ACCOUNTING & FINANCE

Accounting & Finance 57 (2017) 1635–1670

Media sentiment, institutional investors and probability of stock price crash: evidence from Chinese stock markets

Yanjian Zhu[a], Zhaoying Wu[b], Hua Zhang[a], Jing Yu[c]

[a]*School of Economics, Zhejiang University, Hangzhou, China*
[b]*School of Economics, Fudan University, Shanghai, China*
[c]*Department of Accounting and Finance, University of Western Australia, Crawley, WA, Australia*

Abstract

Using a large sample of firm-level media reports data, we examine whether and how media reports affect the probability of stock price crash in China. We find that positive media reports reduce the probability of stock price crash, while the relationship between negative reports and the probability of price crash is U-shaped. The probability of stock price crash is more sensitive to the media reports in SOEs and large firms. Furthermore, we find evidence to support the media management behaviour of institutional investors. Such behaviour significantly changes the probability of stock price crash. However, we only observe the media management behaviour of institutional investors in firms held by non-block institutions, in support of the notion that transient investors behave opportunistically and reap short-term investment gains through media management.

Key words: Institutional investors; Media management; Media sentiment; Probability of stock price crash

JEL classification: G14, G23, G34

doi: 10.1111/acfi.12355

We would like to thank the two referees who provided very good suggestions to our study. We would like to thank the 2017 China Accounting and Finance Conference organized by the Central University of Finance and Economics. Yanjian Zhu is thankful for the financial support provided by the National Natural Science Foundation of China (No. 71403238) and the Art Education & Research Supporting Program of Zhejiang University. Please address correspondence to Hua Zhang via email: prozhwa@126.com

1. Introduction

The probability of firm-specific stock price crash is defined as the likelihood of the occurrence of extreme negative returns in future (Jin and Myers, 2006; Kim *et al.*, 2010, 2011). Two significant events, the global financial crisis of 2008 and the extreme fluctuations of Chinese stock markets in 2015, are representative of the damage caused by market crashes. Stock market crash threatens the economy by aggravating the cost of external capital and eroding shareholders wealth. It is important to study the factors that may lead to firm-specific stock price crashes, especially in emerging stock markets where markets are under development and volatile in nature due to weak institutional infrastructure and lack of market efficiency.

Prior studies attribute the occurrence of stock price crash to one type of the information asymmetry problem, which arises from managers concealing negative information (Hutton *et al.*, 2009; Kothari *et al.*, 2009). Hiding negative news from the public inevitably leads to share price overvaluation and therefore creates a price bubble. However, the price will eventually collapse when managers are no longer able to delay or conceal negative news and the accumulated bad news has to be released into the market over a short period of time.

This study examines the role of media reports on the likelihood of stock price crash due to its effects on information asymmetry problem. On the one hand, media reports expedite the information transmission from managers to outside investors and act as a strong external discipline to corporate managers. The timely information releases through the press mitigates the information asymmetry problem, and its monitoring role increases the private costs of managers suppressing bad news (Dyck and Zingales, 2004; Dyck *et al.*, 2008; Kong *et al.*, 2013). Both channels alleviate information asymmetry problem and reduce the probability of stock price crash. On the other hand, media reports not only affect information volume but also investor sentiment. Negative media reports could amplify investor pessimism within a short window, inducing stock price crash (Chan, 2003; Tetlock, 2007; Tetlock *et al.*, 2008; Joe *et al.*, 2009). Thus, the role of media reports on the probability of stock price crash is an open question.

We examine this research question using a comprehensive sample of media reports from the news headlines of news.baidu.com for all listed Chinese firms between 2012 and 2016. Baidu is most ideal for capturing the overall market sentiment for at least two reasons. (i) After Google's exit in 2010, China's search engine market was almost monopolised by Baidu with its market share exceeding 70 percent in 2017.[1] There is no doubt that Baidu is the most comprehensive news channel that aggregates news from a wide range of over 500 news providers. Its comprehensiveness ensures that we capture all the

[1] http://gs.statcounter.com/search-engine-market-share/all/china.

available opinions, including misleading ones, to conduct an unbiased news sentiment analysis, yet focusing on any single news provider compromises this objective and distorts our sentiment measure towards the views of the news provider. (ii) More importantly, while there are studies using proprietary databases such as *Chinese Main Newspaper Database*, Chinese stock market is dominated by unsophisticated retail investors. For example, retail investors account for 75 percent of trading activities in our sample. Retail investors do not have access to these fee-payable news channels but can search Baidu at no financial costs. Therefore, the fee-payable news databases are well suited to disseminate the corporate views via corporate news releases rather than to aggregate market perceptions over corporate news; yet, it is the latter that eventually drives the stock price formation process and induces the probability of future stock price crash. Unsurprisingly, an emerging stream of literature uses search engine data to perform financial market research; the prominent examples include Da *et al.* (2011) and Choi and Varian (2012) using Google and Cai *et al.* (2015) and Shao *et al.* (2015) using Baidu.

Four measures of the probability of stock price crash used in this study are the negative skewness, up-down volatility of returns, the Value-at-Risk measure using a parametric approach and tail risk measure in Kelly and Jiang (2014). We find a significant negative relationship between positive media reports and probability of stock price crash, whereas the relationship between negative reports and probability of stock price crash is non-monotonic across these four probability of crash measures. The effects of the media coverage on the probability of stock price crash vary between SOEs and non-SOEs and between large and small firms.

Institutional ownership may be an important mediating factor that explains the observed relationship between media reports and probability of stock price crash. Mullainathan and Shleifer (2005) and Gurun and Butler (2012) show that media reports may be biased due to the demands of related parties including institutional investors. For example, poorly managed firms in the process of Initial Public Offering (IPO) tend to pay certain public relationship fees in exchange for either paid news or paid silence of the financial media (Fang, 2014). When it comes to corporate governance, institutional investors always face a trade-off between exit and voice. They can either dump the shares at a favourable price at the first sight of trouble or intervene business decisions. In China, institutional investors are less motivated to intervene company businesses given that these investors only held about 25 percent of all the shares by end of 2015 (Maug, 1998). On the other hand, the costs of opportunistic trading are low for institutional investors in China, where investor protection is weak. As a result, we anticipate these investors to be more opportunistic in China. They have incentives to manage media reports and materialise short-term trading profits.

We find consistent evidence that institutional investors manage media reports to earn excess returns, especially when they cannot fully participate in the corporate decisions of the listed firms. Further analysis suggests that one adverse effect of managing media reports is an increase in the probability of stock price crash. This is consistent with the notion that the abnormal increase in positive reports or the decrease in negative reports associated with institutional trading increases the expected probability of concealing true information from investors. Delayed true information release then leads to future stock price collapse. But we also find that such an effect weakens in firms with block institutions.

Our study contributes to the literature in the following ways. First, unlike previous papers that collect data of media reports mainly from the Chinese Main Newspaper Database, we use a Web crawler program to obtain media reports from the Internet as investors rely more on the Internet for information about listed firms, especially with respect to reports and attitudes of the media. Second, we distinguish among the sentiments of the various media reports and study the varied impacts of positive and negative reports on the probability of stock price crash. Third, we investigate the role of institutional investors in the relationship between media reports and the probability of stock price crash. Contrary to the conventional view that institutional investors are rational investors and help stabilise the stock market, our results suggest they capitalise transient profits by managing media report. The managed media reports inflate the probability of stock price crash as a consequence.

The remainder of the study is organised as follows. In Section 2, we describe the data, variables and methodology, while Section 3 presents the main empirical results and Section 4 concludes the study.

2. Data and methodology

2.1. Data sources

Our sample covers all listed firms from year 2012 to 2016. The split share structure reform was started in April 2005. The majority of firms completed the reform by the end of year 2009. Most of the previously non-tradable shares became tradable after 2 years' lock-up periods, that is by the end of year 2011. Our sample starts from 2012 to avoid the impact of the reform.

Being a major structural change in Chinese stock market, the split share structure reform period likely introduces two opposing effects on the probability of future stock price crash, making it hard to draw an affirmative conclusion about how media sentiment affects probability of price crash across time. The split share structure reform converted a substantial amount of non-tradable shares into tradable shares, resulting in significantly improved stock liquidity. On the one hand, increased stock liquidity may increase the probability of stock price crash if the managers in firms with liquid stocks tend to withhold bad news in fear of the company stock selling by transient traders (Chang *et al.*, 2017). On the other hand, improved liquidity attracts informed traders as they can camouflage their trading activities in liquid market environments, yet the presence of informed traders strengthens the monitoring of corporate management and increases the costs of withholding price-sensitive news from investors (Hou *et al.*, 2012). As such, the probability of stock price crash may decline following the split share structure reform. Another important reason to commence our sample in 2012 is the Google's exit in China's search engine market in 2010. Prior to 2011, it is unclear whether Baidu or Google, or both are the dominant news channel for investors especially for sophisticated institutional and foreign investors.

Firm-specific media reports are collected from the news headlines of the website news.baidu.com, which is the dominant news search engine in China. Baidu news collects news from over 500 sources and distributes them to readers without editing. In this sense, Baidu news can provide most comprehensive news which is least likely to be biased towards the attitudes of the news provider.

Our sample includes the A-share listed companies from year 2012 to year 2016. We collect the aggregate holdings of all institutional investors and other information from the *China Stock Market and Accounting Research* (CSMAR) database. We collect the in-depth institutional holdings data from WIND database. We exclude the financial and insurance firms, the ST/*ST[2] firms and firms with negative equity. We also exclude firms with <30 weekly observations per year (Xu *et al.*, 2013; Wang *et al.*, 2015). Our final sample consists of 12,114 firm-year observations.

2.2. Measurement of media sentiment

Quantifying the tone and sentiment of financial news reports is hard, not to mention the lack of reliable Chinese word dictionaries adapted to the realm of business. Although the most credible classifications could be to assign personnel to thoroughly read and analyse every news report (see such practices in Zhang *et al.* (2014), Cai *et al.*(2015) and Yi *et al.* (2017)), this approach is impossible to deploy when facing a sample of more than 4.5 million news reports from 2012 to 2016. As a result, this study employs two distinct estimation methods for the textual analysis of news sentiment. The first approach is motivated by the contention raised by Loughran and McDonald (2011) that 'a word categorization scheme derived for one discipline might not translate effectively into a discipline with its own dialect'. Strikingly, they find that almost three-fourths of the words identified as negative by the commonly used Harvard dictionary are words typically not considered negative in financial contexts on a large sample of corporate 10-K

[2] 'ST' is the abbreviation for 'Special Treatment'. Listed firms marked with 'ST' had negative net income in recent consecutive 2 years. Listed firms marked with '*ST' had negative net income in recent consecutive 3 years.

news releases. The novel contribution of our study is to create a Chinese word dictionary that is suited in the financial context. Specifically, we randomly select about 2,000 news reports that can be clearly classified as positive or negative reports. We read each of these reports and manually select Chinese words that reflect positive and negative tones. These words appear most frequently in these reports and are used to build our own word dictionary.[3] The news tone, either positive or negative, is then classified on basis of our self-defined dictionary.

Noting the arbitrary nature in the choices of words with the first approach, the second method is based on the technique of machine learning algorithms commonly used in the finance literature (Antweiler and Frank, 2004; Das and Chen, 2007). To execute this approach, we first select 2,000 headlines from our sample of more than 4.5 million news articles, making sure that half of the selected articles have a positive tone and the other half have a clear negative tone. We add these reports with clear tones to the text pool of the Natural Language Processing (NLP) tool. Next, the Bayesian method is applied to calculate the probability of a sentence to be a positive report after splitting the sentence into words. The news headlines with a probability larger than 70 percent are classified as positive reports, and those with a probability <10 percent are classified as negative reports. Those with a probability in between are viewed as neutral reports.

2.3. Probability of stock price crash

To ensure robustness, four alternative measures of probability of price crash are employed in this study. The first two measures are negative return skewness and up-down return volatility following Chen *et al.* (2001) and Xu *et al.* (2014). Their construction is elaborated as below:

First, we run the following regressions to obtain the idiosyncratic return of each stock in each week.

$$r_i = \alpha_i + \beta_1 r_{M,t-2} + \beta_2 r_{M,t-1} + \beta_3 r_{M,t} + \beta_4 r_{M,t+1} + \beta_5 r_{M,t+2} + \varepsilon_{i,t} \tag{1}$$

where $r_{i,t}$ is the stock return of firm i in week t and $r_{M,t}$ is the weighted average market return in week t. We calculate $W_{i,t} = \ln(1 + \varepsilon_{i,t})$ to represent the idiosyncratic return of stock i in week t.

[3] We establish a dictionary of positive words, including 'mairu', 'xianghao', 'zeng', 'licaichanpin', 'fenhong', 'jiangdichengben', 'shunli', 'jiedaodingdan', 'jingliuru', 'zhanlue', 'youhua', 'zhang', 'jisulashen', 'buzhu', 'youshi', 'chengzhang', 'huoyi', 'chuangxin', 'shengji', 'kuozhang',etc. The dictionary of negative words includes 'jianchi', 'pao', 'mengliedaya', 'guanlianjiaoyi', 'huaizhang', 'xiahua', 'susong', 'di', 'die', 'sun', 'danbao', 'cuo', 'weiyue', 'cizhi', 'chufa', 'jingliuchu', 'jiaoting', 'weigui', 'buzu', etc.

We then calculate the following two variables to measure the probability of a stock price crash.

$$NCSKEW_{i,t} = -\left[n(n-1)^{3/2}\sum W_{i,t}^3\right] / \left[(n-1)(n-2)\left(\sum W_{i,t}^2\right)^{3/2}\right] \quad (2)$$

$$DUVOL_{i,t} = \log\left\{\left[(n_u-1)\sum W_{i,t}^2\right] / \left[(n_d-1)\sum W_{i,t}^2\right]\right\} \quad (3)$$

where n equals the total number of trading weeks of stock i and n_u (n_d) is the number of trading weeks with weekly returns higher (lower) than the mean return of the year. We use the variables *NCSKEW* and *DUVOL* to represent the negative return skewness and the up-down return volatility. A higher *NCSKEW* or *DUVOL* indicates a greater probability of stock price crash.

The two measures were originally proposed by Chen *et al.* (2001), which are based on the theoretical model from Hong and Stein (2003). The market condition in China is consistent with the settings in Hong and Stein (2003). In Hong and Stein (2003), investors have different opinions and stock prices are more likely to represent the attitudes of optimistic investors due to the short-sale constraints. Chinese stock markets are relatively young and populated by individual investors. Investors are more likely to have diverse views on the fundamental value of listed firms. In addition, there are few means to take advantage of overvalued shares in China. Short-sale of stocks was strictly prohibited in China before October 2008 and is still partially restricted over our sample period.

The third measure is based on a parametric approach and formulated following Liang and Park (2010) as below:

$$VaR_t(\alpha,\tau) = -D_{r,t}^{-1}(\alpha). \quad (4)$$

Setting α at the 5 percent significance level and time horizon τ as 1 year, D denotes the cumulative distribution function of weekly stock return r_i at time t at the 95 percent confidence level over the 1-year time horizon.[4] The construction of the final measure, *Lambda*, follows Kelly and Jiang (2014) as below:

$$\lambda_t = \frac{1}{K_t}\sum_{k=1}^{K_t}\ln\frac{R_{k,t}}{u_t}. \quad (5)$$

$R_{k,t}$ is the kth weekly return that falls below an extreme value threshold u_t during year t, and K_t is the total number of such exceedances within year t. We also define u_t as the fifth percentile of the cross section each period.

[4] We would like to thank the anonymous reviewer for making this suggestion.

We also calculate quarterly *VaR*s and *Lambda*s using daily returns.

2.4. Firm-level variables

The variable of *INST* represents the proportion of institutional investors' ownership in a company at the year end. The institutions include funds, Qualified Foreign Institutional Investors, security brokerage firms, insurance companies, social insurance funds, trusts, financial companies and banks. For each firm in each year, *POSI_NUM* and *NEGA_NUM* are the number of positive and negative media reports classified according to our manually collected tone words dictionary. *POSI_NUM_NLP* and *NEGA_NUM_NLP* are for the reports classified with NLP method. *POSIMEDIA, NEGAMEDIA, POSI_NLP and NEGA_NLP* are their logarithm values.

We control for the common factors that are closely related to the volatility of returns, the corporate governance and the other features of listed firms. These control variables are defined as follows.

AR: annual average of the difference between the weekly stock return and the market return.
RETURN: annual stock return.
STATE: one for SOEs and zero for non-SOEs.
BM: book-to-market ratio at year end.
ILLIQ: annual average of 10^{10} times of the monthly stock return over the monthly trading volume (Amihud, 2002).
SIZE: logarithm of the total assets at year end.
VOLATILITY: annual standard deviation of weekly stock returns.
A_SHARE: percentage of tradable A-shares out of total shares.
CURR_VALUE: logarithm of the market capitalisation of all tradable shares at year end.
LEVERAGE: total liabilities over total assets at year end.
ROA: earnings before interest and tax over the average total assets.
NONTRD: proportion of non-tradable shares over the total number of shares outstanding at year end.
TANGIBLE: net fixed assets over total assets.
AGE: number of days from IPO date to the last day of the year.
RET: average weekly stock return of firm *i* in year *t*.
SIGMA: standard deviation of weekly idiosyncratic returns of firm *i* in year *t*.
OPTIMISM: analysts' optimistic bias.

$OPT_{i,j,t} = (FEPS_{i,j,t} - AEPS_{i,t})/\left|AEPS_{i,t-1}\right|$ where $FEPS_{i,j,t}$ is the forecast value of the earnings per share of firm i provided by analyst j in year t and $AEPS_{i,j,t}$ is the actual value of the earnings per share of firm *i* in year *t*. *OPTIMISM* is defined as the proportion of analysts with non-negative $OPT_{i,j,t}$.

ABACC $ABACC_{i,t-1} = (|DAC_{i,t-1}| + |\ DAC-\ i,\ t\ -\ 2\ | + |\ DAC_{i,t-3}|)/3$, where $|DAC|$ is the absolute value of discretionary accruals as calculated by the adjusted Jones model (Dechow *et al.*, 1995).

Table 1, which displays the descriptive statistics of the main variables of our firm-year observations, indicates that the media, on average, reports more positive news than negative news regarding listed firms and that the quantity of positive and negative reports is greatly dispersed. Furthermore, the maximum numbers of positive and negative reports are 445 and 389, respectively. It indicates that the media tends to herd to some specific firms for news. The proportion of institutional ownership also varies greatly across our sample, although individual investors still dominate the Chinese stock markets. On average, the sample firms have a low proportion of tangible assets and a high degree of analysts' optimistic bias. Approximately 63 percent of the firms are SOEs. Additionally, non-tradable shares account for 26.1 percent and tradable A-shares account for ~72.2 percent of all shares outstanding. Finally, it is determined that listed firms perform well, on average, with an average ROA of ~4.7 percent.

2.5. Methodology

2.5.1. Media reports and probability of stock price crash

We first run the basic model in Equation (6) to examine the impact of the number of positive reports on the probability of a stock price crash. The variable *CRASH* represents the four measures of the probability of stock price crash, *NCSKEW*, *DUVOL, VaR* and *Lambda*. In Equation (7), we add the variable $NEGAMEDIA^2_{i,t-1}$ to examine the nonlinear relationship between the quantity of negative reports and probability of stock price crash. In addition, we add the interaction terms between the numbers of positive and negative reports and the state dummy and size, $POSIMEDIA \times STATE$, $POSIMEDIA \times SIZE$, $NEGAMEDIA \times STATE$ and $NEGAMEDIA \times SIZE$, to examine the different roles of positive and negative reports in affecting the probability of stock price crash in SOEs versus non-SOEs and in large firms versus small firms. We control for industry effect and year fixed effects and run fixed effect panel data models.

$$
\begin{aligned}
CRASH_{i,t} = {} & \beta_0 + \beta_1 \cdot POSIMEDIA_{i,t-1} + \gamma_1 \cdot SIZE_{i,t-1} + \gamma_2 \cdot LEVERAGE_{i,t-1} \\
& + \gamma_3 \cdot ROA_{i,t-1} + \gamma_4 \cdot BM_{i,t-1} + \gamma_5 \cdot CRASH_{i,t-1} + \gamma_6 \cdot RET_{i,t-1} \\
& + \gamma_7 \cdot SIGMA_{i,t-1} + \gamma_8 \cdot OPTIMISM_{i,t-1} + \gamma_9 \cdot ABACC_{i,t-1} \\
& + \sum YEAR + \sum INDUSTRY + \varepsilon_{i,t}
\end{aligned}
\tag{6}
$$

Table 1
Descriptive statistics

Variables	No. of observations	Mean	Std.	*Min*	*Max*	Median	25%	75%
AR	12,099	0.004	0.015	−0.072	1.139	0.002	−0.002	0.0078
POSI_NUM	12,110	60.260	50.318	0	445	48	23	85
NEGA_NUM	12,110	25.370	26.176	0	389	18	9	34
POSI_NUM_NLP	12,104	129.597	110.918	0	645	96	46	184
NEGA_NUM_NLP	12,104	85.271	68.133	0	618	68	37	114
INST	11,351	0.253	0.243	0.000	0.944	0.162	0.040	0.433
AGE	12,098	7.776	1.082	0.690	9.122	8.053	7.168	8.653
RET	12,056	−0.001	0.012	−0.240	0.275	−0.003	−0.007	0.003
SIGMA	12,041	0.057	0.028	0.003	0.602	0.050	0.039	0.069
TANGIBLE	12,101	0.223	0.168	0.000	0.971	0.188	0.093	0.318
LEVERAGE	12,101	0.439	0.342	−0.195	13.397	0.420	0.248	0.602
SIZE	12,101	21.944	1.342	14.942	29.192	21.764	21.015	22.660
BM	10,662	0.958	1.029	0.001	15.473	0.623	0.367	1.134
ABACC	11,428	0.407	2.168	0.004	119.576	0.239	0.149	0.382
OPTIMISM	10,149	0.838	0.260	0.000	1.000	0.971	0.786	1
RETURN	11,340	0.283	0.657	−0.732	15.211	0.156	−0.136	0.534
CURR_VALUE	12,120	15.079	1.135	12.166	21.283	15.021	14.304	15.768
ROA	12,101	0.047	0.232	−1.558	20.788	0.039	0.014	0.072
NONTRD	12,101	0.261	0.281	0.000	0.957	0.155	0.000	0.500
A_SHARE	12,101	0.722	0.280	0.032	1.000	0.789	0.490	1.000
ILLIQ	12,120	0.621	13.435	−180.722	620.791	0.005	−0.386	0.211
VOLATILITY	12,080	0.073	0.089	0.010	5.269	0.059	0.048	0.084
STATE	11,798	0.630	0.460	0	1	1	0	1

This table reports the descriptive statistics of the sample firms. *AR* is the annual average of the difference between the weekly stock return and the market return. *POSI_NUM* and *NEGA_NUM* are the number of positive and negative media reports classified according to our manually collected tone words dictionary, covering each firm in each year. *POSI_NUM_NLP* and *NEGA_NUM_NLP* are for the reports classified with machine learning NLP method. *POSIMEDIA, NEGAMEDIA, POSI_NLP and NEGA_NLP* are their logarithm values. *INST* is the proportion of institutional investors' ownership at year end. *AGE* is the number of days from IPO date to the last day of the year. *RET* is the average weekly stock return of firm i in year t. *SIGMA* is the standard deviation of weekly idiosyncratic returns of firm i in year t. *TANGIBLE* is the net fixed assets over total assets. *LEVERAGE* is the total liabilities over the total assets at year end. *SIZE* is the logarithm of the total assets at year end. *BM* is the book-to-market ratio at year end. *ABACC* represents the absolute value of discretionary accruals calculated with adjusted Jones model. *OPTIMISM* is the analysts' optimistic bias. *RETURN* is the annual stock return. *CURR_VALUE* is the logarithm of the market capitalisation of all tradable shares at year end. *ROA* is the earnings before interest and tax over the average total assets. *NONTRD* is the proportion of non-tradable shares over the total number of outstanding shares at year end. *A_SHARE* is the percentage of the tradable A-shares in total shares. *ILLIQ* represents the illiquidity index. *VOLATILITY* is the annual standard deviation of weekly stock returns. *STATE* is one for SOEs, and zero for non-SOEs.

$$
\begin{aligned}
CRASH_{i,t} = {} & \beta_0 + \beta_1 \cdot NEGAMEDIA_{i,t-1} + \beta_2 \cdot NEGAMEDIA^2_{i,t-1} \\
& + \gamma_1 \cdot SIZE_{i,t-1} + \gamma_2 \cdot LEVERAGE_{i,t-1} + \gamma_3 \cdot ROA_{i,t-1} \\
& + \gamma_4 \cdot BM_{i,t-1} + \gamma_5 \cdot CRASH_{i,t-1} + \gamma_6 \cdot RET_{i,t-1} + \gamma_7 \cdot SIGMA_{i,t-1} \\
& + \gamma_8 \cdot OPTIMISM_{i,t-1} + \gamma_9 \cdot ABACC_{i,t-1} + \sum YEAR \\
& + \sum INDUSTRY + \varepsilon_{i,t}
\end{aligned} \tag{7}
$$

2.5.2. Institutional investors and active media management

We investigate the institutional investors' media management behaviour according to the intermediate effect test (Baron and Kenny, 1986). We assume that institutional investors can promote the future excess return of a stock by managing the media reports on that stock. In this study, the media reports act as the intermediate factor.

Specifically, we first run the following Equation (8) to examine the relationship between institutional ownership and future excess return. The intermediate effect test continues if institutional ownership is significantly associated with the future excess return.

$$
\begin{aligned}
AR_{i,t} = {} & \beta_0 + \beta_1 \cdot INST_{i,t-1} + \beta_2 \cdot CURRVALUE_{i,t-1} + \beta_3 \cdot BM_{i,t-1} \\
& + \beta_4 \cdot RETURN_{i,t-1} + \beta_5 \cdot ILLIQ_{i,t-1} + \beta_6 + \cdot \beta VOLATILITY_{i,t-1} \\
& + \beta_7 \cdot ASHARE_{i,t-1} + \beta_8 \cdot OPTIMISM_{i,t-1} + \beta_9 \cdot ABACC_{i,t-1} \\
& + \sum YEAR + \sum INDUSTRY + \varepsilon_{i,t}
\end{aligned} \tag{8}
$$

Second, we run Equation (9) to investigate the relationship between institutional ownership and the quantity of media reports in the next period. The intermediate effect test continues if institutional ownership is significantly associated with either the positive or negative media reports in the next period.

$$
\begin{aligned}
POSIMEDIA_{i,t} = {} & \gamma_0 + \gamma_1 \cdot INST_{i,t-1} + \gamma_2 \cdot SIZE_{i,t-1} + \gamma_3 \cdot LEVERAGE_{i,t-1} \\
& + \gamma_4 \cdot ROA_{i,t-1} + \gamma_5 \cdot NONTRD_{i,t-1} + \gamma_6 \cdot AGE_{i,t-1} \\
& + \gamma_7 \cdot TANGIBLE_{i,t-1} + \gamma_8 \cdot OPTIMISM_{i,t-1} \\
& + \gamma_9 \cdot ABACC_{i,t-1} + \sum YEAR + \sum INDUSTRY + \varepsilon_{i,t}
\end{aligned} \tag{9}
$$

Finally, we include both the quantity of media reports and the institutional ownership in Equation (10) to explain the future excess return. If both variables are significantly related to future excess return, there exists a partial intermediate effect. If institutional ownership is not related but the quantity of media reports is significantly related to the excess stock returns, there exists a complete intermediate effect, indicating that the relationship between institutional ownership and future excess returns is completely explained by the media reports. In both cases, we contend that the institutional investors manage the media reports to promote future excess returns.

$$\begin{aligned} AR_{i,t} = {} & \theta_0 + \theta_1 \cdot POSIMEDIA_{i,t} + \theta_2 \cdot INST_{i,t-1} + \theta_3 \cdot CURRVALUE_{i,t-1} \\ & + \theta_4 \cdot BM_{i,t-1} + \theta_5 \cdot RETURN_{i,t-1} + \theta_6 \cdot ILLIQ_{i,t-1} \\ & + \theta_7 \cdot VOLATILITY_{i,t-1} + \theta_8 \cdot ASHARE_{i,t-1} + \theta_9 \cdot OPTIMISM_{i,t-1} \\ & + \theta_{10} \cdot ABACC_{i,t-1} + \sum YEAR + \sum INDUSTRY + \varepsilon_{i,t} \end{aligned} \tag{10}$$

We compute the Sobel (1982) statistics (MacKinnon and Dwyer, 1993; MacKinnon *et al.*, 1995) to test the intermediate effect of the media reports in explaining the role of institutional investors in promoting future excess returns. The Sobel statistics is computed as the z-value which equals $a{*}b$/SQRT ($b^2{*}s_a{}^2 + a^2{*}s_b{}^2$), where a is the regression coefficient γ_1 for the association between institutional ownership and the quantity of media reports, s_a is the standard error of a, b is the coefficient θ_1 for the association between the quantity of media reports and the future excess return, and s_b is the standard error of b. We also use nonparametric bootstrapping procedure to do the Sobel test for robustness check.

2.5.3. Active media management and probability of stock price crash

We use the predicted values of the dependent variables in Equation (9), $POSIM\hat{E}DIA_t$ and $NEGA\hat{M}EDIA_t$, to represent the expected quantity of media reports after the institutional investors' media management. We then test the relationship between the predicted quantity of media reports and the probability of stock price crash. The relationship provides evidence as to whether the media management by institutional investors changes the probability of stock price crash.

3.. Empirical results

3.1. Attitudes of media reports and probability of stock price crash

Table 2 displays the results regarding the relationship between the attitudes of media reports and the probability of stock price crash. We include the square term of the quantity of media reports to study the potential nonlinear

Table 2
Panel regressions of the measures of media sentiment on the probability of stock price crash

Panel A: media sentiments measured with self-defined dictionary

	Dependent variables							
	1		2		3		4	
	$NCSKEW_t$	$DUVOL_t$	$NCSKEW_t$	$DUVOL_t$	$NCSKEW_t$	$DUVOL_t$	$NCSKEW_t$	$DUVOL_t$
$POSIMEDIA_{t-1}$	−0.094***	−0.018***	−0.023	0.007				
	(−5.3)	(−4.13)	(−0.38)	(0.47)				
$POSIMEDIA_{t-1}{}^2$			0.018	0.017				
			(1.02)	(0.77)				
$NEGAMEDIA_{t-1}$					0.081***	0.012***	−0.013**	−0.004**
					(5.37)	(3.22)	(−2.21)	(−2.28)
$NEGAMEDIA_{t-1}{}^2$							0.013**	0.002***
							(2.31)	(2.68)
$SIZE_{t-1}$	−0.085**	−0.045***	−0.090**	−0.044***	−0.096**	−0.045***	−0.103***	−0.044***
	(−2.18)	(−4.61)	(−2.31)	(−4.55)	(−2.41)	(−4.53)	(−2.56)	(−4.41)
$LEVERAGE_{t-1}$	0.205*	0.042	0.185	0.040	0.191	0.062	0.182	0.060
	(1.66)	(1.35)	(1.5)	(1.28)	(1.24)	(1.59)	(1.18)	(1.56)
ROA_{t-1}	−1.296***	−0.079	−1.231***	0.073	−1.523***	0.155*	−1.535***	−0.157*
	(−4.02)	(−0.97)	(−3.80)	(0.89)	(−4.71)	(1.91)	(−4.74)	(−1.93)
BM_{t-1}	−0.215***	−0.021***	−0.212***	−0.021***	−0.233***	−0.026***	−0.233***	−0.026***
	(−8.06)	(−3.15)	(−7.97)	(−3.11)	(−8.72)	(−3.9)	(−8.74)	(−3.91)
$CRASH_{t-1}$	−0.225***	−0.200***	−0.223***	−0.200***	−0.232***	−0.199***	−0.232***	−0.200***
	(−16.12)	(−15.97)	(−16.04)	(−15.96)	(−16.61)	(−15.93)	(−16.64)	(−15.94)
RET_{t-1}	−9.480***	−1.740***	−9.489***	−1.734***	−9.838***	1.881***	−9.932***	−1.895***
	(−5.57)	(−4.53)	(−5.58)	(−4.52)	(−5.72)	(4.85)	(−5.77)	(−4.87)
$SIGMA_{t-1}$	6.492***	1.637***	6.499***	1.637***	6.294***	1.596***	6.326***	1.600***

(continued)

Table 2 (continued)

Panel A: media sentiments measured with self-defined dictionary

	Dependent variables							
	1		2		3		4	
	$NCSKEW_t$	$DUVOL_t$	$NCSKEW_t$	$DUVOL_t$	$NCSKEW_t$	$DUVOL_t$	$NCSKEW_t$	$DUVOL_t$
	(8.68)	(8.73)	(8.70)	(8.74)	(8.4)	(8.51)	(8.44)	(8.52)
$OPTIMISM_{t-1}$	0.234***	0.011	0.237***	0.011	0.229***	0.013	0.229***	0.013
	(4.44)	(0.85)	(4.49)	(0.86)	(4.32)	(1.00)	(4.31)	(1.00)
$ABACC_{t-1}$	0.009	0.002	0.009	0.002	0.009	0.002	0.010	0.002
	(1.28)	(0.82)	(1.27)	(0.81)	(1.32)	(1.08)	(1.32)	(1.08)
Intercept	−2.238***	0.859***	−2.184***	0.864***	−2.359***	0.888***	−2.424***	0.881***
	(−2.71)	(4.16)	(−2.64)	(4.18)	(−2.79)	(4.2)	(−2.86)	(4.16)
Time effect	Controlled	Controlled	Controlled	Controlled	Controlled	Controlled	Controlled	Controlled
Industry effect	Controlled	Controlled	Controlled	Controlled	Controlled	Controlled	Controlled	Controlled
F	70.28***	44.93***	64.29***	40.9***	68.91***	44.56***	62.83***	40.55***
R^2 (within)	0.089	0.059	0.089	0.059	0.088	0.059	0.088	0.059
No. of observations	9,620	9,618	9,620	9,618	9,548	9,546	9,548	9,546

Panel B: media sentiments measured with machine learning NLP method

	Dependent variables							
	1		2		3		4	
	$NCSKEW_t$	$DUVOL_t$	$NCSKEW_t$	$DUVOL_t$	$NCSKEW_t$	$DUVOL_t$	$NCSKEW_t$	$DUVOL_t$
$POSI\_NLP_{t-1}$	−0.086***	−0.018***	−0.099	−0.0004				
	(−4.52)	(−3.87)	(−1.43)	(−0.02)				

(continued)

Table 2 (continued)

Panel B: media sentiments measured with machine learning NLP method

	Dependent variables							
	1		2		3		4	
	$NCSKEW_t$	$DUVOL_t$	$NCSKEW_t$	$DUVOL_t$	$NCSKEW_t$	$DUVOL_t$	$NCSKEW_t$	$DUVOL_t$
$POSI\_NLP^2_{t-1}$			0.024	0.002				
			(0.78)	(1.13)				
$NEGA\_NLP_{t-1}$					0.062***	0.013***	−0.136*	−0.025
					(3.28)	(2.72)	(−1.67)	(−1.21)
$NEGA\_NLPt$ - 1^2							0.026**	0.005*
							(2.49)	(1.89)
$SIZE_{t-1}$	−0.060	−0.049***	−0.060	−0.049***	−0.083**	−0.044***	−0.088**	−0.043***
	(−1.54)	(−5.06)	(−1.55)	(−5.06)	(−2.19)	(−4.69)	(−2.33)	(−4.57)
$LEVERAGE_{t-1}$	0.183	0.037	0.160	0.034	0.204*	0.043	0.195	0.041
	(1.49)	(1.2)	(1.31)	(1.12)	(1.67)	(1.39)	(1.6)	(1.34)
ROA_{t-1}	−1.414***	−0.121	−1.41***	−0.120	−1.528***	0.147*	−1.513***	−0.144*
	(−4.48)	(−1.53)	(−4.47)	(−1.52)	(−4.86)	(1.87)	(−4.81)	(−1.83)
BM_{t-1}	−0.219***	−0.022***	−0.214***	−0.021***	−0.229***	−0.024***	−0.230***	−0.024***
	(−8.23)	(−3.29)	(−8.05)	(−3.21)	(−8.65)	(−3.63)	(−8.67)	(−3.65)
$CRASH_{t-1}$	−0.231***	−0.204***	−0.231***	−0.204***	−0.233***	−0.203***	−0.232***	−0.203***
	(−16.68)	(−16.42)	(−16.69)	(−16.43)	(−16.84)	(−16.37)	(−16.78)	(−16.36)
RET_{t-1}	−9.013***	−1.672***	−9.040***	−1.674***	−9.064***	1.702***	−9.238***	−1.727***
	(−5.32)	(−4.38)	(−5.33)	(−4.39)	(−5.34)	(4.45)	(−5.44)	(−4.52)
$SIGMA_{t-1}$	5.972***	1.542***	5.817***	1.526***	5.761***	1.489***	5.656***	1.469***
	(8.08)	(8.34)	(7.86)	(8.23)	(7.79)	(8.04)	(7.63)	(7.92)
$OPTIMISM_{t-1}$	0.236***	0.013	0.238***	0.013	0.235***	0.013	0.235***	0.013
	(4.51)	(0.98)	(4.54)	(0.99)	(4.49)	(0.97)	(4.49)	(0.97)
$ABACC_{t-1}$	0.010	0.002	0.009	0.002	0.010	0.002	0.010	0.002
	(1.33)	(0.90)	(1.31)	(0.89)	(1.36)	(0.92)	(1.37)	(0.93)

(continued)

Table 2 (continued)

Panel B: media sentiments measured with machine learning NLP method

	Dependent variables							
	1		2		3		4	
	$NCSKEW_t$	$DUVOL_t$	$NCSKEW_t$	$DUVOL_t$	$NCSKEW_t$	$DUVOL_t$	$NCSKEW_t$	$DUVOL_t$
Intercept	−1.733**	0.923***	−1.426*	0.954***	−2.126***	0.85***	−1.905**	0.891***
	(−2.14)	(4.56)	(−1.74)	(4.67)	(−2.65)	(4.24)	(−2.36)	(4.42)
Time effect	Controlled	Controlled	Controlled	Controlled	Controlled	Controlled	Controlled	Controlled
Industry effect	Controlled	Controlled	Controlled	Controlled	Controlled	Controlled	Controlled	Controlled
F	70.26***	46.34***	64.63***	42.24***	69.24***	45.34***	63.55***	41.56***
R^2 (within)	0.087	0.059	0.088	0.060	0.089	0.058	0.087	0.059
No. of observations	9,756	9,754	9,756	9,754	9,772	9,770	9,772	9,770

Panel C: probability of stock price crash measured with VaR in annual frequency

	Dependent variables							
	1		2		3		4	
	VaR_t		VaR_t		VaR_t		VaR_t	
$POSI\_NLP_{t-1}$	−0.010***	−0.002**	−0.010**	−0.012***				
	(−8.2)	(−2.08)	(−2.48)	(−3.03)				
$POSI\_NLP^2_{t-1}$			0.000	0.001				
			(0.09)	(0.52)				
$NEGA\_NLP_{t-1}$					0.012***	0.006***	0.003	−0.0004
					(12.4)	(5.52)	(0.9)	(−0.14)
$NEGA\_NLP^2_{t-1}$							0.002***	0.001*
							(2.82)	(1.90)

(continued)

Table 2 (continued)

Panel C: probability of stock price crash measured with VaR in annual frequency

	Dependent variables							
	1		2		3		4	
	VaR_t		VaR_t		VaR_t		VaR_t	
$SIZE_{t-1}$	0.064***	0.072***	0.064***	0.072***	0.065***	0.072***	0.066***	0.072***
	(25.07)	(28.88)	(25.00)	(28.7)	(24.76)	(28.09)	(24.93)	(28.13)
$LEVERAGE_{t-1}$	0.023***	0.013*	0.023***	0.015*	0.029***	0.014	0.028***	0.013
	(2.77)	(1.67)	(2.77)	(1.85)	(2.86)	(1.38)	(2.75)	(1.31)
ROA_{t-1}	−0.168***	−0.181***	−0.168***	−0.176***	−0.142***	−0.176***	−0.141***	−0.175***
	(−7.87)	(−8.82)	(−7.82)	(−8.54)	(−6.68)	(−8.52)	(−6.61)	(−8.47)
BM_{t-1}	−0.045***	−0.047***	−0.045***	−0.047***	−0.046***	−0.047***	−0.046***	−0.047***
	(−25.52)	(−27.82)	(−25.5)	(−27.92)	(−26.2)	(−27.77)	(−26.25)	(−27.79)
VaR_{t-1}		−0.376***		−0.38***		−0.363***		−0.362***
		(−24.58)		(−24.71)		(−23.26)		(−23.16)
RET_{t-1}	−1.581***	−2.017***	−1.581***	−2.018***	−1.427***	−1.934***	−1.411***	−1.922***
	(−15.97)	(−20.86)	(−15.97)	(−20.87)	(−14.29)	(−19.58)	(−14.11)	(−19.42)
$SIGMA_{t-1}$	0.363***	1.346***	0.364***	1.357***	0.398***	1.323***	0.394***	1.317***
	(7.36)	(21.7)	(7.36)	(21.83)	(8.08)	(21.35)	(8.00)	(21.23)
$OPTIMISM_{t-1}$	−0.017***	−0.014***	−0.017***	−0.014***	−0.018***	−0.014***	−0.018***	−0.014***
	(−4.85)	(−4.16)	(−4.85)	(−4.22)	(−5.02)	(−4.22)	(−5.04)	(−4.24)
$ABACC_{t-1}$	0.000	−0.001	0.000	−0.001	0.000	−0.001	0.000	−0.001
	(1.04)	(−1.17)	(1.04)	(−1.16)	(1.01)	(−1.16)	(1.01)	(−1.15)
Intercept	−1.268***	−1.398***	−1.268***	−1.403***	−1.283***	−1.399***	−1.292***	−1.404***
	(−23.24)	(−26.55)	(−23.23)	(−26.64)	(−23.06)	(−25.96)	(−23.19)	(−26.03)
Time effect	Controlled	Controlled	Controlled	Controlled	Controlled	Controlled	Controlled	Controlled
Industry effect	Controlled	Controlled	Controlled	Controlled	Controlled	Controlled	Controlled	Controlled

(continued)

Table 2 (continued)

Panel C: probability of stock price crash measured with VaR in annual frequency

	Dependent variables							
	1		2		3		4	
	VaR_t		VaR_t		VaR_t		VaR_t	
F	275.29***	328.89***	247.73***	299.78***	283.45***	328.54***	256.15***	299.11***
R^2 (within)	0.256	0.313	0.256	0.314	0.263	0.315	0.2641	0.316
No. of observations	9622	9622	9622	9622	9550	9550	9550	9550

Panel D: Probability of crash measured with *Lambda* in annual frequency

	Dependent variables			
	1	2	3	4
	$Lambda\_Annual_t$	$Lambda\_Annual_t$	$Lambda\_Annual_t$	$Lambda\_Annual_t$
$POSI\_NLP_{t-1}$	−0.0065**	−0.0088		
	(−2.24)	(−0.70)		
$POSI\_NLP^2_{t-1}$		0.00002		
		(0.01)		
$NEGA\_NLP_{t-1}$			0.0089***	0.0175*
			(3.08)	(1.65)
$NEGA\_NLP^2_{t-1}$				0.0031**
				(2.34)
$SIZE_{t-1}$	0.0349***	0.0349***	0.0336***	0.0336***
	(5.90)	(5.91)	(5.80)	(5.79)
$LEVERAGE_{t-1}$	−0.0323*	−0.0294	−0.0338*	−0.0338*
	(−1.73)	(−1.58)	(−1.82)	(−1.81)

(continued)

Table 2 (continued)

Panel D: Probability of crash measured with *Lambda* in annual frequency

	Dependent variables			
	1 $Lambda\_Annual_t$	2 $Lambda\_Annual_t$	3 $Lambda\_Annual_t$	4 $Lambda\_Annual_t$
ROA_{t-1}	−0.2327***	−0.2322***	−0.2382***	−0.2382***
	(−4.83)	(−4.82)	(−4.97)	(−4.97)
BM_{t-1}	−0.0364***	−0.0371***	−0.0356***	−0.0356***
	(−8.98)	(−9.12)	(−8.82)	(−8.81)
$Lambda\_Annual_{t-1}$	−0.2149***	−0.2132***	−0.2124***	−0.2124***
	(−17.12)	(−16.95)	(−16.90)	(−16.88)
RET_{t-1}	0.5312**	0.5323**	0.5111**	0.5110**
	(2.30)	(2.30)	(2.21)	(2.21)
$SIGMA_{t-1}$	1.3307***	1.3069***	1.3073***	1.3073***
	(11.25)	(11.01)	(11.02)	(10.99)
$OPTIMISM_{t-1}$	−0.0039	−0.0041	−0.0038	−0.0038
	(−0.48)	(−0.52)	(−0.48)	(−0.48)
$ABACC_{t-1}$	0.0005	0.0005	0.0005	0.0005
	(0.43)	(0.46)	(0.44)	(0.44)
Intercept	−0.4644***	−0.5048***	−0.4276***	−0.4277***
	(−3.74)	(−4.02)	(−3.48)	(−3.46)
Time effect	Controlled	Controlled	Controlled	Controlled
Industry effect	Controlled	Controlled	Controlled	Controlled
F	92.66***	84.79***	93.46***	84.95***
R^2 (within)	0.1121	0.1127	0.1128	0.1128
No. of observations	9,761	9,761	9,775	9,775

Table 2 (continued)

Panel E: VaRs and Lambdas at quarterly frequency

	Dependent variables							
	1 VaR_Q_t	2 VaR_Q_t	3 VaR_Q_t	4 VaR_Q_t	1 $Lambda\_Q_t$	2 $Lambda\_Q_t$	3 $Lambda\_Q_t$	4 $Lambda\_Q_t$
$POSI\_NLP_{t-1}$	−0.0077*** (−13.39)	−0.001*** (−3.36)			−0.014*** (−8.21)	−0.0359*** (−5.70)		
$POSI\_NLP^2_{t-1}$		−0.0008 (−0.38)				0.0007 (0.82)		
$NEGA\_NLP_{t-1}$			0.0102*** (18.03)	0.003 (1.19)			0.0227*** (13.56)	−0.0329*** (−4.49)
$NEGA\_NLP^2_{t-1}$				0.002*** (5.45)				0.0074*** (7.79)
$SIZE_{t-1}$	−0.007*** (−6.12)	−0.007*** (−6.17)	−0.0084*** (−7.25)	−0.008*** (−7.36)	−0.064*** (−18.38)	−0.0638*** (−18.29)	−0.0628*** (−18.40)	−0.0610*** (−17.82)
$LEVERAGE_{t-1}$	−0.0073* (−1.92)	−0.006* (−1.69)	−0.0077** (−2.02)	−0.007* (−1.9)	0.037*** (3.26)	0.0307*** (2.72)	0.0386*** (3.44)	0.0367*** (3.27)
ROA_{t-1}	0.0242*** (2.50)	0.024*** (2.53)	0.0140 (1.46)	0.0151 (1.57)	0.105*** (3.66)	0.1028*** (3.61)	0.1271*** (4.49)	0.1231*** (4.35)
BM_{t-1}	0.0006 (0.70)	0.0004 (0.5)	0.0014* (1.79)	0.0015* (1.84)	0.014*** (5.90)	0.0152*** (6.39)	0.0125*** (5.28)	0.0123*** (5.22)
VaR_Q_{t-1}	−0.078*** (−14.54)	−0.079*** (−14.58)	−0.082*** (−15.24)	−0.083*** (−15.35)				
$Lambda\_Q_{t-1}$					−0.054*** (−9.95)	−0.0563*** (−10.30)	−0.0566*** (−10.37)	−0.0580*** (−10.63)
RET_{t-1}	−0.9592*** (−20.36)	−0.962*** (−20.41)	−1.001*** (−21.27)	−1.013*** (−21.52)	−4.289*** (−30.65)	−4.2801*** (−30.56)	−4.2126*** (−30.11)	−4.1684*** (−29.79)
$SIGMA_{t-1}$	0.9752*** (43.08)	0.970*** (42.76)	0.9421*** (41.66)	0.936*** (41.37)	2.457*** (36.91)	2.4986*** (37.46)	2.5390*** (38.10)	2.5671*** (38.50)

(continued)

Table 2 (continued)

Panel E: VaRs and Lambdas at quarterly frequency

	Dependent variables							
	1 VaR_Q_t	2 VaR_Q_t	3 VaR_Q_t	4 VaR_Q_t	1 $Lambda\_Q_t$	2 $Lambda\_Q_t$	3 $Lambda\_Q_t$	4 $Lambda\_Q_t$
$OPTIMISM_{t-1}$	0.0044***	0.004***	0.004***	0.0039***	−0.0015	−0.0006	−0.0003	−0.0001
	(2.78)	(2.70)	(2.49)	(2.46)	(−0.32)	(−0.12)	(−0.07)	(−0.03)
$ABACC_{t-1}$	0.0002	0.0002	0.0002	0.0002	−0.0007	−0.0007	−0.0006	−0.0006
	(0.99)	(1.02)	(0.87)	(0.86)	(−1.05)	(−1.11)	(−0.96)	(−0.95)
Intercept	0.2372***	0.227***	0.273***	0.259***	1.299***	1.3754***	1.2323***	1.2868***
	(9.35)	(9.06)	(11.09)	(10.53)	(17.69)	(18.59)	(16.98)	(17.66)
Time effect	Controlled	Controlled	Controlled	Controlled	Controlled	Controlled	Controlled	Controlled
Industry effect	Controlled	Controlled	Controlled	Controlled	Controlled	Controlled	Controlled	Controlled
F	250.34***	228.68***	266.67***	245.33***	168.82***	159.91***	181.24***	170.57***
R^2 (within)	0.0677	0.0681	0.0718	0.0726	0.0467	0.0486	0.0499	0.0516
No. of observations	36,868	36,868	36,914	36,914	36,890	36,890	36,936	36,936

In this table, we examine the relationship between the sentiment of media reports and the probability of stock price crash. Panels A and B use *NCSKEW* and *DUVOL* as proxies for the probability of stock price crash where Panel A measures media sentiment using self-defined dictionary and Panel B quantifies news sentiment using the NLP method. In Panel C, we use VaR as an alternative measure of the probability of price crash, while in Panel D, we use the tail risk measure, *Lambda*, as the measure of probability of crash. Panel E repeats the analysis reported in Panels C and D using quarterly VaR and tail risk measures. Industry and year fixed effects are also included. The *t*-values are presented in brackets. *, **, ***indicate significance at the 10, 5 and 1 percent levels.

relationship. In Panel A, we use NCSKEW and DUVOL to measure the probability of price crash and use the media sentiments classified with the dictionary method. In Panel B, we use the media sentiments classified with machine learning NLP method. In Panel C, we use VaR to measure the probability of stock price crash. In Panel A of Table 2, Model 1 indicates that the quantity of positive reports is significantly negatively related to the probability of stock price crash. As expected, positive media reports promote investors' confidence or reduce the information asymmetry of listed firms over the period of 1 year, thus reducing the probability of stock price crash. However, we cannot differentiate the two effects.

Turning to the negative news tones, the combined results in Models 3 and 4 of Panel A indicate that there is a significant U-shaped relationship between the quantity of negative reports and the probability of stock price crash. When small, the quantity of negative reports is negatively correlated with the probability of stock price crash. It indicates that more negative reports promote information quality more than they damage investors' confidence, thus reducing the probability of stock price crash when the quantity remains low. The degree of information asymmetry accumulates for those firms with fewer negative reports. However, we observe a significant positive relationship between the quantity of negative reports and the probability of stock price crash when the quantity is substantial.

One caveat of the news sentiment variables defined based on the dictionary method arises from the arbitrary choices of pessimistic and optimistic words selected in the dictionary. To alleviate this concern, we adopt more robust and better grounded measures of news tone and sentiment on basis of the NLP method following the steps outlined in Section 2. The unreported correlation coefficients of the *POSIMEDIA* and *NEGAMEDIA* computed using two different news sentiment classification methods are 0.79 and 0.70, respectively, confirming the reliability of the news sentiment captured using self-defined dictionary. In Panel B, with the new *POSIMEDIA* and *NEGAMEDIA* measures based on the NLP method, we continue to observe strong and consistent results that positive news sentiment is associated with a lower likelihood of future price crash while there is a non-monotonic relationship between negative news sentiment and stock price crash risk. Noting that the NLP method is widely used in the finance literature, we choose to use this method as the primary approach in categorising news tone and sentiment. Nevertheless, our results remain unchanged even if the news sentiment is defined according to our self-defined dictionary.

To ensure the robustness of our stock price crash measures, we employ two alternative price crash measures by varying the formulation method and frequency. In particular, Panel C introduces a new stock price crash risk measures, *VaR*, at an annual frequency from weekly stock returns and Panel D employs the tail risk as the proxy for probability of stock price crash, and reassuringly, the primary results are still unaffected.

With simple variants of Panels C and D, Panel E attempts to compute quarterly VaR and tail risk measure, *Lambda*, based on daily stock returns. It is clear from the results of estimating quarterly *VaR* and *Lambda* that our main findings remain unaffected. However, the quarterly *NCSKEW* and *DUVOL* measures cannot be succinctly calculated in that the construction of the two measures relies on estimating yearly regressions using weekly stock returns (Jin and Myers, 2006; Hutton *et al.*, 2009; Kim *et al.*, 2011; Xu *et al.*, 2014; Zhang *et al.*, 2016) while quarterly regressions on weekly observations would not be meaningful due to the lack of statistical power. Running quarterly regressions on daily stock returns would also be difficult to justify for three reasons: (i) daily stock returns would introduce unnecessary transitory noises that stem from market fluctuations such as a number of 'black swan' (i.e. flash crash) events; (ii) it is plausible to anticipate news, especially bad news, to travel slowly in an emerging stock market like China where capital markets are less efficient (Morck *et al.*, 2000), and daily time horizon in this case would be too short for firm-level news to be transmitted into stock prices; and (iii) investors in China are not allowed for same-day trading, thus casting doubt on the real impact of the price crash risk derived from daily returns.

We then include the interaction terms between the state dummy and the quantity of media reports and between the firm size and the quantity of the media reports in our regressions on the probability of stock price crash. Table 3 presents the results. Model 1shows that the interaction term between the state dummy and the quantity of positive reports is significantly positively related to the probability of stock price crash. It indicates that the probability of stock price crash of SOEs is less sensitive to positive media reports than that of non-SOEs. The signs of the significant coefficients of the interaction terms in Model 2 are consistent with our findings presented in Table 2. When a large quantity of negative media reports is published, the probability of a stock price crash increases significantly more for SOEs than for non-SOEs, indicating that the probability of stock price crash of SOEs is more sensitive to negative media reports than that of non-SOEs.

SOEs legally belong to all Chinese citizens although they are actually controlled by the government or government agencies. Investors pay more attention to SOEs than non-SOEs although they might have equal performance. In addition, SOEs have invisible political power which makes them capable of releasing more positive media reports and hiding negative reports. According to You *et al.* (2017), the number of firms covered by state-owned media is three times of those covered by market-oriented media. These media providers are more likely to help SOEs report more positive news and hide more negative news. Investors become less sensitive to the positive media reports, while more sensitive to the negative reports of SOEs than to those of non-SOEs. Thus, the positive media reports of SOEs will reduce the probability of crash to a lesser extent. But the U-shaped relationship between negative reports and probability of crash in SOEs is more prominent than that in non-SOEs.

Table 3
Different roles of positive and negative reports on stock price crash risk in SOEs versus non-SOEs and in large firms versus small firms

	Dependent variables							
	1		2		3		4	
	$NCSKEW_t$	$DUVOL_t$	$NCSKEW_t$	$DUVOL_t$	$NCSKEW_t$	$DUVOL_t$	$NCSKEW_t$	$DUVOL_t$
$POSIMEDIA_{t-1}$	−0.083***	−0.017***			−0.588**	−0.086**		
	(−4.01)	(−3.26)			(−2.01)	(−2.17)		
$NEGAMEDIA_{t-1}$			−0.065	−0.018			−0.326**	−0.419*
			(−1.1)	(−1.19)			(−2.33)	(−1.68)
$NEGAMEDIA^2_{t-1}$			0.002	0.001			0.081**	0.060*
			(0.18)	(0.37)			(2.48)	(1.74)
$POSIMEDIA_{t-1} \times STATE_{t-1}$	0.037**	0.005***						
	(2.42)	(2.70)						
$NEGAMEDIA_{t-1} \times STATE_{t-1}$			−0.148*	−0.046**				
			(−1.76)	(−2.18)				
$NEGAMEDIA^2_{t-1} \times STATE_{t-1}$			0.040**	0.008*				
			(2.43)	(1.93)				
$POSIMEDIA_{t-1} \times SIZE_{t-1}$					−0.031**	−0.003*		
					(−2.33)	(−1.93)		
$NEGAMEDIA_{t-1} \times SIZE_{t-1}$							−0.017*	−0.019*
							(−1.73)	(−1.67)
$NEGAMEDIA^2_{t-1} \times SIZE_{t-1}$							0.004*	0.003**
							(1.92)	(2.46)
$SIZE_{t-1}$	−0.086**	−0.043***	−0.098**	−0.043***	−0.031	−0.033**	−0.024	−0.016
	(−2.16)	(−4.37)	(−2.41)	(−4.18)	(−0.49)	(−2.11)	(−0.32)	(−0.85)
$LEVERAGE_{t-1}$	0.218*	0.041	0.200	0.059	0.173	0.045	0.170	0.062
	(1.75)	(1.32)	(1.28)	(1.52)	(1.4)	(1.44)	(1.1)	(1.6)
ROA_{t-1}	−1.407***	−0.092	−1.646***	−0.173**	−1.330***	−0.076	−1.552***	−0.150*

(continued)

Table 3 (continued)

	Dependent variables							
	1		2		3		4	
	$NCSKEW_t$	$DUVOL_t$	$NCSKEW_t$	$DUVOL_t$	$NCSKEW_t$	$DUVOL_t$	$NCSKEW_t$	$DUVOL_t$
	(−4.21)	(−1.11)	(−4.92)	(−2.06)	(−4.12)	(−0.93)	(−4.8)	(−1.85)
BM_{t-1}	−0.213***	−0.021***	−0.234***	−0.027***	−0.203***	−0.022***	−0.232***	−0.026***
	(−7.94)	(−3.11)	(−8.73)	(−3.95)	(−7.53)	(−3.26)	(−8.71)	(−3.98)
$CRASG_{t-1}$	−0.223***	−0.199***	−0.232***	−0.199***	−0.222***	−0.201***	−0.232***	−0.200***
	(−15.78)	(−15.66)	(−16.35)	(−15.68)	(−15.95)	(−15.99)	(−16.64)	(−15.96)
RET_{t-1}	−9.432***	−1.771***	−9.552***	−1.913***	−9.751***	−1.721***	−9.80***	−1.880***
	(−5.46)	(−4.55)	(−5.47)	(−4.85)	(−5.72)	(−4.48)	(−5.69)	(−4.83)
$SIGMA_{t-1}$	6.534***	1.662***	6.372***	1.637***	6.423***	1.643***	6.264***	1.605***
	(8.58)	(8.71)	(8.35)	(8.57)	(8.59)	(8.76)	(8.36)	(8.55)
$OPTIMISM_{t-1}$	0.248***	0.015	0.241***	0.017	0.235***	0.011	0.227***	0.013
	(4.63)	(1.13)	(4.46)	(1.28)	(4.46)	(0.84)	(4.28)	(0.99)
$ABACC_{t-1}$	0.009	0.001	0.009	0.002	0.009	0.002	0.010	0.002
	(1.26)	(0.8)	(1.31)	(1.06)	(1.27)	(0.82)	(1.35)	(1.07)
Intercept	−2.293***	0.817***	−2.335***	0.852***	0.287	0.609*	−0.736	0.271
	(−2.72)	(3.88)	(−2.71)	(3.95)	(0.21)	(1.79)	(−0.45)	(0.66)
Time effect	Controlled	Controlled	Controlled	Controlled	Controlled	Controlled	Controlled	Controlled
Industry effect	Controlled	Controlled	Controlled	Controlled	Controlled	Controlled	Controlled	Controlled
F	62.42***	39.93***	52.32***	33.85***	64.43***	40.92***	54***	34.55***
R^2 (within)	0.089	0.059	0.089	0.059	0.089	0.059	0.089	0.059
No. of observations	9,376	9,374	9,305	9,303	9,620	9,618	9,548	9,546

In this table, we add the interaction terms between the numbers of positive and negative reports and the state dummy and firm size, *POSIMEDIA*STATE*, *POSIMEDIA*SIZE*, *NEGAMEDIA*STATE* and *NEGAMEDIA*SIZE*. We examine the different roles of positive and negative reports in affecting stock price crash risk in SOEs versus non-SOEs and in large firms versus small firms. The *t*-values are presented in brackets. *, ** and *** indicate significance at the 10, 5 and 1 percent levels, respectively.

Models 3 and 4 of Table 3 reveal that the signs of the significant coefficients of the interaction terms between both the quantity of positive and negative reports and the size of the firm are the same as those of the quantity of positive and negative media reports. Large firms attract more attention from investors and media. Investors are more sensitive to both positive reports and negative reports regarding large firms than small firms. Thus, the relationship between both positive and negative reports and the probability of crash is more significant in large firms.

3.2. Institutional investors, media management and excess returns

In this section, we report the results regarding the media management by institutional investors and the effect of such management on their excess returns. Table 4 reports the results of the intermediate effect test as discussed in Section 2.5. Panel A presents the results for the sample of all institutional investors. Panel B presents the results about the bootstrapping Sobel statistics. Panel C presents the results for the two subsamples with or without block institutional holdings.

In Panel A, the regression result for Equation (8) indicates that institutional ownership is significantly positively related to the excess returns of the next year. The stocks of firms with more institutional ownership perform better in the following year, and this better performance benefits these institutional investors if they buy and hold the stock. Regression results for Equation (9) indicate that institutional ownership is also positively related to the quantity of positive reports and negatively related to the quantity of negative reports of the next year after controlling for other firm-specific features. Institutional investors tend to facilitate the publication of positive reports and restrain the disclosure of negative reports.

Equation (10) includes both institutional ownership and the quantity of positive/negative media reports as independent variables. Regression results indicate that both the quantity of positive/negative media reports and institutional ownership are significantly related to excess returns. In addition, the significant Sobel statistics indicate the existence of a significant partial intermediate effect among the group of regressions. In general, institutional investors tend to facilitate the publication of significantly more positive reports or fewer negative reports on the firms in which they invest. Both activities significantly promote the returns of these firms in which they invest. Accordingly, we find some evidence to support the media management activities of the institutional investors.

The Sobel test has the problem with the product of coefficients when the sample size is different in the models used to estimate the intermediate effects. We use nonparametric bootstrapping procedure to do the Sobel test for robustness check. The bootstrapping is accomplished by taking 1,000 random samples of size n (where n is the original sample size) from the data, sampling

Table 4
Media sentiment, stock returns and institutional ownership

Panel A: Full sample			
Equation (8)			
Dependent variable: AR_t			
$INST_{t-1}$	0.014*** (28.08)	No. of observations	9,431
$CONTROLS_t$	Controlled	F	171.95***
INDUSTRY/YEAR	Controlled	R^2 (within)	0.1796
Equation (9)			
Dependent variable: $POSI\_NLP_t$		Dependent variable: $NEGA\_NLP_t$	
$INST_{t-1}$	0.305*** (8.00)	$INST_{t-1}$	−0.284*** (−5.92)
$CONTROLS_t$	Controlled	$CONTROLS_t$	Controlled
INDUSTRY/YEAR	Controlled	INDUSTRY/YEAR	Controlled
No. of observations	9,621	No. of observations	9,775
F	76.74***	F	22.62***
R^2 (within)	0.0872	R^2 (within)	0.0269
Equation (10)			
Dependent variable: AR_t		Dependent variable: AR_t	
$POSI\_NLP_t$	0.0019*** (11.1)	$NEGA\_NLP_t$	−0.0036*** (−18.67)
$INST_{t-1}$	0.0086*** (14.75)	$INST_{t-1}$	0.0095*** (16.65)
$CONTROLS_t$	Controlled	$CONTROLS_t$	Controlled
IND/YEAR	Controlled	IND/YEAR	Controlled
No. of observations	9621	No. of observations	9775
F	54.19***	F	91.51***
R^2 (within)	0.0699	R^2 (within)	0.1105
Sobel test	6.49***	Sobel test	5.64***

Panel B: bootstrapping statistics for Sobel test

	a*b in the original sample	Simulated random distribution of a*b			
		Mean	95% percentile	97.5% percentile	99.5% percentile
Positive reports	0.00012563***	−0.0000863	−0.000012	0.00000033	0.00002924
Negative reports	0.00043945*	0.00033187	0.0004356	0.00045295	0.00049625

(continued)

Table 4 (continued)

Panel C:Subsamples with and without block institutional ownership

Subsamples	Dependent variable / Independent variable	AR_t / $INST_{t-1}$	$POSI\_NLP_t$ / $INST_{t-1}$	AR_t: $POSI\_NLP_t$	AR_t: $INST_{t-1}$	Sobel test z-value	Positive reports' mediating effect
Block	I	0.0283*** (8.08)	0.0198 (0.11)	0.0013 (1.02)	0.0283*** (8.05)	0.11	No
Non-block	II	0.0159*** (19.38)	0.0918* (1.85)	0.0015*** (5.57)	0.0158** (19.26)	1.76*	Partial mediation effect

Subsamples	Dependent variable / Independent variable	AR_t / $INST_{t-1}$	$NEGA\_NLP_t$ / $INST_{t-1}$	AR_t: $NEGA\_NLP_t$	AR_t: $INST_{t-1}$	Sobel test z-value	Negative reports' mediating effect
Block	I	0.0283*** (8.08)	−0.2745 (−1.30)	−0.00003 (−0.03)	0.0283*** (8.03)	0.03	No
Non-block	II	0.0159*** (19.38)	−0.1774*** (−3.51)	−0.0025*** (−9.32)	0.0154*** (19.00)	3.28***	Partial mediation effect

In this table, we present the results about the intermediate effect test regarding the media management behaviour of institutional investors. The intermediate effect test and the Sobel statistics are described in Section 2.5. Panel A of Table 4 presents the results of the intermediate effect test for the sample of all institutional investors. Panel B presents the bootstrap test of the Sobel statistics. Panel C presents the results for subsamples of firms with and without block institutional ownership. The t-values are presented in brackets. *, **, ***indicate significance at the 10, 5 and 1 percent levels, respectively.

with replacement and computing the intermediate effect, $a*b$, in each sample. We then have a simulated distribution of randomised $a*b$. We list the value of $a*b$ estimated with the original sample and the mean value, the 95%, 97.5% and 99.5% percentiles of the simulated random distribution in Panel B of Table 4. We can see that for the positive reports, $a*b$ is significantly larger than the 99.5% percentile of the simulated random distribution. For the negative reports, the value is significantly larger than the 95% percentile of the distribution. Our main results are consistent with the parametric test. In this table, the value $a*b$ is estimated by running pooled OLS with our original sample and different from the value in Panel A of Table 4 in which we run two-way fixed models. However, the main results do not change.

Different types of investors exhibit very different incentives to manage media. When trading off the costs and benefits of monitoring, block institutional investors are often long-term investors who are more likely to intervene business decisions and prevent managers from self-entrenchment (Holderness, 2003; Harford *et al.*, 2017; Bena *et al.*, 2017). But, non-block investors tend to be transient investors due to the low costs of their selling of company shares and thus may engage in opportunistic activities for immediate trading profits through manipulating short-term media sentiment. With the in-depth institutional holdings data from WIND database, we are able to distinguish between block and non-block institutional investors. We classify the full sample into two subsamples according to whether or not the firm is held by block institutional investors, where an institutional investor is deemed as a block investor if its equity holding of the company greater than 3 percent.[5] We then repeat the regressions of Equations (8), (9) and (10) on these two subsamples. The results in Panel C of Table 4 show that the institutional investors' media management behaviour is significant for firms held by non-block investors but insignificant for firms held by block investors. On the assumption that non-block investors tend to have a short investment horizon and weak incentives to monitor, our evidence suggests that transient institutional investors behave opportunistically by distorting short-term media sentiment in order to achieve immediate trading profits.

3.3. Media management and probability of stock price crash

As described in Section 2.5, we run a two-stage least squares regression in this section. In the first stage, we run the regression of Equation (9) to obtain the predicted value of the dependent variable that represents the expected quantity of media reports after the media management by institutional investors. In the second stage, we use the predicted value in the first stage as the independent variable and run regressions on the probability of stock price crash to test the effect of the media reports managed by institutional investors on the probability of stock price crash. Panel A of Table 5 presents the results for the full sample, and Panel B of Table 5 presents the results for the block and non-block ownership subsamples in Section 3.2.

The results in Panel A of Table 5 indicate that the predicted quantities of both positive and negative reports are significantly negatively related to the probability of stock price crash. More positive media reports occurring after the media management by institutional investors reduce the probability of stock price crash which is consistent with the findings presented in Section 3.1. However, fewer negative media reports due to the media management of institutional investors significantly increase the probability of stock price crash, which is in sharp contrast to the results of the benchmark regressions presented in Section 3.1. Institutional investors hide negative news and restrain negative

[5] The selection of the 3 percent cut-off for block ownership is unconventional in the literature wherein the 5 percent cut-off is commonly used. This is because Chinese stock market is still underdeveloping and predominant of retail and small institutional investors. The institutional ownership exceeding 5 percent is scarce in our sample. An average institution holds about 0.2 percent of the company. Only four firms have an equity ownership exceeding 5 percent. We therefore have to lower the threshold for a meaningful subsample analysis.

Table 5
Impact of media management behaviour on stock price crash risk

Panel A: Full sample

	Dependent variable			
	$NCSKEW_t$	$DUVOL_t$	$NCSKEW_t$	$DUVOL_t$
$\widehat{POSI\_NLP}_t$	−0.4745***	−0.1175***		
	(−2.90)	(−2.86)		
$\widehat{NEGA\_NLP}_t$			−2.6604***	−0.4868***
			(−13.25)	(−9.62)
$SIZE_{t-1}$	−0.1077	−0.0916***	−0.4093***	−0.1336***
	(−1.37)	(−4.63)	(−7.63)	(−9.90)
$LEVERAGE_{t-1}$	−0.0005	−0.0052	−0.1247	−0.0203
	(0.01)	(−0.16)	(−0.99)	(−0.64)
ROA_{t-1}	0.5710	−0.1077	−0.1506	−0.1704*
	(1.15)	(−0.87)	(−0.43)	(−1.92)
BM_{t-1}	−0.2409***	−0.0260***	−0.3211***	−0.0400***
	(−8.90)	(−3.84)	(−11.72)	(−5.80)
$CRASH_{t-1}$	−0.2345***	−0.2052***	−0.2305***	−0.2057***
	(−16.63)	(−16.05)	(−16.56)	(−16.24)
RET_{t-1}	8.3684***	1.5861***	3.5117**	0.7215*
	(4.71)	(3.95)	(1.99)	(1.78)
$SIGMA_{t-1}$	−5.7833***	−1.5308***	−3.0912***	−1.0427***
	(−7.71)	(−8.14)	(−4.02)	(−5.38)
$OPTIMISM_{t-1}$	0.3377***	0.0366**	0.5315***	0.0656***
	(5.32)	(2.30)	(9.31)	(4.56)
$ABACC_{t-1}$	0.0113	0.0023	0.0174**	0.0034*
	(1.54)	(1.23)	(2.39)	(1.83)

(continued)

Table 5 (continued)

Panel A: Full sample

	Dependent variable			
	$NCSKEW_t$	$DUVOL_t$	$NCSKEW_t$	$DUVOL_t$
Intercept	0.5163	1.4994***	17.6493***	4.4642***
	(0.43)	(4.95)	(10.41)	(10.46)
Time effect	Controlled	Controlled	Controlled	Controlled
Industry effect	Controlled	Controlled	Controlled	Controlled
F	69.92***	44.22***	88.25***	53.16***
R^2 (within)	0.0897	0.0587	0.1106	0.0697
No. of observations	9,495	9,493	9,495	9,493

Panel B: Subsamples with and without block institutional ownership

	Dependent variable							
	Block				Non-block			
	$NCSKEW_t$	$DUVOL_t$	$NCSKEW_t$	$DUVOL_t$	$NCSKEW_t$	$DUVOL_t$	$NCSKEW_t$	$DUVOL_t$
$\widehat{POSI\_NLP}_t$	−0.5597	−0.1207			−0.4203*	−0.169***		
	(−1.52)	(−1.19)			(−1.76)	(−2.80)		
$\widehat{NEGA\_NLP}_t$			−0.0524	−0.0393			−2.578***	−0.497***
			(−1.32)	(−0.34)			(−9.26)	(−7.04)
$SIZE_{t-1}$	0.2045	−0.0894	−0.3502	−0.1458**	−0.0234	−0.105***	−0.136**	−0.0768***
	(0.89)	(−1.39)	(−1.62)	(−2.39)	(−0.21)	(−3.80)	(−2.21)	(−4.92)
$LEVERAGE_{t-1}$	−0.0369	−0.0713	−0.3068	−0.1186	−0.2178	−0.1063*	−0.694***	−0.161***
	(−0.05)	(−0.37)	(−0.46)	(−0.62)	(−0.93)	(1.79)	(−3.03)	(−2.76)
ROA_{t-1}	2.4449	0.0932	−0.8611	−0.3153	1.1083	−0.4551**	−0.4015	−0.397***
	(1.35)	(0.18)	(−0.56)	(−0.73)	(1.38)	(−2.25)	(−0.78)	(−3.03)

(continued)

Table 5 (continued)

Panel B: Subsamples with and without block institutional ownership

	Dependent variable							
	Block				Non-block			
	$NCSKEW_t$	$DUVOL_t$	$NCSKEW_t$	$DUVOL_t$	$NCSKEW_t$	$DUVOL_t$	$NCSKEW_t$	$DUVOL_t$
BM_{t-1}	−1.148***	−0.1356*	−1.1267***	−0.1301	−0.191***	−0.0162	−0.275***	−0.031***
	(−4.00)	(−1.67)	(−3.96)	(−1.60)	(−4.14)	(−1.38)	(−5.91)	(−2.64)
$CRASH_{t-1}$	−0.434***	−0.484***	−0.4754***	−0.489***	−0.281***	−0.238***	−0.277***	−0.236***
	(−6.05)	(−7.23)	(−6.61)	(−7.28)	(−14.49)	(−13.33)	(−14.5)	(−13.31)
RET_{t-1}	15.826**	4.2822**	3.9548	3.6400*	6.885***	0.6639	1.8197	−0.1200
	(2.07)	(2.34)	(0.48)	(1.83)	(3.01)	(1.27)	(0.79)	(−0.23)
$SIGMA_{t-1}$	−3.136	−1.4580	−1.5409	−1.4698	−5.561***	−1.435***	−3.391***	−0.985***
	(−0.78)	(−1.28)	(−0.38)	(−1.27)	(−5.30)	(−5.42)	(−3.2)	(−3.66)
$OPTIMISM_{t-1}$	0.2042	−0.0897	0.9631***	−0.0188	0.314***	0.047*	0.636***	0.067***
	(0.62)	(−0.97)	(2.91)	(−0.2)	(2.80)	(1.67)	(7.1)	(2.95)
$ABACC_{t-1}$	−1.389*	−0.2732	0.3552	−0.1117	0.0108	0.0023	0.020**	0.0034*
	(−1.69)	(−1.19)	(0.43)	(−0.48)	(1.34)	(1.15)	(2.53)	(1.68)
Intercept	−0.8856	2.6638**	10.9699**	3.4455**	−0.2911	1.614***	11.543***	3.316***
	(−0.22)	(2.38)	(2.13)	(2.37)	(−0.18)	(3.94)	(6.29)	(7.14)
Time effect	Controlled	Controlled	Controlled	Controlled	Controlled	Controlled	Controlled	Controlled
Industry effect	Controlled	Controlled	Controlled	Controlled	Controlled	Controlled	Controlled	Controlled
F	14.21***	9.8***	14.78***	9.61***	42.18***	26.38***	51.65***	30.84***
R^2 (within)	0.3651	0.284	0.3743	0.2802	0.1034	0.0673	0.1237	0.0778
No. of observations	694	694	694	694	5,168	5,167	5,168	5,167

We use the predicted value of the dependent variable in Equation (9) to represent the expected quantity of media reports after institutional investors' media management. We then test the relationship between the predicted quantity of media reports and the stock price crash risk. Panel A of Table 5 presents the results for the full sample. Panel B of Table 5 presents the results for the subsamples of firms with and without block institutional ownership. The t-values are presented in brackets. *, ** and *** indicate significance at the 10, 5 and 1 percent levels, respectively.

reports on the firms in which they invest. This behaviour aggravates the information asymmetry of these firms and increases the probability of price crash of their stocks. The regulatory authority of China should be made aware of these phenomena. Results in Panel B of Table 5 suggest that the relationship between the predicted quantity of positive/negative reports and the probability of stock price crash is again driven by firms held by non-block and potentially transient investors, corroborating our findings in Panel C of Table 4.

3.4. Robustness tests

We perform a battery of robustness tests as follows. First, instead of using the number of media reports, we use the ratio of the number of positive/negative reports to the total number of reports. This is to mitigate the concern of our results being driven by firm size and reputation. Second, we change the fixed effects panel regressions to the ordinary least squares method and use clustered standard errors in the analysis. Third, given that the extreme turbulence in Chinese stock market in 2015, we re-estimate the tests excluding the firm-years in 2015. We continue to document the negative relationship between the positive reports and the probability of stock price crash and the U-shaped relationship between the negative reports across all of these robustness tests. The results of robustness tests are available upon request.

4. Conclusion

Media reports play an important role in the external governing of listed firms. Media reports may also be manipulated by institutional investors for their own purpose, that is, to earn excess returns. Different types of media reports affect the probability of a stock price crash in different ways. The media management by institutional investors also influences the probability of stock price crash. We find that positive reports help reduce the probability of stock price crash and that there is a U-shaped relationship between the quantity of negative reports and the probability of stock price crash. In general, negative reports are more likely to aggravate the probability of stock price crash in Chinese stock markets. We also find evidence to support the institutional investors' media management behaviour. Further analysis reveals that such media management behaviour stems from the firms held only by transient institutional investors.

The rapid development of the Internet and mobile media has greatly accelerated the speed of information transmission. Equity investors should learn to discriminate news from media reports. Accordingly, regulatory authorities should pay sufficient attention to those institutional investors who exhibit short-term investment strategies, as well as to their media management behaviour. These institutional investors' media management behaviours aggravate the information asymmetry, cause a high probability of stock price crash, damage the benefits of retail investors and impede market efficiency.

References

Amihud, Y., 2002, Illiquidity and stock returns: cross-section and time-series effects, *Journal of Financial Markets* 5(1), 31–56.

Antweiler, W., and M. Frank, 2004, Is all that talk just noise? The information content of internet stock message boards, *Journal of Finance* 59(3), 1259–1294.

Baron, R. M., and D. A. Kenny, 1986, The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations, *Journal of Personality and Social Psychology* 51(6), 1173–1182.

Bena, J., M. A. Ferreira, P. Matos, and P. Pires, 2017, Are foreign investors locusts? The long-term effects of foreign institutional ownership, *Journal of Financial Economics* 126(1), 122–146.

Cai, G., Z. Shao, and X. Xu, 2015, Is there collusion between enterprises and media: indirect evidence from media reports of Chinese Listed Companies, *Management World (In Chinese)* 7, 158–169.

Chan, W. S., 2003, Stock price reaction to news and no-news: drift and reversal after headlines, *Journal of Financial Economics* 70(2), 223–260.

Chang, X., Y. Chen, and L. Zolotoy, 2017, Stock liquidity and stock price crash risk, *Journal of Financial and Quantitative Analysis* 52(4), 1605–1637.

Chen, J., H. Hong, and J. C. Stein, 2001, Forecasting crashes: trading volume, past returns, and conditional skewness in stock prices, *Journal of Financial Economics* 61 (3), 345–381.

Choi, H., and H. Varian, 2012, Predicting the present with google trends, *Economic Record* 88, 2–9.

Da, Z., J. Engelberg, and P. Gao, 2011, In search of attention, *Journal of Finance* 66, 1461–1499.

Das, S. R., and M. Y. Chen, 2007, Yahoo! for Amazon: sentiment extraction from small talk on the web, *Management Science* 53(9), 1375–1388.

Dechow, P. M., R. G. Sloan, and A. P. Sweeney, 1995, Detecting earnings management, *Accounting Review* 70(2), 193–225.

Dyck, A., and L. Zingales, 2004, Control premiums and the effectiveness of corporate governance systems, *Journal of Applied Corporate Finance* 16(3), 51–72.

Dyck, A., N. Volchkova, and L. Zingales, 2008, The Corporate governance role of the media: evidence from Russia, *Journal of Finance* 63(3), 1093–1135.

Fang, J., 2014, Information disclosure, governance environment and media alienation: based on the initial discovery of IPO paid silence, *Management World (In Chinese)* 11, 95–104.

Gurun, U. G., and A. W. Butler, 2012, Don't believe the hype: local media slant, local advertising, and firm value, *Journal of Finance* 67(2), 561–598.

Harford, J., A. Kecskes, and S. Mansi, 2017, Do long-term investors improve corporate decision making? *Journal of Corporate Finance*. https://doi.org/10.1016/j.jcorpfin.2017.09.022

Holderness, C. G., 2003, A survey of blockholders and corporate control, *Economic Policy Review* 9, 51–64.

Hong, H., and J. C. Stein, 2003, Differences of opinion, short-Sales constraints, and market crashes, *Review of Financial Studies* 16(2), 487–525.

Hou, W., J. Kuo, and E. Lee, 2012, The impact of state ownership on share price informativeness: the case of the split share structure reform in China, *British Accounting Review* 44(4), 248–261.
Hutton, A. P., A. J. Marcus, and H. Tehranian, 2009, Opaque financial reports, R^2, and crash risk, *Journal of Financial Economics* 94(1), 67–86.
Jin, L., and S. C. Myers, 2006, R^2 around the world: new theory and new tests, *Journal of Financial Economics* 79(2), 257–292.
Joe, J. R., H. Louis, and D. Robinson, 2009, Managers' and investors' responses to media exposure of board ineffectiveness, *Journal of Financial & Quantitative Analysis* 44(3), 579–605.
Kelly, B., and H. Jiang, 2014, Tail risk and asset prices, *Review of Financial Studies* 27 (10), 2841–2871.
Kim, J. B., Y. Li, and L. Zhang, 2010, Corporate tax avoidance and stock price crash risk: firm-level analysis, *Journal of Financial Economics* 100(3), 639–662.
Kim, J. B., Y. Li, and L. Zhang, 2011, CFOs versus CEOs: equity incentives and crashes, *Journal of Financial Economics* 101(3), 713–730.
Kong, D., S. Liu, and Q. Ying, 2013, The role of media in corporate behavior, *Management World (In Chinese)* 7, 145–162.
Kothari, S. P., S. Shu, and P. D. Wysocki, 2009, Do managers withhold bad news?, *Journal of Accounting Research* 47(1), 241–276.
Liang, B., and H. Park, 2010, Predicting hedge fund failure: a comparison of risk measures, *Journal of Financial and Quantitative Analysis* 45(1), 199–222.
Loughran, T., and B. McDonald, 2011, When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks, *Journal of Finance* 66(1), 35–65.
MacKinnon, D. P., and J. H. Dwyer, 1993, Estimating mediated effects in prevention studies, *Evaluation Review* 17, 144–158.
MacKinnon, D. P., G. Warsi, and J. H. Dwyer, 1995, A simulation study of mediated effect measures, *Multivariate Behavioral Research* 30, 41–62.
Maug, E., 1998, Large shareholders as monitors: Is there a trade-off between liquidity and control, *Journal of Finance* 53(1), 65–98.
Morck, R., B. Yeung, and W. Yu, 2000, The information content of stock markets: why do emerging markets have synchronous stock price movements?, *Journal of Financial Economics* 58(1–2), 215–260.
Mullainathan, S., and A. Shleifer, 2005, The market for news, *American Economic Review* 95(4), 1031–1053.
Shao, X., M. He, P. Jiang, Y. Xue, and J. Liao, 2015, The management of media public relation, investor sentiment and security offerings, *Journal of Financial Research (In Chinese)* 9, 190–206.
Sobel, M. E., 1982, Asymptotic confidence intervals for indirect effects in structural equation models, *Sociological Methodology* 13, 290–312.
Tetlock, P. C., 2007, Giving content to investor sentiment: the role of media in the stock market, *Journal of Finance* 62(3), 1139–1168.
Tetlock, P., M. Saar-Tsechansky, and S. Macskassy, 2008, More than words: quantifying language to measure firms' fundamentals, *Journal of Finance* 63(3), 1437–1467.
Wang, H., F. Cao, and K. Ye, 2015, Monitoring or tunneling? The proportion of the proportion held by the big shareholders and the risk of the crash of the stock price, *Management World (In Chinese)* 2, 45–57.
Xu, N., S. Yu, and Z. Yi, 2013, The herd behavior of institutional investors and the risk of stock price crash, *Management World (In Chinese)* 7, 31–43.
Xu, N., X. Li, Q. Yuan, and K. C. Chan, 2014, Excess perks and stock price crash risk: evidence from China, *Journal of Corporate Finance* 25(2), 419–434.

Yi, Z., Z. Pan, N. Mao, and X. Li, 2017, Strategic media disclosure and wealth transfer: evidence from managers' stock selling, *Economic Research (In Chinese)* 4, 166–180.
You, J., B. Zhang, and L. Zhang, 2017, Who captures the power of the pen?, *Review of Financial Studies* 31(1), 43–96.
Zhang, S., L. Yu, Y. Li, and X. Chen, 2014, Media coverage and PEAD in China's A-share stock market: a limited attention perspective, *Journal of Financial Research (In Chinese)* 7, 154–170.
Zhang, M., L. Xie, and H. Xu, 2016, Corporate philanthropy and stock price crash risk: evidence from China, *Journal of Business Ethics* 139(3), 595–617.