used in the field of text analysis because of its simplicity and ease of operation.

3.3.2. Machine learning-based methods

The VSTOXX index is constructed based on machine learning methods. Compared with the emotional dictionary method of dictionaries, the machine learning method improves the scientific nature and is reflected in more accurate classification. However, the difficulty of operation has increased significantly, which is reflected in the fact that the construction of the feature training set requires manual operation [12].

3.3.3. Deep learning-based methods

In recent years, with the development of machine learning, the concept of deep learning has been proposed, and its advantages in text analysis have gradually emerged. Relying on the neural network model, some scholars proposed the LSTM-CNN feature fusion model. This model is a fusion of a convolutional neural network and a bidirectional long short-term memory model [12]. The advantage of the convolutional neural network model is to extract local features of text vectors, and the advantage of the bidirectional long short-term memory model is to extract global features related to the text context. The LSTM-CNN feature fusion model combines the features extracted by the two complementary models to solve the problem of the single convolutional neural network model ignoring the contextual semantic and grammatical information of the word. At the same time, the problems of gradient disappearance or gradient dispersion in traditional recurrent neural networks are avoided. It can be seen that this model improves the accuracy of classification. Compared with emotional lexicon and machine learning models, deep learning models are more scientific and universal.

3.3.4. The scope of application of the three mainstream methods

The dictionary-based sentiment lexicon method is suitable for texts that are short in length and have weak contextual connections. Machine learning-based methods are suitable for texts with relatively small amounts of data. Methods based on deep learning are suitable for texts with large amounts of data, and different models can be selected, improved, or integrated according to different text characteristics [13].

4. Research Model

4.1. OLS Model (Ordinary Least Squares Method)

The first row in Table 1 shows the results obtained by applying the OLS model in "Investor Sentiment and Stock Price Crash Risk: Evidence from China" by Yunqi Fan et al. The results in the first and third columns show that when DUVOL and NCSKEW are used as explained variables, the regression coefficients of investor sentiment obtained are both significantly positive at the 1% significance level. However, the results of the robustness test are not ideal. The two investor sentiment regression coefficients obtained are both significantly positive at the 10% significance level.

The OLS model has been applied to the issue of stock price crash risk for a long time. However, because the lagged value of the dependent variable is used as an explanatory value, OLS produces biased estimates. Therefore, many scholars have optimized the OLS model and established It has introduced other models and enriched the application scenarios [14, 15].

4.2. Panel Vector Autoregressive Model

The second row in Table 1 shows the results obtained by applying the panel vector autoregressive model in “Crisis sentiment and banks’ stock price crash risk: A missing piece of the puzzle?” by Tzomakas Christos et al. The results from the first to the fourth columns show that when DUVOL and NCSKEW are used as explained variables, the regression coefficients of investor sentiment obtained are all significantly positive at the 1% significance level. The results of the robustness test are relatively ideal, and the two investor sentiment regression coefficients obtained are significantly positive at the 1% and 10% significance levels respectively.

Based on the OLS model, the PVAR model applies the Generalized Method of Moments (GMM) to make the lagged values of the dependent variable and the independent variable an instrument. The panel vector autoregressive model is more scientific than the OLS model [12].

Variables in the PVAR system are treated as endogenous and allow for potential unobserved individual heterogeneity across the countries and periods examined, allowing the relationship between each variable and its lagged term to be fully studied. When using this model, to obtain the optimal lag length of the variable, Moment and Model Selection Criteria (MMSC) need to be applied for GMM estimation. After this, the baseline PVAR estimate and Granger causality test are performed for each indicator that represents the stock price crash risk, and the orthogonal impulse response function of each indicator is checked to see whether it is consistent with the baseline PVAR estimate. After this, a robustness check is performed, first by variable replacement, and the above steps are repeated. We then continue to check the robustness of the baseline PVAR estimates by controlling for macroeconomic factors.

4.3. Unbalanced Two-Way Fixed Effects Model

The third row in Table 1 shows the results obtained by applying the non-equilibrium two-way fixed effects model in "Investor Sentiment Mining Based on Deep Machine Learning and Its Impact on Stock Price Crash Risk" by YIN Hai-yuan et al. The results from the first to the fourth columns show that when DUVOL and NCSKEW are used as explained variables, the regression coefficients of investor sentiment obtained are all significantly positive at the 1% significance level. The results of the robustness test are ideal, and the two investor sentiment regression coefficients obtained are both significantly positive at the 1% significance level.

Due to the reverse causal relationship between the explained variable and the explanatory variable, the model needs to consider the endogeneity problem [11]. The endogeneity test was conducted using the instrumental variable method, propensity score matching method (PSM), and DID test. Variable substitution and placebo tests were used to check the robustness of the model. Different from the panel vector autoregressive model, this model examines heterogeneity in the following ways: 1. Explore the impact of differences in company size through regression results under different company sizes. 2. Study the impact of differences in ownership concentration through the regression results of different ownership concentrations. 3. Analyze the impact of differences in short-selling restrictions through regression results under different short-selling systems. 4. Through the regression results under different marketization levels, observe the impact of differences in marketization levels. 5. Study the mediating effect of liquidity through the regression results of the mediating effect of stock liquidity.

4.4. Comparison of Three Mainstream Models

As a classic model, the OLS model still has a large number of applications in the academic community. Compared with the unbalanced two-way fixed effects model, the panel vector autoregressive model does not need to be tested for endogeneity because its variables are considered endogenous. Compared with the panel vector autoregressive model, the non-equilibrium two-way fixed effects model can conduct more expanded research and facilitate the testing of heterogeneity and mediation effects, because its essence is a basic regression model. Table 1 shows the results for NCSKEW and DUVOL using three different models, as well as the results of the robustness check.

Table 1. Running results of the three models.

	Results 1 for NCSKEW	Results 2 for NCSKEW	Results 1 for DUVOL	Results 2 for DUVOL	Results 1 of robustness check	Results 2 of robustness check
OLS model	0.609*** (3.90)	no	0.488*** (3.64)	no	0.115* (1.66)	0.998* (1.65)
Panel vector autoregressive model	0.338*** (0.094)	10.865*** (1.165)	0.020*** (0.007)	0.238*** (0.050)	0.011* (0.007)	0.794*** (0.101)
Unbalanced two-way fixed effects model	0. 382*** (4. 234)	0. 182*** (4. 337)	0. 355*** (4. 071)	0. 169*** (4. 165)	0. 030*** (2. 864)	0. 014*** (2. 909)

5. Conclusion

This study finds a positive relationship between investor sentiment and stock price crash risk. Investors rely not just on bank-level or macro fundamentals, but also on their perceptions of risk. As crisis sentiment rises, a large portion of investors will begin to implement stop-loss strategies and even sell at very low prices to minimize losses. As a result, the likelihood of a collapse in stock prices has increased significantly.

Most previous studies discuss stock price crash risk from the company level, focusing on factors such as CEO power, brand capital, and ESG performance. In recent years, with the development of technology, the types and amounts of alternative data that can be obtained have increased. It provides a new direction for the study of stock crash risk, and many scholars have drawn relevant conclusions in new fields through empirical research. The optimistic tone of Internet media is positively related to the risk of future stock price crashes, and companies' online interactions with investors can significantly reduce the risk of future stock price crashes. The role of alternative data does not stop there. For example, data such as the number of times a web page is forwarded and the average time readers spend on the web page are of practical significance for studying investor sentiment. In addition, how to use sensors to obtain data about individual investors will also be a future research direction. Proper use of alternative data can help gauge investor sentiment and, in turn, help study stock price crash risks.

This study found that the results of some models cannot well demonstrate the relationship between investor sentiment and stock price crash risk. This is because there is reverse causality between investor sentiment and stock price crash risk, and not all models can handle endogeneity issues well.

This study also has some limitations. When the model is the same, applying different data sets to measure investor sentiment will result in different operating results of the model. When the data sets are the same, applying different models will also yield different results. In the future, we can study whether there is an optimal model and whether there is an optimal data set. In addition, there may not be an independent optimum, and the optimum is based on the appropriate combination of data set and model.

Overall, this study provides insights into the impact of investor sentiment on stock price crash risk and provides a basis for future research and practice. Finally, we hope that this research can make a certain contribution to the development of related fields and trigger more meaningful discussions.

References

[1] Zhou L, Han J. Literature review on stock price crash risk. Journal of Chongqing University of Science and Technology: Social Science Edition, 2019 (1): 41-44.

[2] He J, Xu F. Research on the stock market systemic risk of financial technology development. Statistics and Information Forum, 2024, 39(03): 40-52.

[3] Xiong Z. Literature review on stock price crash risk. Shanxi Agricultural Economics, 2020, (01): 167-168.

[4] Yifei G, Yanchun X, Xuehua X, et al. Management Earnings Forecasts Bias, Internal Control, and Stock Price Crash Risk: New Evidence from China. Emerging Markets Finance and Trade, 2023, 59(8): 2331-2343.

[5] Zhao R, Fan R, Xiong X, et al. Media Tone and Stock Price Crash Risk: Evidence from China. Mathematics, 2023, 11(17).

[6] Liangcheng W, Wenhui Z, Tao P, et al. Employee quality and stock price crash risk. Asia-Pacific Journal of Accounting Economics, 2023, 30(4): 1047-1066.

[7] Jie L, Zhongqiang Z, Yongjie Z, et al. Information interaction among institutional investors and stock price crash risk based on multiplex networks. International Review of Financial Analysis, 2023, 89.

[8] Yi L, Pengfei W, Wei Z. Does online interaction between firms and investors reduce stock price crash risk? The British Accounting Review, 2023, 55(4).

[9] Shuo Z, Jun G, Kaijun C. Digital transformation and risk of share price crash: Evidence from a new digital transformation index. Finance Research Letters, 2023, 58(PB).

[10] Li M, Shen C, Wen M. The Effect of Firm-Specific Environmental Punishment on Stock Price Crash Risk: Evidence from China. SAGE Open, 2023, 13(4).

[11] Haiyuan, Y, Zaohong N. Mining investor sentiment based on deep learning and its impact on stock price crash risk. Journal of Central University of an, 2024, (03): 36-56.

[12] Christos T, Dimitrios A, Apostolos K, et al. Crisis sentiment and banks' stock price crash risk: A missing piece of the puzzle? Journal of International Financial Markets, Institutions Money, 2023, 87.

[13] Helong L, Changsong R, Xinru L, et al. A review of research on text sentiment in financial markets. Data Analysis and Knowledge Discovery, 2023, 7(12): 22-39.

[14] Bashir U, Kayani N U, Khan S, et al. Investor sentiment and stock price crash risk: The mediating role of analyst herding. Computers in Human Behavior Reports, 2024, 13100371.

[15] Yunqi F, Fangzhao Z, Yunbi A, et al. Investor Sentiment and Stock Price Crash Risk: Evidence from China. Global Economic Review, 2021, 50(4): 310-339.