Standing in the Limelight:

Sophisticated Active Attention and Managerial Bad News Hoarding*

TAO CHEN, *Nanyang Technological University*†

JIMMY CHENGYUAN QU, *Nanyang Technological University*

January 2021

* We gratefully acknowledge helpful comments from Simba Chang, David Feldman (the discussant), Jun-koo Kang, Qifei Zhu, and seminar and conference participants at the Nanyang Technological University, and the 2020 Australasian Finance and Banking Conference. We thank James Ryans for sharing data on EDGAR searching volume and Elisabeth Kempf for sharing data on shareholder distraction. We remain responsible for any errors and omissions.

† Corresponding author. Tel: +65-67905785. Email: jtchen@ntu.edu.sg. Address: Nanyang Business School, Nanyang Technological University, S3-B1A-08, 50 Nanyang Avenue, Singapore 639798.

Standing in the Limelight:

Sophisticated Active Attention and Managerial Bad News Hoarding

ABSTRACT

This paper examines the effect of active attention from sophisticated market participants on managerial bad news hoarding. Using EDGAR searching volume, we first find a positive impact of sophisticated active attention on managerial bad news hoarding measured by stock price crash risk, which can not be explained by general attention from Google and firms' information supply. Further evidence from option market reaction, management guidance, financial reporting quality, and accounting conservatism further confirms managers' tendency to hide bad news under greater pressure from sophisticated attention. Two natural experiments are implemented to provide a causal inference. By providing systematic evidence on the impact of sophisticated active attention on managerial bad news hoarding, this paper sheds light on the pressure effect of external attention on managers' disclosure strategies that have been underexplored in prior literature.

Keywords: sophisticated active attention, EDGAR searching volume, stock price crash risk, bad news hoarding, pressure effect

1 Introduction

How will managers react when they are standing in the limelight? A recent case in point seems to provide some evidence. In February 2020, the short-seller Muddy Waters Research cited an anonymous report questioning the financials of Luckin Coffee Inc. (Ticker: LK), China's most popular chain of coffeehouses listed in NASDAQ, which aroused great attention in the market. At that time, managers of Luckin announced that the report was misleading and denied all the negative information.[1] However, two months later, in April 2020, Luckin was found to have committed fabricating sales transactions, leading to a crash of about 5 billion U.S. dollars in market capitalization. Similar cases are not rare in the financial market nowadays. In a more widely known case, Jeffrey Skilling, the CEO of Enron, the "America's Most Innovative Company" named by *Fortune*,[2] chose to hide the bad news when the recession hit in 2000 made Enron a great loss, finally leading to a huge crash in the stock market and the subsequent bankruptcy. Based on these cases, it is natural to ask: Are famous firms more likely to hide bad news because of their higher exposure to the public? Does this external pressure push managers to hide bad news? As the information transformation becomes increasingly efficient these days, it is critical to understand the interplay between managers and the attention of market participants, especially from financial experts. This paper intends to provide systematic evidence on the impact of active attention from sophisticated market participants on managerial bad news hoarding behavior.

Extant studies suggest the effects of catering motivation on corporate financial reporting decisions (Healy and Palepu 2001; Verrecchia 2001; Graham et al. 2005; Dichev et al. 2013). In order to cater to stakeholders by hitting earnings benchmarks, managers choose to inflate earnings

[1] Luckin Coffee Responds to Anonymous Report Containing Misleading and False Allegations, *GlobeNewswire,* February 3, 2020.
[2] The World's Most Admired Companies, *Fortune*, October 2, 2000.

even at the expense of firms' long-term value (Graham et al. 2005). As such, when hit by bad news, managers may choose to hide it in the belief that the negative effect will be offset by good news in the future (Verrechia 2001; Graham et al. 2005; Kothari et al. 2009). Since the cost of releasing bad news is higher when firms are more exposed to the public (Baloria and Heese 2018), releasing bad news under external stakeholders' attention increases managers' probability of job loss (Francis et al. 2004). Therefore, we expect managers to hide bad news when they receive greater market attention. Because sophisticated market participants are generally equipped with substantial financial knowledge and more potent in financial markets, their attention should have a direct impact on managers' bad news hoarding. In this paper, we first focus on managers' bad news hoarding behavior measured by firms' stock price crash risk (e.g., Jin and Myers 2006).[3] Firms with greater negative skewness returns imply a higher possibility of negative stock crashes, indicating the sudden release of bad news. Accordingly, since managers' value in the managerial labor market is highly associated with firms' performance (Fama 1980), pressure from meeting earnings benchmarks and managers' career concerns are expected to drive managers to hide bad news (Graham et al. 2005; Baginski et al. 2018), thus increasing stock price crash risk (Jin and Myers 2006).

A recently-emerged measure, EDGAR searching volume gives us a direct measure to gauge the effect of sophisticated active attention. There are several benefits of using EDGAR searching data as sophisticated active attention. First, different from market-based attention measures, measures based on EDGAR searching information provide us with a direct way to investigate the attention of sophisticated market participants from non-market sources, which alleviates the spurious regression problem (e.g., Loghran and McDonald 2017; Ryans 2017). Second, different

[3] Based on the skewness of stock returns, stock price crash risk is validated by Piotroski et al. (2015) and widely used in the literature.

from information content from the supply side, EDGAR searching data reflect the active attention of sophisticated market participants from the demand side, which conveys more information about markets' attitudes (e.g., Drake et al. 2015; Li and Sun 2018; Drake et al. 2017; Chen, Cohen et al. 2020; Chen, Kelly et al. 2020; Iliev et al. forthcoming). Further, apart from active attention reflected as information acquisition of processed data in media platforms (e.g., Drake et al. 2012; Ben-Rephael et al. 2017), the active attention of sophisticated market participants proxied by EDGAR searching volume is greater effective to firms because EDGAR viewers are more skilled and experienced to affect managerial behaviors (Loghran and McDonald 2017).[4]

To test our hypothesis, in the baseline regressions, we follow the literature and measure stock price crash risk by two widely-used variables, down-up volatility ratio (*DUVOL)* and negative conditional skewness (*NCKSEW*) (e.g., Hutton et al. 2009, Kim et al. 2011, Chen et al. 2017). Measures of sophisticated attention in most of our analysis are based on the calculation as Ryans (2017), Drake et al. (2015), and Loughran and McDonald (2017). Using a large sample of U.S. listed firms for the period 2003-2015, we first find a significantly positive impact of sophisticated active attention on managerial bad news hoarding measured by stock price crash risk, which is consistent with our prediction. This evidence is robust to different model specifications and alternative measures of stock price crash risk and sophisticated active attention. To eliminate the possibility that the effect comes from attention through mass media and firms' information supply, we further control for general attention measured by Google searching volume and the number of filings in EDGAR in each firm-year. As a result, we find that the impact of sophisticated attention is not driven by general attention and firms' information supply. By testing the effect of

[4] In this paper, we investigate active attention from all sophisticated market participants in order to provide evidence on the total effect of sophisticated attention under the interaction among different types of sophisticated market participants in information acquisition (e.g., Chen, Kelly et al. 2020).

sophisticated active attention from different types of filings, we find that the active attention to filings with irregular releasing time and shorter releasing cycle has greater effect on stock price crash risk, suggesting that attention to unexpected information has a stronger effect on managerial behavior.

Further, we implement several heterogeneity tests based on different types of filings and different firm-level characteristics. Also, we find that the effect of sophisticated attention on stock price crash risk is stronger for firms with more investment opportunities, more financial constraints, and firms failing to beat analysts' expectations in the current year, suggesting that managers with greater catering motivations are more likely to hide bad news under greater attention pressure.

To further validate the use of stock price crash risk in measuring managerial bad new hoarding, we employ measures of option market reaction, management guidance, accounting information quality, and accounting conservatism as our dependent variables. Consistent with our hypothesis, we find that firms under greater sophisticated active attention are more likely to have higher implied volatility smirk, fewer bad news disclosures in management guidance, lower financial reporting quality, and higher accounting conservatism, which further extends our understanding of how sophisticated active attention reshapes managerial bad new hoarding behavior. Further analysis on executive compensation shows that higher option compensation and lower non-equity incentive plan compensation can attenuate the effect of sophisticated active attention, suggesting that managers with more long-term incentives and fewer short-term incentives are less likely to hold bad news when standing in the limelight.

Although the positive relation between sophisticated active attention and stock price crash risk is robust to alternative measures and model specifications, the potential omitted variable and reverse causality problems are not entirely eliminated. We alleviate these concerns by implementing two quasi-natural experiments. The first quasi-natural experiment is based on the Tier 3 adoption of eXtensible Business Reporting Language (XBRL) in 2011. According to prior works, the adoption of XBRL reduces the cost of information processing (Kim et al. 2019) and increases information acquisition through EDGAR (Chen and Zhou 2018). Under a difference-in-differences research design following Fang et al. (2012), we employ the regulatory shock of XBRL adoption and document a significant increase in managerial bad news hoarding measured by stock price crash risk in the treatment group relative to control group, which is consistent with our results in OLS regressions. A placebo test using the bootstrap simulation method validates this experiment. The second quasi-natural experiment follows Kempf et al.'s (2017) shareholder distraction shock based on the argument that institutional investors allocate attention across firms subject to a limited constraint as attention is a scarce resource (Kahneman and Tversky 1979; Fitch et al., 2015). Building on Kempf et al.'s (2017) attention distraction measure, we observe a negative effect of shareholder distraction on stock price crash risk, which lends further supports to the baseline results. By implementing two quasi-natural experiments, we find supporting evidence that the impact of sophisticated active attention on stock price crash risk is causal.

This paper contributes to the literature in three ways. First, this paper adds to the literature on managers' bad news disclosures. According to prior works, pressure from both managerial labor markets and financial markets pushes managers to hide bad news (Fama 1980; Gibbons and Murphy 1992; DeFond and Park 1999; Nagar 1999; Fee and Hadlock 2000; Kothari et al. 2009; Baginski et al. 2018; Bao et al. 2018). Since firms failing to meet external expectations will be severely punished by financial markets and leads to market value destroy and increased cost of capital (e.g., Burgstahler 1997; Skinner and Sloan 2002), releasing bad news at an incorrect time is highly associated with a higher probability of future job loss (Francis et al. 2004). This motivation is further confirmed by surveys on executives (Graham et al. 2005; Dichev et al. 2013). In this paper, we provide a direct evidence on managers' bad news disclosure strategies under greater pressure from sophisticated outsiders. In addition, this paper is also related to a boarder literature on managers' disclosure strategies (e.g., Healy and Wahlen 1999; Healy and Palepu 2001; Verrecchia 2001).

Second, this paper extends the literature on stock price crash risk by investigating the effect of sophisticated active attention on managerial bad news hoarding measured by stock price crash risk. Prior literature investigates the determinants of stock price crash risk like financial analysts, stock market, accounting decisions, managers' features, and institutional features (e.g., Kim et al. 2011; Kim et al. 2016; Kim and Zhang 2016; DeFond 2015; Kim et al. 2018; Chang et al. 2018). Our work is one of the first works to provide systematic evidence on the impact of active attention on firms' future crash risk. Consistent with the literature on stock price crash risk (e.g., Jin and Myers 2006; Hutton et al. 2009; Kim et al. 2011), we find evidence to support our conjecture that managers inflate earnings and hold bad news when receiving greater sophisticated attention, thus increasing stock crash risk. Furthermore, different from prior studies testing the monitoring effect of outsiders on stock price crash risk, our findings provide an attention pressure mechanism that outsiders can exert on managerial bad news hoarding. Different from Baloria and Heese (2018), who document a positive effect of media slant on managerial bad news suppressing proxied by stock price crash risk, our work examines the attention from the demand side of information flow as well as from a more professional part of market participants. Using novel measures of active attention, we complement Baloria and Heese's (2018) findings by showing that the exposure to sophisticated market participants motivates managers' bad news hoarding behavior.

Third, we contribute to the literature on investor attention and information acquisition by showing a dark side of attention from sophisticated market participants. The information role of investor attention has been documented in both theoretical and empirical works (e.g., Hirshleifer and Teoh 2003; Peng 2005; Peng 2005; Peng and Xiong 2006; Hirshleifer, Lim, and Teoh 2009). Using EDGAR searching volume, many recent works investigate the impact of information acquisition (active attention) on capital markets and corporate actions (e.g., Drake et al. 2012; Drake et al. 2015; Bozanic et al. 2017; Li et al. 2018; Chen, Cohen et al. 2020; Chen, Kelly et al. 2020; Iliev et al. forthcoming). However, few papers discuss the pressure effect of active attention that may distort managerial behavior (e.g., Graham et al. 2005). Different from literature showing that investor attention can alleviate information asymmetry in the market (e.g., Drake et al. 2012; Drake et al. 2015) and discipline managers' behavior (Iliev et al. forthcoming), our paper adds to this strand of literature by showing that, at a higher cost of releasing bad news when firms are more exposed to sophisticated market participants, managers are more likely to hide bad news under greater sophisticated active attention. What is more, our work also adds to the literature of shareholder distraction originated from Kempf et al. (2017). According to Kempf et al. (2017), shareholder distraction leads to greater managerial opportunism, which destroys firm value. By contrast, our paper shows that less attention pressure can decrease managers' incentives to hold bad news and reduce the probability of a sudden crash risk, which further extends our understanding of shareholder distraction.

The remainder of this paper is organized as follows. Section 2 reviews related literature and develops the hypotheses. Section 3 defines the main variables, source of our sample, and summary statistics. Section 4 reports the main results on the effect of sophisticated attention on stock price crash risk. Section 5 shows the results of two quasi-natural experiments. Section 6 concludes.

2 Related literature and empirical prediction

In theoretical and survey works, catering to stakeholders by hitting earnings benchmarks is the most important factor considered by managers when they choose their disclosure strategies (Healy and Wahlen 1999; Healy and Palepu 2001; Graham et al. 2003; Dichev et al. 2013). Failing to meet external expectations, firms will be severely punished by stock markets in the form of destroying market value, and by debt markets through increasing cost of capital (e.g., Burgstahler 1997; Skinner and Sloan 2002). Therefore, due to career and reputation concerns, managers are more opportunistic to hide negative news (e.g., DeFond and Park 1999; Nagar 1999; Fee and Hadlock 2000; Kothari et al. 2009). As a result, managers under greater pressure are more likely to hide bad news and accumulate it until being detected, thus creating a higher probability of crash risks (Jin and Myers 2006).

Investor attention has been proved to have a significant impact on the capital market.[5] In recent literature, EDGAR searching volume data are used in various studies focusing on the effect of information acquisition from sophisticated market participants. In studies on the general effect of information acquisition on EDGAR, Li and Sun (2018) document that the investor can get an abnormal return of 8% that does not reverse in the long run by constructing a zero investment portfolio based on the attention in EDGAR. Ryans (2018) finds that EDGAR downloads of firms' comment letters can predict future earnings, probability of restatement, and the rate of future write-downs. Bauguess et al. (2013) find that pre-IPO firms with more page viewers experience higher

[5] Many theoretical works discuss the importance of investor attention to market under-reaction and argue that attention from investors leads to the slow price adjustment (e.g., Hirshleifer and Teoh 2003, Peng 2005, Peng 2005, Peng and Xiong 2006). In empirical studies, anomalies like post-earnings announcement drift (Hirshleifer, Lim, and Teoh 2009), the slow reaction to earnings news on Fridays (Dellavigna and Pollet 2009), the 52-week high puzzle (Li and Yu 2012), and the under-reaction to liquid stocks (Bali et al. 2014) are discussed as the consequences of investor inattention.

negative price revision. Drake et al. (2015) find that information acquisition is highly correlated with corporate events (e.g., restatements, earnings announcements, and stock performance) and the information environment. Drake et al. (2017) use EDGAR searching volume as a measure of investor attention to investigate the comovement of investor attention and its consequences. All of these papers agree that information acquisition from EDGAR alleviates information asymmetry and increases market efficiency. Since page viewers can be identified in EDGAR searching data through request IP, many other papers also investigate the information acquisition behavior of various types of market participants and find the consequences of investors' information acquisition in different types. For example, in research on institutional investors, Drake et al. (2020) document that stock performance is associated with information acquisition more from institutional investors than retail investors. Iliev et al. (forthcoming) observe the monitoring role of investors from their attention through EDGAR. For auditors, Drake et al. (2019) find that auditors are more likely to resort to disclosure benchmarking and learn from other auditors through their financial filings when facing cases with higher reporting complexity, financial reporting risk, and litigation risk. From the perspective of governments, Bozanic et al. (2017) investigate the information acquisition of IRS; Li et al. (2018) examine the searching from the Federal Reserve. These studies together prove the usefulness of EDGAR from the perspective of market participants. However, few works document how attention through EDGAR can reshape managers' disclosure strategies. We try to fill this gap by testing the impact of attention through EDGAR on managerial bad news hoarding measured by stock price crash risk.

Based on the arguments that corporate financial reporting strategy is a tradeoff between firms' cost and benefits (Healy and Wahlen 1999; Graham et al. 2005) and that the greater attention from sophisticated market participants increases the cost of bad news releasing (Baloria and Heese

2018), we expect managers to hide bad news when they receive greater sophisticated attention and release them when they are less exposed to the sophisticated market participants. Drawing upon the aforementioned arguments, we predict a positive impact of attention from sophisticated market participants on managerial bad news hoarding and propose the following hypothesis.

HYPOTHESIS 1. *Active attention from sophisticated market participants motivates managerial bad news hoarding, all else being equal.*

If this hypothesis is correct, we would observe an increase of future stock price crash risk under greater sophisticated active attention. Furthermore, we would observe consistent results that other measures proxied for managerial bad news are also affected by sophisticated active attention in the same direction.

Since the prediction of the positive relation between sophisticated active attention and managerial bad news hoarding is based on the catering motivation of managers, the impact of sophisticated active attention on managerial bad news hoarding is expected to be stronger for firms with greater catering motivation because managers of these firms are more unwilling to lose their reputation and firms' interests when bad news comes. This argument leads to our second hypothesis.

HYPOTHESIS 2. *The positive impact of sophisticated active attention on managerial bad news hoarding is stronger for firms with greater catering motivation, all else being equal.*

Empirical findings that are consistent with Hypothesis 2 lend further support to catering motivation in Hypothesis 1. In contrast, if the impact of sophisticated active attention on managerial bad news hoarding is not driven by managers' catering motivation, we will not observe evidence that is consistent with Hypothesis 2.

3 Data and summary statistics

Data sources

Our sample comes from multiple sources. Firm-level financial data come from the COMPUSTAT database. Stock price and return data come from the Center for Research in Security Prices (CRSP). Institutional holdings data come from Thomson Reuters Institutional Holdings (13f). Analyst coverage data come from Institutional Brokers Estimate System (I\B\E\S). The EDGAR searching volume data come from James Ryans' EDGAR Log File Data.[6] Information on SEC filings come from EDGAR. In the heterogeneity test, executive compensation data are from Execucomp. To provide further evidence, we also use option data from Option Metrics, management guidance data from I\B\E\S Guidance. Following most extant literature on stock price crash risk, we only include observations that satisfy the following criteria: (1) Book equity is positive; (2) Year-end stock price is above one U.S. dollar; (3) At least 26 observations are available in CRSP weekly data for each firm-year; (4) Variables used in the research are available; (5) Each firm should at least have 2-year consecutive observations. What is more, firms in financial industry (SIC codes 6000-6999) and utilities industry (SIC codes 4900-4999) are excluded from the sample. Finally, our sample includes firm-years that meet our requirement during the period 2003-2015 when variables of EDGAR searching volume are available, and the page requests are completely recorded over the years.[7] All continuous variables are winsorized at 1^{st} and 99^{th} percentiles to alleviate the potential disturbance from outliers. Finally, our sample consists of

[6] This data are available in James Ryans' personal website and are arranged in several zip files according to calendar years. In each file, firms' CIK number, file accession key, the date of file being viewed, and the number of page view are included. Noted that only files viewed by non-robot viewers are recorded in these files, eliminating the noise from Internet crawlers that are generally regarded to have no searching preference.

[7] In Ryans' files, we find few observations in 2016. In order to avoid potential selection problem in our analysis, we choose not to use those observations. The results are similar if those observations are included.

21,617 observations for 2003-2015.

Measuring firm-specific crash risk

According to prior literature (e.g., Hutton et al. 2009), in order to isolate firm-specific risk from common risk, we first estimate the following model based on CRSP weekly return data in each firm-year:[8]

$$r_{i,t} = \alpha_i + \beta_1 r_{m,i,t-2} + \beta_2 r_{m,i,t-1} + \beta_3 r_{m,i,t} + \beta_4 r_{m,i,t+1} + \beta_4 r_{m,i,t+2} + \varepsilon_{i,t} \quad (1)$$

where $r_{i,t}$ is a firm's weekly return; $r_{m,i,t}$ is the weekly value-weighted market return. According to Dimson (1979) and prior studies on crash risk, we include the lead and lag terms of market returns to eliminate the impact of nonsynchronous trading. To estimate Equation (1), we require that at least 26 weeks of stock return be available for each firm-year. Following extant literature on crash risk, we estimate firm-specific weekly return $W_{i,t}$ as the logarithm of 1 plus the residual term of Equation (1), namely $W_{i,t} = ln(1 + \varepsilon_{i,t})$.

Two measures of stock price crash risk are used in baseline regressions. The first one is the negative conditional skewness of return (*NCSKEW*) developed by Chen et al. (2001). Specifically, *NCSKEW* for a firm *i* in year *j* is defined as the negative of the ratio of the third momentum of $W_{i,t}$ to its standard deviation raised to the third power, as shown below:

$$NCSKEW = -\frac{n(n-1)^{3/2} \sum W_{i,t}^3}{(n-1)(n-2)(\sum W_{i,t}^2)^{3/2}}$$

As mentioned in prior literature, multiplying -1 here is to make sure that *NCSKEW* increases as the stock crash risk becomes higher.

The second measure is the down-to-up volatility (*DUVOL*), which is calculated as the

[8] The results are similar if we use another model presented in related papers (e.g. Chang et al. 2017)

$$r_{i,t} = \alpha_i + \beta_1 r_{m,i,t\text{-}1} + \beta_2 r_{ind,i,t\text{-}1} + \beta_3 r_{m,i,t} + \beta_4 r_{ind,i,t} + \beta_5 r_{m,i,t+1} + \beta_6 r_{ind,i,t+1} + \varepsilon_{i,t}$$

where only one-period lead and lag variables are included, and value-weighted industry index are also added in order to control for industrial common factors.

logarithm of the ratio of standard deviation in weeks with negative returns to the standard deviation in weeks with positive returns.

$$DUVOL = ln\left[\left((n_u - 1)\sum_{Down} W_{it}^2\right) / \left((n_d - 1)\sum_{Up} W_{it}^2\right)\right]$$

where n_d and n_u are the number of weeks with negative returns and weeks with positive returns, respectively.

In robustness tests, we also include another two measures frequently used in extant literature (Kim et al. 2011; Jin and Myers 2006), *CRASH* and *COUNT*. *CRASH* is a dummy variable indicating the extreme losses in a firm-year. The value of *CRASH* equals to 1 if the firm experiences a firm-specific return falling three times or more of its standard deviation below the average return within the year, and equals to 0 otherwise. *COUNT* refers to the balance of extremely negative and positive returns. Following Jin and Myers (2006), *COUNT* is calculated as the difference between the frequency of firm-specific returns falling three times of its standard deviation or more below the average return within the year and the frequency of firm-specific returns rising three times of its standard deviation or more above the average return within the year. In this paper, we use measures of stock price crash risk in year t+1 as our dependent variables to alleviate reverse causality problems. In order to understand whether the association between sophisticated active attention and crash risk comes from potential common factors in the stock market, we further measure crash risk from option market. Following Kim and Zhang (2016) and Kim et al. (2018), we use implied volatility smirk as the measure of Ex-ante crash risk. The implied volatility smirk is calculated as the difference between the implied volatility of the OTM put option and implied volatility of the ATM call option. The option data come from Option Metrics.

Measuring sophisticated active attention

In this paper, we use the three measures of EDGAR searching volume to proxy sophisticated attention (or sophisticated information acquisition). In extant literature, three methods are used as criteria to count the non-robust requests, deriving the counting variables of non-robot requests as: (1) *lm* from Loughran and McDonald (2017); (1) *drt* from Drake et al. (2015); and (3) *ryans* from Ryans (2017). Specifically, Loughran and McDonald (2017) identify non-robust page viewers under the assumption that human does not download more than 50 items in a day. Drake et al. (2015) require that human does not download more than 5 items per minute. Ryans (2017) looses the two criteria mentioned above to 500 items/day and 25 items/minute and introduces another restriction that human does not search more than 3 firms in a minute. In order to alleviate the concern that our results will be biased for the skewness of counting variables, we use the logarithm form of these counting variables as our proxies of sophisticated attention. Specifically, *ESV_LM*, *ESV_DRT*, *ESV_RYANS* are defined as the logarithm of 1 plus *lm*, *drt*, and *ryans*, respectively. Table A2 presents the two-way distributions among *ESV_LM*, *ESV_DRT*, and *ESV_RYANS*, from which we find that the three variables are quite similar to each other. Besides, from the density of the upper triangle and lower triangle in the tables, we find that the values of *ESV_RYANS*, *ESV_DRT*, and *ESV_LM* are in descending order, which is consistent with our common knowledge that stricter criteria result in lower value of EDGAR searching volume. According to Ryans (2017), *ESV_RYANS* performs the best among the three measures, especially after the adoption of XBRL. Therefore, we choose *ESV_RYANS* to be the main independent variable in our analysis.[9]

Control variables

Following prior studies on information acquisition and stock price crash risk (e.g., Drake et al. 2015, Hutton et al. 2009, Jin and Myers 2006), we include a set of control variables that may

[9] The other two measures are used in robustness tests, and we find similar results.

be correlated with both the dependent and independent variables. The control variables include: (1) firm size (*SIZE)* measured as the logarithm of 1 plus inflation-adjusted book value of total assets; (2) Tobin's Q (*TOBINQ)* measured as market value of total assets over book value of total assets; (3) cash flow scaled by book value of total assets (*CASH_FLOW)*; (4) book leverage (*BLEV)* calculated as the total liability over book value of total assets; (5) capital expenses scaled by book value of total assets *(CAPX)*; (6) growth of sales scaled by book value of total assets (*GSALE)*; (7) analyst coverage (*NAN)* calculated as the logarithm of 1 plus the arithmetic mean of the 12 monthly numbers of analysts following a firm in a fiscal year; (8) institutional ownership (*IO*) measured by the percent of share held by institutional investors; (9) the mean of firm-specific weekly return (*RET)*; (10) the standard deviation of firm-specific weekly return (*SIGMA)*; (11) change of monthly turnover (*DTURN)*; (12) negative conditional skewness of return (*NCSKEW)*; and (13) opacity proxy based on modified Jones model (*OPAQUE)*. The detailed definitions of these variables are listed in Appendix Table A1. In all our regressions, industries are classified by Fama-French 48 industries based on Fama and French (1997). Standard errors are adjusted for heteroskedasticity and clustered by firm.

Summary statistics

Table 1 reports the summary statistics and correlation matrix of variables used in this paper. Panel A of Table 1 shows the summary statistics of all variables used. Panel B of Table 1 provides the Pearson pairwise correlation matrix of dependent and independent variables. As shown in panel A of Table 1, the summary statistics of variables are similar in magnitude to those in prior studies, which means that the sample we use does not have a structural difference from samples used in other related research.

Interestingly, we find that the dependent variables *DUVOL* and *NCSKEW* are a little bit higher than those used in previous literature. This phenomenon might happen because our sample starts in 2003, which is later than many prior studies. Not surprisingly, this is consistent with the finding of Irvine and Pontiff (2009) that documents the increasing trend in idiosyncratic volatility in the past 40 years resulting from the increasingly competitive external environment over the years. What is more, the financial crisis that happened around 2008 might be another reason for the higher crash risk.[10]

[Insert TABLE 1 about here]

Panel B of Table 1 shows the univariate comparison between groups with high and low *ESV_RYANS*, the EDGAR searching volume measured by Ryans (2017).[11] The observations are sorted into two groups according to the median of *ESV_RYANS* over the years. We report univariate comparisons of firms' characteristics between two groups and their corresponding T-statistics and p-value. Firms' characteristics include measures of EDGAR searching volume, measures of stock price crash risk, and control variables introduced in the previous section.

Panel B of Table 1 shows significant differences in firms' characteristics at 1% significant level between groups with different levels of sophisticated attention. Consistent with Drake et al. (2015), this result illustrates that firms' exposure to sophisticated attention is, to some degree, determined by their characteristics. Specifically, firms with larger size (higher *SIZE*), more investment opportunities (higher *TOBINQ*), more cash flow (higher *CASH_FLOW*), higher book leverage (higher *BLEV*), higher investment (higher *CAPX*), higher sales performance (higher *GSALE*), higher firm-specific stock return (higher *RET*), lower firm-specific stock volatility (lower *SIGMA*), higher increase of turnover (higher *DTURN*), and less opacity (lower *OPAQUE*) are more likely to receive greater sophisticated attention. This evidence shows the necessity to include those

[10] Fortunately, the exclusion of observation in 2008 does not change our results.
[11] The results are similar if we use *LM* or *DRT*.

variables into our regression models in order to alleviate omitted variable problems. What is more, from the comparison of dependent variables between the two groups, we can find preliminary evidence of the positive relationship between crash risk and sophisticated information acquisition, which will be illustrated in detail in the following analysis.

In panel C of Table 1, the pairwise correlations between dependent variables and independent variables show positive associations at 1% level of significance among all these variables, which also reveals basic evidence that sophisticated attention has a positive association with firm-specific crash risk.

4 Main results

Baseline regression

In baseline regression analysis, we employ the following regression specification based on extant literature on stock price crash risk and EDGAR searching volume (Hutton et al. 2009; Kim et al. 2011; Drake et al. 2015; Gibbons et al. forthcoming). The basic model specification is as follows:

$$CrashVar_{i,t+1} = \alpha + \beta_1 ESV_{i,t} + \gamma Controls + FE + \varepsilon_{i,t} \tag{2}$$

where $CrashVar_{i,t+1}$ denotes measures of future stock market crash risk (*DUVOL* and *NCSKEW*); $ESV_{i,t}$ represents measures of EDGAR searching volume (including *ESV_RYANS*, *ESV_LM*, and *ESV_DRT*); $\varepsilon_{i,t}$ is the error term. Since Ryans (2017) shows that *ESV_RYANS* has the best empirical performance (especially after the adoption of XBRL), we use *ESV_RYANS* as our main independent variable in the following analysis.[12] Control variables are mentioned in Section 3.4. *FE* denotes the fixed effects in regression models. In order to alleviate the effect of firm-, time-, and industry-level invariant factors, we include Firm- and Industry×year-fixed effects in most of

[12] In robustness tests, we show that our results are similar if we use *ESV_LM* and *ESV_DRT*, instead.

our regression models.[13] We run OLS regressions to estimate the coefficients. Table 2 presents the results of baseline regressions.

[Insert TABLE 2 about here]

From Table 2, we find that the coefficients on *ESV_RYANS* are all positive and significant at 1% level across simple/full model, different dependent variables (*DUVOL* and *NCSKEW*), and different fixed effects settings. This implies that firms will experience higher crash risk if they receive greater attention from sophisticated market participants, which is consistent with our catering hypothesis that sophisticated attention has a positive effect on firms' future stock market crash risk. The coefficients on control variables are similar to those in extant literature.

In results for future *DUVOL*, the coefficient on *ESV_RYANS* is about 0.026 in the full sample. In terms of economic magnitude, a one-standard-deviation increase in *ESV_RYANS* is associated with a 1.0894×0.026=0.0283 increase in firms' future *DUVOL*. Compared to the sample standard deviation of *DUVOL* (0.3566), the magnitude of the coefficient on *ESV_RYANS* is both statistically and economically significant. Similarly, the economic magnitude of *ESV_RYANS* on *NCSKEW* is 1.0894×0.068=0.0741, which also suggests a statistically and economically significant effect given the sample standard deviation of *NCSKEW*.

From results in columns (2) and (4) (also columns (6) and (8)), we find differences of coefficients of interests between different sets of fixed effects in the full models, which suggests that, although not affecting the main results, controlling for invariant effect on industry-level is necessary for our regression analysis. Therefore, we include Firm- and Industry×year-fixed effects in the following analysis (apart from difference-in-differences analysis). All the results remain when Firm- and Year- fixed effects are included instead.

[13] We also include Firm- and Year- fixed effects and Industry- and Year- fixed effects for robustness tests.

Robustness tests

In this section, we conduct several tests to ensure the robustness of our baseline results. First, we choose another two frequently mentioned proxies of stock price crash risk, *COUNT* and *CRASH*. Second, we use different measures of sophisticated information acquisition based on prior studies.

Alternative dependent variables

Besides *DUVOL* and *NCSKEW*, another two measures of stock price crash risk, *CRASH* and *COUNT*, are also frequently used in extant literature. Different from *DUVOL* and *NCSKEW*, *CRASH* and *COUNT* account for the real stock crashes that happen over the years. The definitions of these variables are listed in Appendix Table A1.

[Insert TABLE 3 about here]

Panel A of Table 3 presents results for robustness tests using alternative measures of stock price crash risk. *COUNT* and *CRASH* are used in these regressions as dependent variables. In addition, *COUNT* is further dissected into *COUNT_UP* and *COUNT_DOWN* in order to test the asymmetric effect of EDGAR searching volume on the frequency of sudden jump and sudden crash. Columns (1)-(4) show the results when dependent variables are *COUNT*, *COUNT_UP*, *COUNT_DOWN*, and *CRASH,* respectively. Other model specifications (fixed effect, industry classification, and cluster level) are the same as those in baseline regressions.

From panel A of Table 3, we find that the sign of coefficient on *ESV_RYANS* is positive and significant at 1% significance level for all the dependent variables in this table. In column (1), the coefficient on *ESV_RYANS* is 0.1139, meaning that a one-standard-deviation increase of *ESV_RYANS* is associated with a 0.1241 (=0.1139×1.0894) increase in *COUNT*. Compared to the *COUNT* mean of 0.0414, the magnitude of the estimate is statistically and economically significant.

In columns (2)-(3), the estimates of the independent variable are -0.0430 and 0.0707 for *COUNT_UP* and *COUNT_DOWN* with significance levels at 10% and 1%, respectively. These results show that the effect of sophisticated attention on stock crash risk mainly comes from increasing the frequency of sudden crashes and comes slightly from decreasing the frequency of sudden jumps.

As shown in column (4) of Table 1, a one-standard-deviation increase in the independent variable leads to about a 4% increase in potential crash risk, which further shows the significant effect of *ESV_RYANS* on firms' stock price crash risk in the following year.

Alternative definitions of EDGAR searching volume

As mentioned above, *ESV_LM* and *ESV_DRT* are also used in related literature. Therefore, as robustness tests, we apply those two measures as independent variables. Since there is no consensus on which measure is the best among *ESV_LM*, *ESV_DRT*, and *ESV_RYANS*, we further construct *ESV_FCOMP* as the first principal component score of *ESV_RYANS*, *ESV_DRT*, and *ESV_LM* in a principal component analysis. Panel B of Table 3 shows the regression results for those measures of EDGAR searching volume.

From panel B of Table 3, we find that the estimates of *ESV_LM* and *ESV_DRT* are quite close to the estimate of *ESV_RYANS* in the full model shown in Table 2. Coefficients on *ESV_LM* and *ESV_DRT* are 0.0266 and 0.0241respectively when the dependent variable is future *DUVOL*, and 0.0721 and 0.0624 respectively when the dependent variable is future *NCSKEW*. Compared to coefficients shown in Table 1 (0.0262 and 0.0691 for *DUVOL* and *NCSKEW* respectively with the same model specification), there are only roughly 6% difference from results when we apply alternative measures of EDGAR searching volume. The estimates of *ESV_FCOMP* is a little bit lower than the three measures above, but the magnitude and significance of the coefficients still hold.

Since analysts and institutional investors are identified by Loughran and McDonald (2017) as important EDGAR users, and the effect of information acquisition from the two market participants is proved in prior studies (Chen, Kelly et al. 2020; Gibbons et al, forthcoming), it is necessary to eliminate this effect from measures of EDGAR searching volume. Enlightened by Li and Sun (2018), measures of abnormal EDGAR searching volume are the residuals from Equation (3):

$$ESV_{i,t} = \delta_0 + \delta_1 SIZE_{i,t} + \delta_2 TOBINQ_{i,t} + \delta_3 NAN_{i,t} + \delta_4 IO_{i,t} + IndustryFE + \varepsilon_{i,t} \tag{3}$$

where *ESV* refers to *EDGAR* searching volume measures *ESV_RYANS*, *ESV_LM*, and *ESV_DRT*. We denote the residuals as *ABN_ESV_RYANS*, *ABN_ESV_LM*, *ABN_ESV_DRT* for *ESV_RYANS*, *ESV_LM*, and *ESV_DRT,* respectively. Panel B of Table 5 shows the results when those measures are introduced as independent variables.

From panel C of Table 3, we find that our baseline results hold in terms of magnitude and significance. There is only a slight change when we apply the measures of abnormal EDGAR searching volume, which is not surprising since we have already included analysts and institutional investors in our baseline regressions.

Alternative explanations

Controlling for general attention based on Google searching index

Besides information acquisition through EDGAR, information acquisition through general searching platforms is also discussed in extant literature. Prior studies use searching volume from information platforms like Yahoo! Finance, Google, Bloomberg, and Twitter as proxies to analyze the real effect of investor attention on firms' operating performance and market performance (Drake et al. 2012; Ben-Rephael et al. 2017; Bartov et al. 2018). Different from EDGAR which provides raw information of firms, media platforms, such as Yahoo! Finance and Google, are often used by market participants to find processed information on firms. Detailed information useful to sophisticated market participants is always omitted in the processed information platforms. However, this omitted information, such as the tone in annual reports and the change of tone in the Management Discussion and Analysis (MD&A) section, is proved to have a real impact on the stock market (Loughran and McDonald 2011; Feldman et al. 2010). Therefore, it is natural to believe that searching volume from EDGAR should contain extra information compared to the searching volume from general information platforms.

However, if the increase in stock price crash risk comes from attention to general information[14] instead of sophisticated attention since searching volume on EDGAR is associated with information acquisition through mass media, our results that sophisticated attention triggers firms' stock crash risk would be misleading. Therefore, to exclude the alternative explanation that the effect of sophisticated attention on stock price crash risk is driven by the attention of ordinary people, we include a measure of general attention into our analysis. Following Drake et al. (2015), we include Google Searching Index (*Google*) to capture the effect of general attention on stock price crash risk. The Google searching index data come from Google Trends. The variable *Google* is calculated as the average searching index of months over the year, scaled by 100. We report the summary statistics of *Google* in panel A of Table 4.

In order to investigate whether general attention is associated with sophisticated information, we first examine the correlation between Google and measures of EDGAR searching volume. Panel B of Table 4 shows the correlation between *Google* and measures of EDGAR searching volume.

[14] For simplicity, we use the term "general attention" in the following analysis.

[Insert TABLE 4 about here]

From panel B of Table 4, we find a highly significant correlation between *Google* and measures of EDGAR searching volume. Surprisingly, the correlations between Google and *ESV_RYANS*, *ESV_LM*, *ESV_DRT*, and *ESV_FCOMP* are significantly positive, while the correlations between Google and abnormal ESV are significantly negative. Since abnormal ESV variables are residuals from Equation (3), we guess the discrepancy of correlation coefficients comes from information associated with analysts and institutional investors. From the negative correlation coefficients, we find an information substitution between EDGAR and mass media in sophisticated market participants apart from analysts and institutional investors. However, in the following regressions, this discrepancy does not change our findings.

Panel C of Table 4 presents the regression results for *DUVOL,* and panel D of Table 4 presents the results for *NCSKEW.* Control variables contain *Google* and control variables in baseline regressions. The two panels show that our findings in baseline regressions still hold after controlling for general attention measured by *Google*, which means that information acquisition through EDGAR is different from that through mass media channels, which is consistent with Drake et al.'s (2015) findings.

Does information supply drive the results?

The baseline results show that greater sophisticated active attention motivates firms' bad news hoarding. However, if the greater attention simply comes from the more filings that firms submit to SEC (namely, the increase in information supply), our results may be driven by firms' information supply instead of sophisticated active attention. To alleviate such concern, we include the number of firms' SEC filings in the regressions. Table 5 shows the results.

[Insert TABLE 5 about here]

In this test, we include the logarithm form of the total number of filings (*LN_NUMBER_TOTAL*) to control for the total information content that firms submit to SEC and the number of voluntary 8-K filings (*LN_VOL_8K*) identified as Lerman and Livnat (2010)[15] to control for the effect of firms' voluntary disclosure. By including these variables, the results do not change the significance of the coefficient on EDGAR searching volume measured by Ryans (2017), which suggests that our baseline results are not driven by firms' information supply.

Effects of different types of filings

Next, we investigate different effects from types of filings. To simplify, we briefly separate EDGAR filings into three types (10-K, 10-Q, and 8-K&Others) to represent filings releasing in long, medium, and short releasing cycles, respectively. Intuitively, filings released irregularly bring more unexpected information to the market than filings released regularly. Therefore, filings with shorter releasing cycles are expected to contain more unexpected information. Therefore, information acquisition from filings with shorter releasing cycles and irregularly released filings should have a stronger effect on stock price crash risk. In the three types of filings, 10-K is released annually, 10-Q is reported quarterly, and 8-K and other filings are released irregularly over fiscal years. Therefore, we posit that sophisticated attention to 8-K and other filings should have the strongest effect on stock crash risk, and least for attention to 10-K filings.

[Insert TABLE 6 about here]

Table 6 shows that the effect of sophisticated attention in 8-K&Other group is the greatest and most significant among all types of filings. The estimate on *ESV* in 8-K&Others group is 0.0225 (0.0587) at 1% significance level for the dependent variable *DUVOL* (*NCSKEW*). Considering the mean of *DUVOL* and *NCSKEW* in our sample, these estimates are both statistically and

[15] Specifically, an 8-K filling is identified as voluntary if it contains Item 2.02, Item 7.01, and Item 8.01.

economically significant. For 10-Q group, the estimate of *ESV* is 0.0112 (0.0258) for *DUVOL* (*NCSKEW*), which is about half the amount of the corresponding estimate in 8-K&Others group.

What is more, though positively significant, estimates in 10-Q group are only significant at 10% level. For 10-K group, the estimate of *ESV* is 0.0087 (0.0199) for *DUVOL* (*NCSKEW*), which is about 70% of the corresponding estimate in 10-Q group and not significant. These results give evidence to support our prediction that filings with shorter releasing cycles and irregular releasing time have more impact on the effect of sophisticated attention on stock price crash risk.

Heterogeneity analysis

In this section, we test our second prediction by examining the heterogeneous effects of sophisticated attention on stock price crash risk across different catering motivations. Because for firms with bad performance (e.g., failing to hit earnings benchmarks or more financial constraints) face higher pressure from the market (Gramham et al. 2005), releasing bad news will further hurt firms' reputation as well as increasing managers' turnover risk. Similarly, when firms have more investment opportunities, releasing bad news is more costly because of the greater reduction of firms' market reputation and increase in the cost of capital (Burgstahler 1997; Skinner and Sloan 2002), thus leading firms to miss positive NPV projects (Jensen and Meckling 1976). Similarly, firms that are more financially constrained are more intended to withhold bad news when the potential cost of releasing them is higher. Overcall, according to Hypothesis 2, we expect the impact of sophisticated active attention to be stronger for firms failing to hit earnings benchmarks proxied by analysts' consensus expectations, firms with more investment opportunities, and more financially constrained firms.

Hitting earnings benchmarks

We investigate whether hitting earnings benchmark or not plays an important role in bad news accumulation for firms with greater sophisticated attention. According to Graham et al. (2005), most executives regard analysts' consensus expectation as the earnings benchmark. If Hypothesis 1 is correct, concerning their current position and reputation, managers who fail to beat analysts' expectations this year have more incentives to increase earnings and beat analysts' expectations in the next year. Since managers face more pressure from sophisticated market participants when they perform poorly, we posit that the effect of sophisticated attention on stock crash risk should be higher for firms failing to beat analysts' expectations currently. Information on analysts' consensus of expected EPS comes from I\B\E\S. We conduct a subsample analysis of the impact of sophisticated attention on stock price crash risk grouped by whether firms beat analysts' expected EPS. A firm is included in the Beat (Not-beat) group if its EPS is higher (lower) than analysts' consensus EPS. Panel A of Table 7 reports the results.

Consistent with our hypothesis, we observe that coefficients of interest in Beat group are significantly lower than those in Not-beat group, which implies that managers are more worried about their position with greater sophisticated attention when the current performance fails to meet analysts' expectations. The magnitudes of differences are -0.046 for *DUVOL* and -0.098 for *NCSKEW* with 1% significance level, showing the importance of analysts' expectation in forming crash risk under sophisticated attention.

Investment opportunity

In the bad news hoarding story in the extant literature, the cost and benefit of holding bad news determine the bad news accumulation, thus causing future stock crash risk. For firms with more investment opportunities, managers face more blame from outside if they failed to run their business well than managers in firms with fewer investment opportunities. That is to say, the benefit of holding bad news (also, the cost of not holding bad news) is higher for firms with more investment opportunities. Therefore, for firms with more investment opportunities, the effect of sophisticated attention on stock price crash risk should be stronger than firms with fewer investment opportunities. We examine this hypothesis by using subsample regressions.

Panel B of Table 7 presents the subsample analysis of the impact of sophisticated attention on stock price crash risk sorted by investment opportunity. We use Tobin's Q (*TOBINQ*) to proxy for investment opportunities. A firm is included in the High (Low) investment opportunity group if its Tobin's Q is higher (lower) than the median of Tobin's Q over fiscal years. In all the regressions shown in this table, *ESV_RYANS* is used as the measure of EDGAR searching volume. The model specifications are the same as those in baseline regressions.

The results show that the difference between High and Low group is 0.018 (0.039) when the dependent variable is *DUVOL* (*NCSKEW*), and the two differences are significant at 10% and 5% level for *DUVOL* and *NCSKEW,* respectively. This means that the effect of sophisticated attention on stock price crash risk is significantly higher for firms with more investment opportunities than firms with fewer investment opportunities, which is consistent with our inference above.

Financial constraint

In prior studies, when firms are financially constrained, their performance becomes more volatile, thus increasing CEOs' turnover risk (Jenter and Kanaan 2015). Since career concern is one of the main reasons for managers' bad news hoarding, and managers in financially constrained firms are more vulnerable to bad news (especially when firms are followed by greater sophisticated market participants), it is interesting to investigate whether managers of financially constrained firms have more incentives to accumulate bad news under sophisticated attention.

To exam the role of financial constraint in our analysis, we adopt the financial constraint level measured by Whited and Wu (2006) as the sorting variable. A firm is included in the High (Low) financial constraint group if its financial constraint proxy is higher (lower) than the median of the financial constraint measures over fiscal years. Model specifications are the same as baseline regressions. Panel C of Table 7 presents the results of subsample analysis by financial constraint level.

In panel C of Table 7, we find a significant difference in coefficients on *ESV_RYANS* between High and Low groups. On average, the coefficient on *ESV_RYANS* in the High group is 40% higher than that in Low group, which shows that managers in firms with higher financial constraints are more likely to hide bad news when receiving greater sophisticated attention.

In summary, through heterogeneity tests, we examine the incentives of managers' bad news hoarding under sophisticated attention across catering motivation, which is consistent with Hypothesis 2 and supports the catering motivation in Hypothesis 1.

Additional evidence on bad news hoarding

Evidence from ex-ante crash risk in the option market

To further examine the effect of sophisticated active attention on firms' bad news hoarding, we test whether managers' bad news hoarding under greater sophisticated active attention can also be observed from option markets. Following Kim et al. (2016), among others, we use ex-ante crash risk to gauge managerial bad news hoarding. Ex-ante crash risk is measured as firms' option implied volatility smirk (*IV_SKEW*), which is calculated as the difference between the implied volatility of the OTM put option and implied volatility of the ATM call option. As mentioned in Kim et al. (2016), the use of ex-ante crash risk reveals investors' expectation of firms crash risk instead of the real crash risk and provides a larger variation of the dependent variable. The option data come from OptionMetrics. With the same control variables and fixed effects as baseline regressions, the estimation model is given as Equation (4):

$$Ex-ante\ CrashRisk_{i,t} = \alpha + \beta_1 ESV_{i,t} + \gamma Controls + FE + \varepsilon_{i,t} \tag{4}$$

where *Ex-ante CrashRisk* denotes IV_SKEW, ESV denotes EDGAR searching volume measures including ESV_RYANS, ESV_DRT, and ESV_LM. Table 8 shows the results.

[Insert TABLE 8 about here]

In Table 8, we observe that the estimates of interest are all significantly positive across different independent variables and different sets of fixed effects. On average, the coefficients of interest are above 0.005, meanings that a one-standard-error increase in the attention measure results in about a 0.04 increase in option implied volatility smirk. Considering the mean and standard deviation of IV_SKEW in this sample (0.0272 and 0.0269, respectively), this magnitude shows the economic significance of sophisticated active attention. Given that IV_SKEW measures crash risk from an alternative source, the consistent results corroborate the positive relation between sophisticated active attention and firms' crash risk mentioned in Hypothesis 1.

Evidence from management guidance

Following Bao et al. (2018), we use information in management guidance to measure managers' release of bad news. Based on our prediction, if managers hide bad news under greater sophisticated active attention, we would observe a negative association between ESV and the number of bad news released in the management guidance. In this test, we use $FREQ\_BAD\_MG$, the number of bad-news management guidance in each firm-year, as our dependent variable. Following prior works (Bao et al. 2018), a management guidance is defined as a bad news if the EPS of management guidance is lower than the most recent consensus analyst forecast. Information on management guidance and analyst forecast comes from I\B\E\S Database. Table 9 shows the results.

[Insert TABLE 9 about here]

From Table 9, we find a positive relation between ESV and the number of bad news released, and the relation is robust to alternative sets of fixed effects and calculation methods of EDGAR searching volume, which shows that firms under greater sophisticated attention release fewer bad news in management guidance, thus supporting our main hypothesis.

Evidence from financial reporting quality

In order to cater to investors, managers have the incentive to inflate firms' earnings through earnings management so as to hide bad news, even if the methods of earnings management are harmful to firms in the future (Hutton et al. 2009; Graham et al. 2005; Dichev et al. 2013). For accrual-based earnings management, the inflated earnings come from adjustment on accounting items. This adjustment works until it is detected. Investors will discount firms' value once they find firms' earnings management, thus causing a crash in the stock market. Besides, the side effect of accrual-based/ real earnings management further increases firms' stock crash risk by distorting optimal operation decisions. Therefore, if managers receive greater sophisticated attention and use earnings management to inflate earnings in order to cater to investors, firms' stock crash risk will increase. As a result, the quality of financial reporting will decrease in this case.

In order to find further evidence on managerial bad news hoarding from financial reporting quality, we first regress measures of earnings management on EDGAR searching volume with same set of fixed effects and the clustering level. Control variables include firm size (*SIZE*), Tobin's Q (*TOBINQ*), cash flow (*CASH_FLOW*), capital expenditure (*CAPX*), growth of sales (*GSALES*), analyst coverage (*NAN*), and institutional ownership (*IO*).[16] Panel A of Table 10 reports the results.

[Insert TABLE 10 about here]

[16] We drop opacity measure *OPAQUE* in these regressions because *OPAQUE* is defined as the moving average of discretionary accruals, which may distort our results. The results are the similar if *OPAQUE* is not dropped.

The first two columns of panel A show the significant association between sophisticated attention and accrual-based earnings management. Dependent variables in columns (1) and (2) are the absolute value of discretionary accruals based on the modified Jones model, *ABSDA_MJONES*, and Kothari et al. (2006), *ABSDA_KOTHARI*. The rest of the columns reports the relation between sophisticated attention and real earnings management. Column (3) presents the results for the absolute value of real earnings management (*ABSRM*) defined in Zang (2012). Columns (4)-(6) show the results for the absolute value of real earnings management on production (*ABSRM_PROD*), discretionary expenses (*ABSRM_DISC_EXP*), and operating cash flow (*ABSRM_OANCF*), respectively.

In columns (1)-(3) of Panel A, coefficients on *ESV_RYANS* are all positive and significant at 1% significance level, which supports our hypothesis that managers may use earnings management to hide bad news when they receive higher pressure from sophisticated market participants, and the inflated earnings will increase future stock crash risk. Together with other columns of Panel A, we observe that both accrual-based and real earnings management are used to inflate earnings, in which the manipulation of production cost and operating cash flow are used as tools of real earnings management.

In addition, we use alternative measures of accrual quality as our dependent variables. Six measures of accrual quality are used in this analysis. *AQ_DD, AQ_MDD,* and *AQ_FLOS* denote the standard deviation of firm-level residual from models proposed by Dechow and Dichev's (2002), Dechow and Dichev's (2002)[17], and Francis et al. (2005), respectively. In addition, we apply model-free measures, such as total write-down, special item, and change in goodwill, to examine accrual quality. *WDA* is defined as 100 times the total write-down scaled by total assets.

[17] We use the modified version as proposed by McNichols (2002).

SPI is 100 times the special item scaled by total assets. $\triangle GDWL$ denotes 100 times the increasing value of goodwill scaled by total assets. From panel B of Table 10, we find that firms with greater sophisticated active attention are more likely to have lower accrual quality measured as a higher value of *AQ_DD, AQ_MDD,* and *AQ_FLOS*. The decrease in accrual quality can also be observed through a decrease in total write-down and special item, as well as an increase in goodwill. Taken together, as reflected from firms' financial reporting quality, these results corroborate our baseline results that firms with greater sophisticated active attention are more likely to hide bad news.

Evidence from accounting conservatism

We further validate the bad new hoarding by accounting conservatism. Hoping that the impact of bad news can be offset by future good news, managers would choose to hide bad news by decreasing their level of accounting conservatism. In this way, when bad news accumulates, firms' future crash risk will increase (Kim and Zhang 2016). When firms are followed by greater sophisticated market participants, the increasing cost of releasing bad news forces managers to hold them in the hope that investors' attention may be distracted by good news in the future (Graham et al. 2005). What is more, the benefit of releasing good news in advance, another way recognized as a decrease in accounting conservatism, will be larger when firms received greater sophisticated attention. Therefore, if managers under greater sophisticated attention hide bad news, we would observe a negative association between accounting conservatism and sophisticated attention. To investigate firms' accounting conservatism level under sophisticated attention, we use Khan and Watts's (2009) firm-year accounting conservatism measure *C-SCORE* as the dependent variable. As shown in panel C of Table 10, the coefficients of interest are positive across different measures of ESV, which is consistent with our hypothesis.

Taken together, evidence from option market reactions, management guidance, financial reporting quality, and accounting conservatism all point to a positive impact of sophisticated active attention on managerial bad news hoarding, which lends further credence to hypothesis 1.

The role of executive compensation

Since the results support the catering hypothesis, managers' interests (e.g., compensation, career and reputation concerns) seem to be an essential factor that affects managers' bad news hoarding when facing greater outside attention. Therefore, it is interesting to examine the role of executive compensation in this situation. Testing on executives' compensation measures, we find that option compensation and non-equity incentive compensation plan have an impact on managers' bad new hoarding decisions.

[Insert TABLE 11 about here]

Results in Table 11 show that the interaction terms between option compensation and sophisticated attention are significantly negative while those between non-equity incentive compensation plans are significantly positive. The results are similar in executive-average and CEO level of both crash risk measures. Since option compensation focuses on performance in the future but non-equity incentive compensation plan, which often refers to cash bonuses, is related to current performance, these results suggest that managers are less likely to hide bad news under greater attention when the compensation is related more to their further wealth rather than current wealth. This evidence sheds light on the contract design in order to prevent managerial bad news hoarding when firms are under greater attention.

5 Identification: two quasi-natural experiments

In Section 4, we find empirical results of the positive relation between sophisticated attention and stock price crash risk and evidence to support our hypothesis on this relation. However, though we use future crash risk measures as dependent variables and include current *NCSKEW* as the dependent variable to avoid reverse causality, potential endogeneity problems are not entirely ruled out. In order to confirm the causal relation in our analysis, we further conduct two natural experiments. The first experiment comes from the extensive adoption of eXtensible Business Reporting Language (XBRL) in 2011. We also implement a placebo test based on the bootstrap simulation method in order to eliminate potential mechanical and occasional results in this experiment. The second experiment comes from Kempf et al.'s (2017) measure of shareholder distraction, which calculates the portion of shares held by institutional investors who are distracted by shocks from other industries. In both of the experiments, we find significant and consistent results, thus establishing a causal link between sophisticated active attention and firms' bad news hoarding.

Identification: Tier 3 adoption of XBRL

Adoption of XBLR

As an interactive and standard markup language, eXtensible Business Reporting Language (XBRL) facilitates the download and analysis of financial statement information (SEC 2009). [18] After the adoption of XBRL, a surge of literature studies the effect of XBRL on capital market and corporate decisions (e.g., Blankespoor et al. 2014). Although this literature focuses on different topics, it is widely recognized that the adoption of XBRL has a real effect on facilitating investors' information acquisition and improving market transparency (e.g., Chen and Zhou 2018; Kim et al. 2019).[19]

[18] During 2009-2011, the SEC mandated firms to provide financial statements in the form of XBRL over three phase-in periods. The first period (Tier 1 XBRL) started in 2009 for firms with a public float higher than $5 billion. The second period (Tier 2 XBRL) begins in 2010 for firms with a public float higher than $700 million. In the last period (Tier 3 XBRL), an extensive adoption of XBRL is implemented in 2011 when all the firms are mandated to adopt XBRL.

[19] The adoption of XBRL increases investors' information acquisition mainly in two ways. First, it increases the efficiency of understanding the information contented in the financial statement. Second, the use of XBRL decreases the threshold of market participants to process financial information (the users are still required to have basic

Extant literature uses the staggered adoption of Tier 1 XBRL or Tier 2 XBRL as exogenous shocks in DID or RDD framework (e.g., Dong et al. 2016; Kim et al. 2019). However, some papers document that the results from those analyses are partially valid because Tier 3 XBRL, which contains the majority of listed firms, is not taken into account. What is more, since the effect of adoption XBLR is not temporal and the three periods are implemented in consecutive years, the average treatment effect in those tests are biased if firms in the control group are affected by the next period of XBLR adoption (Kim et al. 2019).

In this paper, we use Tier 3 XBRL adoption as an exogenous shock in our analysis and construct treatment and control groups based on the increase of sophisticated attention during the experiment period. There are four reasons why we use Tier 3 XBRL adoption. First, since Tier 3 XBRL refers to the majority part of listed firms, this shock can provide more general evidence in our study than Tier 1 and Tier 2 that only focus on large- and median-size firms. Second, since all firms are required to use XBRL after 2011, our results are unaffected by the contamination in the post-treatment period. Third, since the effect of XBRL adoption is not temporal, the effects from earlier periods can still work in Tier 3, which decreases the bias of our analysis. Fourth, according to prior studies, many popular XBRL tools, such as XBRL Data in Use, Calcbench, and SQL Power XBRL Analytics, did not appear at the time in the early stages of XBRL adoption and investors need time to learn how to use XBRL before using it. As a result, the effect of Tier 1 and Tier 2 is constrained, which is consistent with findings in Blankespoor et al. (2014) and Harris and Morsfield (2012).

Meanwhile, there is a problem when using this shock. Since we cannot guarantee that the

knowledge on finance and accounting, and are also part of sophisticated market participants), thus attracting more investors to search on SEC filings in EDGAR. Though those ways, on average, the attention of sophisticated market participants is higher after the adoption of XBRL.

control group does not receive any treatment during our analysis period (because some of them have already adopted XBRL in Tier 1 and Tier 2), and the effect in earlier periods of the adoption of XBRL is not temporal (Harris and Morsfield 2012), our estimate on the average treatment effect is supposed to be downward-biased. Therefore, we need to be careful to interpret results in this framework.

Difference-in-differences analysis

The difficulty of using the extensive adoption of XBRL is that we do not have a clear boundary between treated and control groups since all the listed firms are affected by the extensive adoption of XBRL in 2011. Imitating the method used in the literature on liquidity where decimalization in 2001 is treated as an exogenous shock (e.g., Fang et al. 2014), we construct treatment and control groups through the following processes. Firms receiving higher increased searching volume during the period 2010-2012 (top half) are included in the treatment group, and firms that receive lower increased searching volume during the same period (bottom half) are included in the control group.

We implement a propensity score matching between treatment and control groups using the observations in 2010, matching for fundamental variables including control variables used in the baseline regressions. We also use the first principal component score *ESV_FCOMP* as a matching variable to control for the difference in sophisticated active attention before the treatment. Dependent variables *DUVOL* and *NCSKEW* are also included to control the alleviate differences in prior stock price crash risk between the groups. Each observation in the treatment group is matched to one observation in the control group by the nearest-neighbor criterion. Finally, the matched sample consists of 6,087 observations. We conduct a balance test to verify that there are no differences between treated and control firms in any of our matching variables.[20] After matching,

[20] Appendix Table A3 shows the results.

we adopt a DID test in a seven-year window centered on 2011 as follows:

$$CrashVar_{i,t+1} = b_0 + b_1 TREAT \times POST + \gamma Controls + FirmFE + YearFE + \varepsilon_{i,t} \quad (5)$$

where *TREAT* is a dummy variable that equals one if the firm is in the treatment group, and zero otherwise; *POST* if the fiscal year is equal to or after 2011, and zero otherwise; Control variables are the same as those in baseline regressions. Firm- and Year- fixed effects are included according to the setting of difference-in-differences tests. The results are shown in Table 12.

[Insert TABLE 12 about here]

Columns (1) and (4) of Table 12 show the results when no control variables are included. The average treatment effect of this experiment is 0.0675 (0.1287) for *DUVOL* (*NCSKEW*) at 1% significance level. The magnitude of the DID estimator on *DUVOL* (*NCSKEW*) is about 18.9% (16.7%) of its standard deviation in the sample, suggesting that the treatment effects are both statistically significant and economically meaningful for both measures of stock crash risk. Columns (3) and (6) present similar results when we add control variables into Equation (5). The results still hold at 1% significant level. Considering that our results are downward-biased, the real effect of sophisticated attention on stock price crash risk is even more significant.

In order to investigate the dynamics of the experiment, we future use a regression framework enlightened by Bertrand and Mullainathan (2003) and Fang et al. (2014) in a seven-year window centered around the year 2011 and estimate the following model:

$$\begin{aligned} CrashVar_{i,t+1} = c_0 &+ c_1 TREAT \times BEFORE^{-1} + c_2 TREAT \times CURRENT \\ &+ c_3 TREAT \times AFTER^{1} + c_2 TREAT \times AFTER^{2\&3} + FirmFE + YearFE + e_{i,t} \quad (6) \end{aligned}$$

where *BEFORE*$^{-1}$ is a dummy variable that equals one if the observation is from one year before 2011 and zero otherwise; *CURRENT* is a dummy variable that equals one if the observation is from the year 2011 and zero otherwise; *AFTER*^{1} is a dummy variable equals one if the observation is from one year after 2011 and zero otherwise; and $AFTER^{2\&3}$ is a dummy variable equals one if the observation is from two or three years after 2011 and zero otherwise. As shown in columns (2) and (5), the coefficients on $TREAT \times BEFORE^{-1}$ are not statistically significant, which suggests that the parallel trend assumption of DID is not violated. We observe coefficients on *TREAT*×*CURRENT*, $TREAT \times AFTER^{1}$, and $TREAT \times AFTER^{2\&3}$ are all positive and statistically significant, but the treatment effect decreases as time passes by. Overall, these results find evidence supporting the causal effect of sophisticated attention on stock price crash risk.

Placebo tests

To rule out the possibility that our DID results come from mechanical reasons or driven by chance. We further conduct a placebo test using the bootstrap simulation method. Specifically, we first randomly select a group of observations as a pseudo treatment group and then use the same method to construct a pseudo control group. The group size of pseudo-treated and control groups are the same as our real difference-in-differences analysis. We require firms not to have missing values in 2010 to make sure that the experiment is processed in the same way as Table 12. Then, we replicate columns (1) and (4) of Table 12 using the pseudo samples for 5,000 times and record the coefficient on *TREAT* ×*POST*. Figure 1 shows the histogram of the coefficients on *TREAT* ×*POST* from 5,000 bootstrap simulations of difference-in-differences analysis. Panel A of Figure 1 shows the results when the dependent variable is future *DUVOL*. Panel B of Figure 1 shows the results when the dependent variable is future *NCSKEW*. In each panel, the red line paralleled to y-axis shows the actual results given in Table 12. The mean and standard deviation are also shown in each panel.

[Insert FIGURE 1 about here]

In panel A of Figure 1, the actual coefficient on *TREAT* ×*POST* is 0.0675, which is about 4 times of the standard deviations (0.0179) far away from the mean (-0.0003) of the simulation distribution. Similarly, in panel B of Figure 1, the actual coefficient on *TREAT* ×*POST* is 0.1287, which is about 3.5 times the standard deviations (0.0383) far away from the mean (-0.0006) of the simulation distribution. This evidence shows that our findings are not driven by mechanical factors or purely driven by chance. The results are similar if we replicate columns (3) and (6) for regressions with control variables.

Identification: exogenous shocks from other industries

In the second quasi-natural experiment, we implement the shareholder distraction measure proposed by Kempf et al. (2017). This measure is based on institutional investors' holdings data and exogenous shocks from other industries. As a result of exogenous shocks from other industries, shareholder distraction increases managerial opportunism (Kempf et al. 2017). This measure is valid in our analysis because shareholders facing shocks from other industries they invest in are more likely to spend less time and energy to check SEC filings on firms without shocks. Using this measure, we examine the causality on how shareholder attention affects firms' bad news hoarding. The shareholder distraction data is from Kempf's website.[21] We use the year-average shareholder distraction (*DISTRACTION*) as the dependent variable. In this analysis, control variables are the same as those in baseline regressions. Firm-fixed effect and Industry ×year-fixed effect are used to control for the effect of firm-, year-, and industry-invariant factors.[22]

[Insert TABLE 13 about here]

From Table 13, we find that the coefficients on shareholder distraction (*DISTRACTION*) are all significantly negative across different dependent variables and model specifications. Consistent with our hypothesis, the significantly negative results establish the causal link between

[21] https://sites.google.com/site/elikempf/research
[22] The results are similar if we include firm- and industry- fixed effects as those used in Kempf et al. (2017).

sophisticated attention and stock price crash risk.

In sum, both quasi-natural experiments provide evidence to identify the causality that sophisticated active attention leads to higher stock price crash risk. Therefore, we further confirm that our baseline results do not suffer from endogeneity problems.

6 Conclusion

Using a large sample of U.S. firms during 2003-2015, we first find that greater sophisticated active attention pushes managers to hide bad news measured by stock price crash risk, and this finding is robust to alternative measures of crash risk, sophisticated active attention, and model specifications. Consistent with our prediction, this evidence suggests that managers are more likely to hide bad news under greater active attention from sophisticated market participants, and this effect is not driven by general attention and firms' information supply. Further findings suggest that managers with greater sophisticated attention are more likely to cater to investors by hiding bad news when they fail to meet current analysts' expectations, have more investment opportunities, and experience tighter financial constraints. In addition, from additional evidence on managerial bad news hoarding, we find that firms with greater attention are observed to experience an increase in ex-ante crash risk measured by option implied volatility smirk, have lower financial reporting quality, release fewer bad news in management guidance, and delay releasing bad news in financial statements, which further corroborates our prediction. Moreover, we find that attention to irregularly released fillings shows more impact on firms and that long-term compensation design can attenuate the impact of sophisticated active attention. Evidence from two quasi-natural experiments based on Tier 3 XBRL adoption and exogenous shocks from other industries establish a causal link between sophisticated active attention and managerial bad news hoarding measured by stock price crash risk. In this paper, we provide systematic evidence on the impact of sophisticated active attention on managerial bad news hoarding, which answers our questions mentioned at the beginning of this paper. By showing managers' bad news hoarding under pressure from sophisticated market participants through active attention, our study sheds light on the pressure effect that attention can affect managerial bad news disclosures that have been underexplored in prior literature.

References

Baginski, Stephen P., John L. Campbell, Lisa A. Hinson, and David S. Koo, 2017, Do Career Concerns Affect the Delay of Bad News Disclosure?, *The Accounting Review* 93 (2): 61–95.

Baloria, Vishal P., and Jonas Heese, 2018, The Effects of Media slant on Firm Behavior, *Journal of Financial Economics* 129, 184–202.

Bali, Turan G., Lin Peng, Yannan Shen, and Yi Tang, 2014, Liquidity Shocks and Stock Market Reactions, *Review of Financial Studies* 27, 1434–1485.

Ball, R. A., 2009, Market and Political/Regulatory Perspectives on the Recent Accounting Scandals, *Journal of Accounting Research* 47, 277–323.

Bao, Dichu, Yongtae Kim, G. M. Mian, and Lixin Su, 2018, Do Managers Disclose or Withhold Bad News? Evidence from Short Interest, *The Accounting Review* 94, 1–26.

Bartov, Eli, Lucile Faurel, and Partha S. Mohanram, 2018, Can Twitter Help Predict Firm-Level Earnings and Stock Returns?, *The Accounting Review* 93, 25–57.

Bauguess, Scott W., John Cooney, and Kathleen W. Hanley, 2013, Investor Demand for Information in Newly Issued Securities, *Working Paper*.

Ben-Rephael, Azi, Zhi Da, and Ryan D. Israelsen, 2017, It Depends on Where You Search. Institutional Investor Attention and Underreaction to News, *The Review of Financial Studies* 30, 3009–3047.

Blankespoor, Elizabeth, Brian P. Miller, and Hal D. White, 2014, Initial evidence on the market impact of the XBRL mandate, *Review of Accounting Studies* 19, 1468–1503.

Bozanic, Zahn, Jeffery L. Hoopes, Jacob R. Thornock, and Braden M. Williams, 2017, IRS Attention, *Journal of Accounting Research* 55, 79–114.

Chang, Xin, Yangyang Chen, and Leon Zolotoy, 2017, Stock Liquidity and Stock Price Crash Risk, *Journal of Financial and Quantitative Analysis* 52, 1605–1637.

Chen, Huaizhi, Lauren Cohen, Umit Gurun, Dong Lou, and Christopher Malloy. 2020. IQ from IP: Simplifying Search in Portfolio Choice. *Journal of Financial Economics* 138 (1): 118–37.

Chen, Joseph, Harrison Hong, and Jeremy C. Stein, 2001, Forecasting crashes. Trading volume, past returns, and conditional skewness in stock prices, *Journal of Financial Economics* 61, 345–381.

Chen, Yong, Bryan Kelly, and Wei Wu. 2020. Sophisticated Investors and Market Efficiency: Evidence from a Natural Experiment. *Journal of Financial Economics* 138 (2): 316–41.

Chen, Gary, and Jie Zhou, 2018, XBRL Adoption and Systematic Information Acquisition via EDGAR, *Journal of Information Systems* 33, 23–43.

Da, Z. H., Joseph Engelberg, and Pengjie Gao, 2011, In Search of Attention, *The Journal of Finance* 66, 1461–1499.

Dechow, Patricia M., and Ilia D. Dichev, 2002, The Quality of Accruals and Earnings, *The Accounting Review* 77, 35–59.

DeFond, Mark L., and Chul W. Park, 1999, The effect of competition on CEO turnover, *Journal of Accounting and Economics* 27, 35–56.

Dellavigna, Stefano, and Joshua M. Pollet, 2009, Investor Inattention and Friday Earnings Announcements, *The Journal of Finance* 64, 709–749.

Dichev, Ilia D., John R. Graham, Campbell R. Harvey, and Shiva Rajgopal. 2013. Earnings Quality: Evidence from the Field. *Journal of Accounting and Economics* 56 (2-3): 1–33.

Dimson, Elroy, 1979, Risk measurement when shares are subject to infrequent trading, *Journal of Financial Economics* 7, 197–226.

Dong, Ming, David Hirshleifer, and Siew H. Teoh, 2012, Overvalued Equity and Financing Decisions, *Review of Financial Studies* 25, 3645–3683.
Dong, Yi, Oliver Z. Li, Yupeng Lin, and Chenkai Ni, 2016, Does Information-Processing Cost Affect Firm-Specific Information Acquisition? Evidence from XBRL Adoption, *Journal of Financial and Quantitative Analysis* 51, 435–462.
Drake, Michael S., Jared Jennings, Darren T. Roulstone, and Jacob R. Thornock, 2017, The Comovement of Investor Attention, *Management Science* 63, 2847–2867.
Drake, Michael S., Phillip T. Lamoreaux, Phillip J. Quinn, and Jacob R. Thornock, 2019 Auditor benchmarking of client disclosures, *Review of Accounting Studies* 24, 393–425.
Drake, Michael S., Bret A. Johnson, Darren T. Roulstone, and Jacob R. Thornock. 2020. Is There Information Content in Information Acquisition? *The Accounting Review* 95 (2): 113–39.
Drake, Michael S., Darren T. Roulstone, and Jacob R. Thornock, 2012, Investor Information Demand. Evidence from Google Searches Around Earnings Announcements, *Journal of Accounting Research* 50, 1001–1040.
Drake, Michael S., Darren T. Roulstone, and Jacob R. Thornock, 2015, The Determinants and Consequences of Information Acquisition via EDGAR, *Contemporary Accounting Research* 32, 1128–1161.
Fama, Eugene F., and Kenneth R. French, 1997, Industry costs of equity, *Journal of Financial Economics* 43, 153–193.
Fang, Vivian W., Xuan Tian, and Sheri Tice, 2014, Does Stock Liquidity Enhance or Impede Firm Innovation?, *The Journal of Finance* 69, 2085–2125.
Fee, C. E., and Charles J. Hadlock, 2000, Management Turnover and Product Market Competition. Empirical Evidence from the US Newspaper Industry, *The Journal of Business* 73, 205–243.
Feldman, Ronen, Suresh Govindaraj, Joshua Livnat, and Benjamin Segal, 2010, Management's tone change, post earnings announcement drift and accruals, *Review of Accounting Studies* 15, 915–953.
Fich, Eliezer M., Jarrad Harford, and Anh L. Tran, 2015, Motivated monitors. The importance of institutional investors′ portfolio weights, *Journal of Financial Economics* 118, 21–48.
Francis, Jennifer, Ryan LaFond, Per Olsson, and Katherine Schipper, 2005, The market pricing of accruals quality, *Journal of Accounting and Economics* 39, 295–327.
Gibbons, Brian, Peter Iliev, and Jonathan Kalodimos, forthcoming, Analyst Information Acquisition via EDGAR, *Management Science*.
Graham, John R., Campbell R. Harvey, and Shiva Rajgopal, 2005, The Economic Implications of Corporate Financial Reporting. *Journal of Accounting and Economics* 40 (1-3): 3–73.
Harris, T.S., and S.G. Morsfield, 2012, An evaluation of the current state and future of XBRL and interactive data for investors and analysts, *Columbia Business School Working Paper*
Healy, Paul M., and Krishna G. Palepu, 2001, Information Asymmetry, Corporate Disclosure, and the Capital Markets: A Review of the Empirical Disclosure Literature, *Journal of Accounting and Economics* 31 (1-3): 405–40.
Healy, Paul M., and James M. Wahlen, 1999, A Review of the Earnings Management Literature and Its Implications for Standard Setting, *Accounting Horizons* 13, 365–383.
Hirshleifer, David, Sonya S. Lim, and Siew H. Teoh, 2009, Driven to Distraction. Extraneous Events and Underreaction to Earnings News, *The Journal of Finance* 64, 2289–2325.
Hirshleifer, David, and Siew H. Teoh, 2003, Limited attention, information disclosure, and financial reporting, *Journal of Accounting and Economics* 36, 337–386.
Hirshleifer, David, Siew H. Teoh, and Jeff J. Yu, 2011, Short Arbitrage, Return Asymmetry, and

the Accrual Anomaly, *Review of Financial Studies* 24, 2429–2461.
Hutton, Amy P., Alan J. Marcus, and Hassan Tehranian, 2009, Opaque financial reports, R2, and crash risk, *Journal of Financial Economics* 94, 67–86.
Iliev, Peter, Jonathan Kalodimos, and Michelle B. Lowry, forthcoming, Investors' Attention to Corporate Governance, *Review of Financial Studies.*
Irvine, Paul J., and Jeffrey Pontiff, 2009, Idiosyncratic Return Volatility, Cash Flows, and Product Market Competition, *Review of Financial Studies* 22, 1149–1177.
Jensen, Michael C., and William H. Meckling, 1976, Theory of the firm. Managerial behavior, agency costs and ownership structure, *Journal of Financial Economics* 3, 305–360.
Jenter, Dirk, and Fadi Kanaan, 2015, CEO Turnover and Relative Performance Evaluation, *The Journal of Finance* 70, 2155–2184.
Jin, L., and S. Myers, 2006, R2 around the world. New theory and new tests, *Journal of Financial Economics* 79, 257–292.
Kahneman, Daniel, and Amos Tversky, 1979, Prospect Theory. An Analysis of Decision under Risk, *Econometrica* 47, 263.
Kempf, Elisabeth, Alberto Manconi, and Oliver Spalt, 2017, Distracted Shareholders and Corporate Actions, *Review of Financial Studies* 30, 1660–1695.
Khan, Mozaffar, and Ross L. Watts, 2009, Estimation and empirical properties of a firm-year measure of accounting conservatism, *Journal of Accounting and Economics* 48, 132–150.
Kim, Jeong-Bon, Leye Li, Louise Y. Lu, and Yangxin Yu, 2016, Financial statement comparability and expected crash risk, *Journal of Accounting and Economics* 61, 294–312.
Kim, Jeong-Bon, Yinghua Li, and Liandong Zhang, 2011, Corporate tax avoidance and stock price crash risk. Firm-level analysis, *Journal of Financial Economics* 100, 639–662.
Kim, Jeong-Bon, Louise Y. Lu, and Yangxin Yu, 2018, Analyst Coverage and Expected Crash Risk. Evidence from Exogenous Changes in Analyst Coverage, *The Accounting Review*.
Kim, Jeong-Bon, and Liandong Zhang, 2016, Accounting Conservatism and Stock Price Crash Risk. Firm-level Evidence, *Contemporary Accounting Research* 33, 412–441.
Kim, Jeong-Bon, Bing Li, and Zhenbin Liu, 2019, Information-Processing Costs and Breadth of Ownership, *Contemporary Accounting Research* 91, 208.
Kothari, S. P., Susan Shu, and Peter D. Wysocki, 2009, Do Managers Withhold Bad News?, *Journal of Accounting Research* 47, 241–276.
Lerman, Alina, and Joshua Livnat, 2010, The new Form 8-K disclosures, *Review of Accounting Studies* 15, 752–778.
Li, Edward X., Gary Lind, K. Ramesh, and Min Shen, 2018, Externalities of Accounting Disclosures. Evidence from the Federal Reserve, *Working Paper*.
Li, Frank W., and Chengzhu Sun, 2018, Information Acquisition and Expected Returns. Evidence from EDGAR Search Traffic, *Working Paper*.
Li, Jun, and Jianfeng Yu, 2012, Investor attention, psychological anchors, and stock return predictability, *Journal of Financial Economics* 104, 401–419.
Loughran, T. I., and Bill McDonald, 2011, When Is a Liability Not a Liability? Textual Analysis, Dictionaries, and 10-Ks, *The Journal of Finance* 66, 35–65.
Loughran, Tim, and Bill McDonald, 2017, The Use of EDGAR Filings by Investors, *Journal of Behavioral Finance* 18, 231–248.
McNichols, Maureen F., 2002, Discussion of the Quality of Accruals and Earnings. The Role of Accrual Estimation Errors, *The Accounting Review* 77, 61–69.
Nagar, Venky, 1999, The Role of the Manager's Human Capital in Discretionary Disclosure,

Journal of Accounting Research 37, 167.
Peng, Lin, 2005, Learning with Information Capacity Constraints, *Journal of Financial and Quantitative Analysis* 40, 307–329.
Peng, Lin, and Wei Xiong, 2006, Investor attention, overconfidence and category learning, *Journal of Financial Economics* 80, 563–602.
Polk, Christopher, and Paola Sapienza, 2008, The Stock Market and Corporate Investment. A Test of Catering Theory, *Review of Financial Studies* 22, 187–217.
Ryans, James, 2017, Using the EDGAR Log File Data Set, *Working Paper*.
Ryans, James P. 2020. Textual Classification of SEC Comment Letters. *Review of Accounting Studies* 59 (3): 1259.
SEC, U.S., 2009, Interactive Data to Improve Financial Reporting.
Skinner, Douglas J., and Richard G. Sloan, 2002, Earnings Surprises, Growth Expectations, and Stock Returns or Don't Let an Earnings Torpedo Sink Your Portfolio, *Review of Accounting Studies* 7, 289–312.
Verrecchia, Robert E, 2001, Essays on Disclosure, *Journal of Accounting and Economics* 32 (1-3): 97–180.

TABLE 1
Summary statistics

Panel A: Summary statistics

Variables	N	MEAN	ST.DEV	P25	MEDIAN	P75
DUVOL	21,617	0.0029	0.3566	-0.2285	0.0028	0.2349
NCSKEW	21,617	0.1061	0.7666	-0.3205	0.0858	0.4979
COUNT	21,617	0.0414	1.3937	0.0000	0.0000	0.0000
CRASH	21,617	0.2495	0.4327	0.0000	0.0000	0.0000
ESV_RYANS	21,617	7.8797	1.0894	7.1148	7.8962	8.6406
ESV_DRT	21,617	8.0162	1.1227	7.2240	8.0110	8.7936
ESV_LM	21,617	7.7398	1.0687	7.0085	7.7489	8.4665
SIZE	21,617	6.2381	2.0075	4.8335	6.1558	7.5854
TOBINQ	21,617	1.9138	1.2252	1.1243	1.5279	2.2689
CASH FLOW	21,617	0.0392	0.1865	0.0255	0.0795	0.1257
BLEV	21,617	0.1961	0.1952	0.0055	0.1565	0.3155
CAPX	21,617	0.0536	0.0615	0.0166	0.0331	0.0654
GSALE	21,617	0.1330	0.4208	-0.0223	0.0766	0.2009
NAN	21,617	1.3408	1.0995	0.0000	1.3863	2.3026
IO	21,617	0.4963	0.3620	0.0893	0.5647	0.8319
RET	21,617	-0.1784	0.2504	-0.2083	-0.0972	-0.0427
SIGMA	21,617	0.0509	0.0319	0.0295	0.0445	0.0653
DTURN	21,617	0.0256	1.0371	-0.3263	0.0049	0.3532
OPAQUE	21,617	0.0741	0.0623	0.0348	0.0571	0.0922

Panel B: Univariate comparison based on EDGAR searching volume (*ESV_RYANS*)

	Low ESV		High ESV		High-Low		
	N	Mean	N	Mean	Differences	T-statistics	p-value
DUVOL	10,806	-0.030	10,811	0.036	0.067	13.800	0.000
NCSKEW	10,806	0.049	10,811	0.164	0.116	11.100	0.000
COUNT	10,806	-0.044	10,811	0.127	0.172	9.050	0.000
CRASH	10,806	0.240	10,811	0.259	0.019	3.200	0.002
ESV_RYANS	10,806	7.280	10,811	8.479	1.199	96.900	0.000
ESV_DRT	10,806	7.415	10,811	8.617	1.203	93.200	0.000
ESV_LM	10,806	7.136	10,811	8.343	1.207	100.600	0.000
SIZE	10,806	5.274	10,811	7.201	1.927	80.400	0.000
TOBINQ	10,806	1.819	10,811	2.009	0.190	11.450	0.000
CASH FLOW	10,806	0.025	10,811	0.053	0.029	11.350	0.000
BLEV	10,806	0.158	10,811	0.235	0.077	29.800	0.000
CAPX	10,806	0.052	10,811	0.056	0.004	4.650	0.000
GSALE	10,806	0.123	10,811	0.143	0.020	3.500	0.001
NAN	10,806	0.787	10,811	1.894	1.107	85.650	0.000
IO	10,806	0.375	10,811	0.618	0.242	52.250	0.000
RET	10,806	-0.214	10,811	-0.143	0.072	21.250	0.000
SIGMA	10,806	0.058	10,811	0.044	-0.013	-30.500	0.000
DTURN	10,806	-0.018	10,811	0.069	0.086	6.100	0.000
OPAQUE	10,806	0.022	10,811	0.008	-0.014	-13.500	0.000

Panel C: Pairwise correlations of dependent and independent variables

Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)
(1) DUVOL	1.000						
(2) NCSKEW	0.958***	1.000					
(3) COUNT	0.718***	0.768***	1.000				
(4) CRASH	0.558***	0.614***	0.756***	1.000			
(5) ESV_RYANS	0.091***	0.068***	0.050***	0.028***	1.000		
(6) ESV_DRT	0.085***	0.063***	0.046***	0.027***	0.996***	1.000	
(7) ESV_LM	0.086***	0.064***	0.047***	0.026***	0.989***	0.986***	1.000

Notes: This table reports summary statistics and correlation matrix. The data used in our research come from multiple sources. Firm-level financial data come from the COMPUSTAT database. Stock price and return data come from Center for Research in Security Prices (CRSP). Institutional holdings data come from Thomson Reuters Institutional Holdings (13f). Analyst coverage data come from Institutional Brokers Estimate System (I\B\E\S). The EDGAR searching volume data come from James Ryans' EDGAR Log File Data. We include observations that satisfy the following criteria: (1) Book equity is positive; (2) Year-end stock price is above 1 U.S. dollar; (3) At least 26 observations are available in CRSP weekly data for each firm-year; (4) Variables used in our research are available; (5) Each firm should at least have 2-year consecutive observations. Firms in financial industry (SIC codes 6000-6999) and utility industry (SIC codes 4900-4999) are excluded from our sample. Our sample includes firm-years that meet our requirement during the period 2003-2015 when variables of EDGAR searching volume data are available and the request numbers of firms recorded in the dataset are stable. All continuous variables are winsorized at 1st and 99th percentiles to alleviate the potential disturbance from outliers. The final sample consists of 21,617 observations. Panel A of this table shows the summary statistics of variables. Panel B shows the univariate comparison between groups with high and low EDGAR searching volume measured by *ESV_RYANS* in Ryans (2017). Panel C of this table presents the Pearson pairwise correlation matrix of dependent and independent variables. Observations are sorted into halves according to the median of *ESV_RYANS* over the years.*** shows significance at the 1% level.

TABLE 2
Results of baseline regressions

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Variables	*DUVOL*	*DUVOL*	*DUVOL*	*DUVOL*	*NCSKEW*	*NCSKEW*	*NCSKEW*	*NCSKEW*
ESV_RYANS	0.0505***	0.0250**	0.0518***	0.0262***	0.1204***	0.0670***	0.1224***	0.0691***
	(0.0088)	(0.0099)	(0.0089)	(0.0099)	(0.0186)	(0.0210)	(0.0189)	(0.0212)
SIZE		0.0940***		0.0980***		0.1886***		0.1965***
		(0.0099)		(0.0101)		(0.0220)		(0.0226)
TOBINQ		0.0557***		0.0562***		0.1135***		0.1162***
		(0.0040)		(0.0041)		(0.0091)		(0.0092)
CASH_FLOW		0.0954***		0.0739***		0.1378***		0.0993*
		(0.0237)		(0.0245)		(0.0531)		(0.0551)
BLEV		-0.0514*		-0.0594*		-0.0637		-0.0829
		(0.0301)		(0.0305)		(0.0666)		(0.0681)
CAPX		0.1242*		0.1166		0.1851		0.1637
		(0.0733)		(0.0748)		(0.1589)		(0.1615)
GSALE		-0.0022		-0.0015		-0.0002		0.0022
		(0.0077)		(0.0081)		(0.0171)		(0.0179)
NAN		0.0296***		0.0309***		0.0543***		0.0587***
		(0.0081)		(0.0083)		(0.0178)		(0.0185)
IO		0.0622***		0.0654***		0.1104**		0.1137**
		(0.0228)		(0.0231)		(0.0505)		(0.0513)
RET		0.0779**		0.0897**		0.1767**		0.1937**
		(0.0394)		(0.0400)		(0.0856)		(0.0868)
SIGMA		-0.1127		-0.0034		0.1933		0.3538
		(0.3952)		(0.4064)		(0.8606)		(0.8819)
DTURN		0.0090***		0.0104***		0.0181***		0.0209***
		(0.0026)		(0.0026)		(0.0055)		(0.0056)
NCSKEW		-0.0530***		-0.0558***		-0.1191***		-0.1258***
		(0.0036)		(0.0037)		(0.0080)		(0.0082)
OPAQUE		0.0490		0.0708		0.1302		0.1565
		(0.0687)		(0.0706)		(0.1535)		(0.1576)
Constant	-0.3953***	-0.9366***	-0.4053***	-0.9773***	-0.8423***	-1.9219***	-0.8581***	-2.0014***
	(0.0697)	(0.0847)	(0.0702)	(0.0858)	(0.1469)	(0.1852)	(0.1489)	(0.1883)
Year FE	Yes	Yes	No	No	Yes	Yes	No	No
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Industry×Year FE	No	No	Yes	Yes	No	No	Yes	Yes
Observations	21,617	21,617	21,617	21,617	21,617	21,617	21,617	21,617
R-squared	0.1885	0.2258	0.2195	0.2560	0.1844	0.2174	0.2121	0.2452

Notes: This table presents regression results for the relation between sophisticated active attention and stock price crash risk. Equation (2) shows the baseline regression:

$$CrashVar_{i,t+1} = \alpha + \beta_1 ESV_{i,t} + \gamma Controls + FE + \varepsilon_{i,t} \tag{2}$$

where *CrashVar* denotes the proxies of stock market crash risk (*DUVOL* and *NCSKEW*); *ESV* represents the measures of EDGAR searching volume (*ESV_RYANS*, *ESV_LM*, and *ESV_DRT*); Control variables contain firm size (*SIZE)*, Tobin's Q (*TOBINQ)*, cash flow (*CASH_FLOW)*, book leverage (*BLEV)*, capital expenses *(CAPX)*, growth of sales (*GSALE)*, analyst coverage (*NAN)*, institutional ownership (*IO)*, mean of firm-specific weekly return (*RET)*, standard deviation of firm-specific weekly return (*SIGMA)*, change of monthly turnover (*DTURN)*, negative conditional skewness of return (*NCSKEW)*, and opacity proxy based on modified Jones model (*OPAQUE)*. Measures of crash risk used here are *DUVOL* and *NCSKEW*. Columns (1)-(4) show the results for future *DUVOL*, and results for future *NCSKEW* are presented in column (5)-(8). Two fixed-effect settings are shown in this table. Column (1), (2), (5), and (6) use firm-fixed effect and year-fixed effect to control for time- and firm- invariant factors. Columns (3), (4), (7), and (8) use firm-fixed effect and industry ×year-fixed effect. Industries are classified by Fama-French 48 industries. Standard errors are adjusted for heteroskedasticity and clustered by firm. The results of the full and simple model are shown in this table. *, **, and *** indicate significance at the 10%, 5%, and 1% level, respectively. Standard errors are shown in the parentheses.

TABLE 3
Robustness tests

Panel A: Alternative dependent variables

	(1)	(2)	(3)	(4)
Variables	*COUNT*	*COUNT_UP*	*COUNT_DOWN*	*CRASH*
ESV_RYANS	0.1139***	-0.0430*	0.0707***	0.0347***
	(0.0384)	(0.0237)	(0.0257)	(0.0117)
Constant	-2.6790***	1.7878***	-0.8930***	-0.4723***
	(0.3346)	(0.2112)	(0.2184)	(0.0999)
Controls	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes
Industry×Year FE	Yes	Yes	Yes	Yes
Observations	21,617	21,617	21,617	21,617
R-squared	0.2145	0.2329	0.2144	0.2079

Panel B: Alternative measures of sophisticated active attention

	(1)	(2)	(3)	(4)	(5)	(6)
Variables	*DUVOL*	*NCSKEW*	*DUVOL*	*NCSKEW*	*DUVOL*	*NCSKEW*
ESV_LM	0.0266***	0.0721***				
	(0.0092)	(0.0197)				
ESV_DRT			0.0241***	0.0624***		
			(0.0093)	(0.0199)		
ESV_FCOMP					0.0170***	0.0450***
					(0.0061)	(0.0131)
Constant	-0.9728***	-2.0000***	-0.9691***	-1.9720***	-0.7690***	-1.4508***
	(0.0807)	(0.1782)	(0.0842)	(0.1851)	(0.0666)	(0.1483)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes
Industry×Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	21,617	21,617	21,617	21,617	21,617	21,617
Adjusted R-squared	0.1083	0.0954	0.1082	0.0952	0.1082	0.0954

Panel C: Abnormal EDGAR searching volume

	(1)	(2)	(3)	(4)	(5)	(6)
Variables	*DUVOL*	*NCSKEW*	*DUVOL*	*NCSKEW*	*DUVOL*	*NCSKEW*
ABN_ESV_RYANS	0.0263***	0.0693***				
	(0.0099)	(0.0212)				
ABN_ESV_LM			0.0267***	0.0722***		
			(0.0092)	(0.0197)		
ABN_ESV_DRT					0.0242***	0.0625***
					(0.0093)	(0.0199)
Constant	-0.8169***	-1.5779***	-0.8135***	-1.5685***	-0.8181***	-1.5811***
	(0.0634)	(0.1419)	(0.0634)	(0.1419)	(0.0634)	(0.1418)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes
Industry×Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	21,617	21,617	21,617	21,617	21,617	21,617
R-squared	0.2560	0.2452	0.2560	0.2453	0.2560	0.2452

Notes: This table presents the results for robustness tests. Panel A shows the results using alternative measures of stock price crash risk. Two other widely mentioned measures of stock market crash risk, *COUNT* and *CRASH*, are used in these tests. The definitions of these variables are listed in Appendix Table A1. *COUNT* is further dissected into *COUNT_UP* and *COUNT_DOWN* in order to test the asymmetric effect of EDGAR searching volume on *COUNT*. In all the regressions in this table, *ESV_RYANS* is used as the measure of EDGAR searching volume. Columns (1)-(4) show the results when dependent variables are *COUNT*, *COUNT_UP*, *COUNT_DOWN*, and *CRASH*, respectively. Panel B of this table shows the results when independent variables are constructed based on alternative methods mentioned in prior studies. *ESV_LM* is the logarithm of 1 plus the searching volume from non-robust page viewers counted by the method established by Loughran and McDonald (2017) in a fiscal year. *ESV_DRT* is the logarithm of 1 plus the searching volume from non-robust page viewers counted by the method established by Drake et al. (2015) on the year basis. *ESV_FCOMP* is the first principal component score of *ESV_RYANS*, *ESV_DRT*, and *ESV_LM*. Panel C of this table shows the results when abnormal EDGAR searching volume measures are introduced as independent variables. Enlightened by Li and Sun (2018), measures of abnormal EDGAR searching volume are the residuals from Equation (3):

$$ESV_{i,t} = \delta_0 + \delta_1 SIZE_{i,t} + \delta_2 TOBINQ_{i,t} + \delta_3 NAN_{i,t} + \delta_4 IO_{i,t} + IndustryFE + \varepsilon_{i,t} \quad (3)$$

where *ESV* refers to EDGAR searching volume measures *ESV_RYANS*, *ESV_LM*, and *ESV_DRT*. We denote the residuals as *ABN_ESV_RYANS*, *ABN_ESV_LM*, *ABN_ESV_DRT* for *ESV_RYANS*, *ESV_LM*, and *ESV_DRT*, respectively. Dependent variables are *DUVOL* and *NCSKEW*. Firm-fixed effect and Industry ×year-fixed effect are used to control for the effect of firm-, year-, and industry-invariant factors. Industries are classified by Fama-French 48 Industries. Standard errors are adjusted for heteroskedasticity and clustered by firm. *, **, and *** indicate significance at the 10%, 5%, and 1% level, respectively. Standard errors are shown in the parentheses.

TABLE 4

Results by controlling for general attention based on Google searching index

Panel A: Summary statistics of Google searching index

Variables	N	MEAN	ST.DEV	P25	MEDIAN	P75
Google	21,617	0.1063	0.1403	0.0000	0.0542	0.1625

Panel B: Correlation between Google searching index and ESV

Variables	*(1) ESV_RYANS*	*(2) ESV_LM*	*(3) ESV_DRT*	*(4) ESV_FCOMP*	*(5) ABN_ESV_RYANS*	*(6) ABN_ESV_LM*	*(7) ABN_ESV_DRT*
Google	0.131***	0.142***	0.118***	0.131***	-0.064***	-0.052***	-0.072***

Panel C: Regression results for *DUVOL*

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
				DUVOL			
Variables	*ESV_RYANS*	*ESV_LM*	*ESV_DRT*	*ESV_FCOMP*	*ABN_ESV_RYANS*	*ABN_ESV_LM*	*ABN_ESV_DRT*
ESV	0.0262***	0.0267***	0.0242***	0.0170***	0.0264***	0.0268***	0.0242***
	(0.0099)	(0.0092)	(0.0093)	(0.0061)	(0.0099)	(0.0092)	(0.0093)
Google	-0.0055	-0.0064	-0.0047	-0.0057	-0.0055	-0.0063	-0.0047
	(0.0333)	(0.0333)	(0.0333)	(0.0333)	(0.0333)	(0.0333)	(0.0333)
Constant	-0.9779***	-0.9734***	-0.9696***	-0.7690***	-0.8170***	-0.8137***	-0.8182***
	(0.0857)	(0.0806)	(0.0841)	(0.0666)	(0.0633)	(0.0634)	(0.0633)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Industry×Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	21,617	21,617	21,617	21,617	21,617	21,617	21,617
R-squared	0.2560	0.2560	0.2560	0.2560	0.2560	0.2560	0.2560

Panel D: Regression results for *NCSKEW*

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
				NCSKEW			
Variables	*ESV_RYANS*	*ESV_LM*	*ESV_DRT*	*ESV_FCOMP*	*ABN_ESV_RYANS*	*ABN_ESV_LM*	*ABN_ESV_DRT*
ESV	0.0693***	0.0723***	0.0625***	0.0451***	0.0695***	0.0725***	0.0627***
	(0.0212)	(0.0198)	(0.0199)	(0.0131)	(0.0212)	(0.0198)	(0.0199)
Google	-0.0158	-0.0186	-0.0135	-0.0166	-0.0158	-0.0185	-0.0135
	(0.0720)	(0.0720)	(0.0719)	(0.0720)	(0.0720)	(0.0720)	(0.0719)
Constant	-2.0032***	-2.0020***	-1.9733***	-1.4507***	-1.5782***	-1.5689***	-1.5815***
	(0.1881)	(0.1779)	(0.1848)	(0.1482)	(0.1417)	(0.1418)	(0.1417)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Industry×Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	21,617	21,617	21,617	21,617	21,617	21,617	21,617
R-squared	0.2452	0.2453	0.2452	0.2453	0.2452	0.2453	0.2452

Notes: This table reports regression results when Google searching index (*Google*) is included as the control variable. Google searching index data come from Google Trends. *Google* is calculated as the average searching index of all months in a fiscal year, scaled by 100. Panel A of this table presents the summary statistics of *Google*. Panel B of this table shows the correlation between Google searching index and measures of EDGAR searching volume. *** shows significance at the 1% level. Panel C of this table presents regression results for *DUVOL*. Panel D of this table presents the regression results for *NCSKEW*. Control variables are the same as those in baseline regressions. We use firm-fixed effect and industry ×year-fixed effect in all the regressions from this table. *IndepVar* denotes dependent variables in our research. Industries are classified by Fama-French 48 industries. Standard errors, shown in the parentheses, are adjusted for heteroskedasticity and clustered by firm. *, **, and *** indicate significance at the 10%, 5%, and 1% level, respectively.

TABLE 5
Results by controlling for information supply

	(1)	(2)	(3)	(4)	(5)	(6)
Variables	*DUVOL*	*DUVOL*	*DUVOL*	*NCSKEW*	*NCSKEW*	*NCSKEW*
ESV_RYANS	0.0263***	0.0252**	0.0254**	0.0694***	0.0660***	0.0664***
	(0.0099)	(0.0099)	(0.0099)	(0.0212)	(0.0212)	(0.0212)
LN_NUMBER_TOTAL	0.0222		0.0218	0.0616		0.0601
	(0.0398)		(0.0398)	(0.0837)		(0.0836)
LN_VOL_8K		0.0178	0.0178		0.0598	0.0596
		(0.0179)	(0.0179)		(0.0389)	(0.0389)
Constant	-0.9929***	-0.9830***	-0.9984***	-2.0449***	-2.0208***	-2.0631***
	(0.0907)	(0.0863)	(0.0912)	(0.1979)	(0.1895)	(0.1991)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes
Industry×Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	21,617	21,617	21,617	21,617	21,617	21,617
R-squared	0.2560	0.2560	0.2560	0.2452	0.2453	0.2453

Notes: This table shows the robustness test after controlling for firms' information supply. The dependent variables are future *DUVOL* and future *NCSKEW*. The independent variable, *ESV_RYANS*, is the EDGAR searching volume calculated by Ryans (2017). We include the logarithm form of the total number of filings (*LN_NUMBER_TOTAL*) and the number of voluntary 8-K filings (*LN_VOL_8K*) identified as Lerman and Livnat (2010) to control for the effect of firms' voluntary disclosure. Columns (1)-(3) show the results of *DUVOL*. Columns (4)-(6) show the results of *NCSKEW*. Control variables are the same as those in the baseline regressions. Firm-fixed effect and Industry ×year-fixed effect are used to control for the effect of firm-, year-, and industry-invariant factors. Industries are classified by Fama-French 48 Industries. Standard errors are adjusted for heteroskedasticity and clustered by firm. *, **, and *** indicate significance at the 10%, 5%, and 1% level, respectively. Standard errors are shown in the parentheses.

TABLE 6
Different types of filings (10-X and 8-K&Others)

	(1)	(2)	(3)	(4)	(5)	(6)
		DUVOL			*NCSKEW*	
Variables	*10-K*	*10-Q*	*8-K&Others*	*10-K*	*10-Q*	*8-K&Others*
ESV_RYANS	0.0087	0.0112*	0.0225***	0.0199	0.0258*	0.0587***
	(0.0061)	(0.0063)	(0.0083)	(0.0130)	(0.0137)	(0.0180)
Constant	-0.8572***	-0.8628***	-0.9405***	-1.6711***	-1.6848***	-1.8997***
	(0.0682)	(0.0676)	(0.0766)	(0.1529)	(0.1515)	(0.1697)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes
Industry × Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	21,617	21,617	21,617	21,617	21,617	21,617
R-squared	0.2557	0.2558	0.2560	0.2448	0.2449	0.2452

Notes: This table shows the effect of sophisticated attention on different types of filings, 10-K, 10-Q, and 8-K&Others. In each filing group, the independent variable *ESV_RYANS* is measured by Ryans (2018). Dependent variables are future *DUVOL* and future *NCSKEW*. Control variables are the same as those used in the baseline model. Firm-fixed effect and Industry ×year-fixed effect are used to control for the effect of firm-, year-, and industry-invariant factors. Industries are classified by Fama-French 48 Industries. Standard errors, shown in the parentheses, are adjusted for heteroskedasticity and clustered by firm. *, **, and *** indicate significance at the 10%, 5%, and 1% level, respectively.

TABLE 7
Heterogeneity analysis

Panel A: Beating analyst expectation

	(1)	(2)	(1)	(2)
	DUVOL		*NCSKEW*	
Variables	Beat	Not Beat	Beat	Not Beat
ESV_RYANS	0.0028	0.0491***	0.0220	0.1201***
	(0.0173)	(0.0165)	(0.0366)	(0.0351)
Constant	-0.8525***	-1.1769***	-1.8850***	-2.3425***
	(0.1491)	(0.1517)	(0.3213)	(0.3320)
Controls	Yes	Yes	Yes	Yes
Ind×Year FE	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes
Equality Test	-0.046***		-0.098***	
"High=Low"	[0.00]		[0.00]	
Observations	7,483	9,168	7,483	9,168
R-squared	0.3930	0.3237	0.3935	0.3165

Panel B: Investment opportunity

	(1)	(2)	(1)	(2)
	DUVOL		*NCSKEW*	
Variables	High	Low	High	Low
ESV_RYANS	0.0341**	0.0156	0.0858**	0.0460
	(0.0158)	(0.0146)	(0.0342)	(0.0312)
Constant	-1.1474***	-1.0022***	-2.2903***	-2.0742***
	(0.1419)	(0.1383)	(0.3166)	(0.2962)
Controls	Yes	Yes	Yes	Yes
Ind×Year FE	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes
Equality Test	0.018*		0.039**	
"High=Low"	[0.06]		[0.03]	
Observations	10,426	10,457	10,426	10,457
R-squared	0.3055	0.3251	0.2944	0.3186

Panel C: Financial constraint

	(1)	(2)	(1)	(2)
	DUVOL		*NCSKEW*	
Variables	High	Low	High	Low
ESV_RYANS	0.0553***	0.0054	0.1139***	0.0373
	(0.0154)	(0.0132)	(0.0338)	(0.0269)
Constant	-1.1775***	-0.8172***	-2.2658***	-1.8689***
	(0.1159)	(0.1441)	(0.2560)	(0.3013)
Controls	Yes	Yes	Yes	Yes
Ind×Year FE	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes
Equality Test	0.049***		0.076***	
"High=Low"	[0.00]		[0.00]	
Observations	10,405	10,447	10,405	10,447
R-squared	0.3245	0.2687	0.3095	0.2692

Notes: This table presents the heterogeneity analysis of the effect of sophisticated attention on stock price crash risk by whether beating analyst expectation or not, investment opportunity, and financial constraint level. Panel A shows the results of subsample tests by whether beating analyst expectation or not. Analyst expectation is measured as analysts' expected EPS consensus for annual reports. Panel B shows the results of subsample tests by investment opportunity. Investment opportunity is measured by Tobin's q. Panel C shows the results of subsample tests by financial constraint level. Financial constraint level is measured by WW-index (Whited and Wu 2006). In panel A, a firm is included in Beat (Not-beat) group if its EPS is higher (lower) than analysts' consensus EPS. In panels B and C, a firm is included in High (Low) group if its value of the measure is higher (lower) than the median value of the measure over fiscal years. In all the regressions shown in this table, *ESV_RYANS* is used as the measure of EDGAR searching volume. Control variables are the same as those in baseline regressions. Firm-fixed effect and industry ×year-fixed effect are used to control for the effect of firm-, year-, and industry-invariant factors. Industries are classified by Fama-French 48 Industries. Standard errors, shown in the parentheses, are adjusted for heteroskedasticity and clustered by firm. *, **, and *** indicate significance at the 10%, 5%, and 1% level, respectively.

TABLE 8
Sophisticated active attention and ex-ante crash risk

	(1)	(2)	(3)	(4)	(5)	(6)
Variables	*IV_SKEW*	*IV_SKEW*	*IV_SKEW*	*IV_SKEW*	*IV_SKEW*	*IV_SKEW*
ESV_RYANS	0.0058***	0.0052***				
	(0.0012)	(0.0012)				
ESV_DRT			0.0049***	0.0045***		
			(0.0011)	(0.0011)		
ESV_LM					0.0055***	0.0050***
					(0.0011)	(0.0011)
SIZE	0.0060***	0.0071***	0.0062***	0.0073***	0.0059***	0.0070***
	(0.0012)	(0.0013)	(0.0012)	(0.0013)	(0.0012)	(0.0013)
TOBINQ	0.0048***	0.0047***	0.0049***	0.0047***	0.0048***	0.0047***
	(0.0005)	(0.0004)	(0.0005)	(0.0004)	(0.0005)	(0.0004)
CASH_FLOW	0.0109***	0.0110***	0.0108***	0.0109***	0.0109***	0.0110***
	(0.0029)	(0.0029)	(0.0029)	(0.0029)	(0.0029)	(0.0029)
BLEV	0.0016	0.0024	0.0018	0.0026	0.0016	0.0025
	(0.0035)	(0.0033)	(0.0035)	(0.0033)	(0.0035)	(0.0033)
CAPX	0.0395***	0.0356***	0.0394***	0.0355***	0.0398***	0.0357***
	(0.0106)	(0.0106)	(0.0107)	(0.0107)	(0.0107)	(0.0107)
GSALE	0.0000	-0.0002	0.0000	-0.0002	0.0001	-0.0002
	(0.0008)	(0.0008)	(0.0008)	(0.0008)	(0.0008)	(0.0008)
NAN	0.0026***	0.0021**	0.0026***	0.0021**	0.0027***	0.0021**
	(0.0009)	(0.0009)	(0.0009)	(0.0009)	(0.0009)	(0.0009)
IO	0.0177***	0.0177***	0.0179***	0.0178***	0.0180***	0.0179***
	(0.0036)	(0.0037)	(0.0036)	(0.0037)	(0.0036)	(0.0037)
RET	0.0289***	0.0262***	0.0292***	0.0265***	0.0287***	0.0261***
	(0.0069)	(0.0072)	(0.0069)	(0.0072)	(0.0069)	(0.0072)
SIGMA	0.3116***	0.2629***	0.3154***	0.2659***	0.3071***	0.2594***
	(0.0626)	(0.0657)	(0.0627)	(0.0657)	(0.0627)	(0.0657)
DTURN	0.0014***	0.0015***	0.0014***	0.0015***	0.0013***	0.0014***
	(0.0003)	(0.0003)	(0.0003)	(0.0003)	(0.0003)	(0.0003)
NCSKEW	-0.0007**	-0.0005*	-0.0007**	-0.0005*	-0.0007**	-0.0005*
	(0.0003)	(0.0003)	(0.0003)	(0.0003)	(0.0003)	(0.0003)

OPAQUE_MJONES	0.0076	0.0072	0.0076	0.0073	0.0076	0.0072
	(0.0081)	(0.0081)	(0.0081)	(0.0082)	(0.0081)	(0.0082)
Constant	-0.1046***	-0.1042***	-0.1004***	-0.1007***	-0.1012***	-0.1010***
	(0.0118)	(0.0118)	(0.0113)	(0.0114)	(0.0114)	(0.0114)
Year FE	Yes	No	Yes	No	Yes	No
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes
Industry×Year FE	No	Yes	No	Yes	No	Yes
Observations	10,752	10,716	10,752	10,716	10,752	10,716
R-squared	0.5658	0.6016	0.5655	0.6013	0.5659	0.6016

Notes: This table shows the association between sophisticated attention and ex-ante crash risk. The regression is given as Equation (4):

$$Ex - ante\ CrashRisk_{i,t} = \alpha + \beta_1 ESV_{i,t} + \gamma Controls + FE + \varepsilon_{i,t} \tag{4}$$

Ex-ante crash risk is measured as firms' implied volatility smirk used in Kim et al. (2019). The implied volatility smirk is calculated as the difference between the implied volatility of the OTM put option and implied volatility of the ATM call option. The option data come from OptionMetrics. Sophisticated attention is measured as the EDGAR searching volume: *ESV_RYANS*, *ESV_DRT*, and *ESV_LM*. Columns (1) and (2) show the results of *ESV_RYANS*. Columns (3) and (4) show the results of *ESV_DRT*. Columns (5) and (6) show the results of *ESV_LM*. Control variables are the same as those in baseline regressions. Firm-fixed effect and industry ×year-fixed effect are used to control for the effect of firm-, year-, and industry-invariant factors. Industries are classified by Fama-French 48 Industries. Standard errors, shown in the parentheses, are adjusted for heteroskedasticity and clustered by firm. *, **, and *** indicate significance at the 10%, 5%, and 1% level, respectively.

TABLE 9

Sophisticated active attention and managerial disclosure of bad news: evidence from management guidance

	(1)	(2)	(3)	(4)	(5)	(6)
Variables	*FREQ_BAD_MG*	*FREQ_BAD_MG*	*FREQ_BAD_MG*	*FREQ_BAD_MG*	*FREQ_BAD_MG*	*FREQ_BAD_MG*
ESV_RYANS	-0.8304***	-0.8996***				
	(0.2885)	(0.2675)				
ESV_DRT			-0.7933***	-0.8971***		
			(0.2737)	(0.2553)		
ESV_LM					-0.8868***	-0.9222***
					(0.2621)	(0.2463)
Constant	2.2425	3.1246	2.1632	3.2753	2.3556	3.0457
	(2.5574)	(2.4426)	(2.5564)	(2.4583)	(2.4092)	(2.3003)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	No	Yes	No	Yes	No	Yes
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes
Industry×Year FE	Yes	No	Yes	No	Yes	No
Observations	5,366	5,366	5,366	5,366	5,366	5,366
R-squared	0.7323	0.7027	0.7323	0.7028	0.7326	0.7030

Notes: This table presents the relation between sophisticated active attention and managerial disclosure of bad news in management guidance. The dependent variable, *FREQ_BAD_MG*, is the number of bad-news management guidance in each firm-year, where bad-news is defined as management guidance lower than the most recent consensus analyst forecast. Information on management guidance and analyst forecast comes from IBES Database. EDGAR searching volume measures (*ESV_RYANS*, *ESV_LM*, and *ESV_DRT*) are used as independent variables. Control variables are the same as those in baseline regressions. Two fixed-effect settings are shown in this table. Columns (1), (3), and (5) use Firm-fixed effect and Industry ×year-fixed effect. Columns (2), (4), and (6) use firm-fixed effect and year-fixed effect to control for time- and firm- invariant factors. Industries are classified by Fama-French 48 industries. Standard errors are adjusted for heteroskedasticity and clustered by firm. The results of the full and simple model are shown in this table. *, **, and *** indicate significance at the 10%, 5%, and 1% level, respectively. Standard errors are shown in the parentheses.

TABLE 10
Evidence from financial reporting quality and accounting conservatism

Panel A: ESV and earnings management

	(1)	(2)	(3)	(4)	(5)	(6)
Variables	*ABSDA_MJONES*	*ABSDA_KOTHARI*	*ABSRM*	*ABSRM_PROD*	*ABSRM_DISC_EXP*	*ABSRM_OANCF*
ESV_RYANS	0.0117***	0.0151***	0.0186***	0.0146***	0.0034	0.0181***
	(0.0021)	(0.0021)	(0.0054)	(0.0047)	(0.0021)	(0.0040)
Constant	-0.0213	0.0106	0.1749***	0.1432***	0.0575***	-0.0198
	(0.0219)	(0.0213)	(0.0483)	(0.0432)	(0.0194)	(0.0330)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes
Industry×Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	21,617	21,617	13,815	13,930	13,815	13,930
R-squared	0.3786	0.3961	0.7091	0.7089	0.6372	0.5830

Panel B: ESV and accrual quality

	(1)	(2)	(3)	(4)	(5)	(6)
Variables	*AQ_DD*	*AQ_MDD*	*AQ_FLOS*	*WDA*	*SPI*	*△GDWL*
ESV_RYANS	0.0125***	0.0166***	0.0047***	-0.0003**	-0.0105***	0.4265***
	(0.0038)	(0.0040)	(0.0014)	(0.0001)	(0.0033)	(0.1281)
Constant	0.2052***	0.1937***	0.0729***	0.0022**	0.0674**	-5.4650***
	(0.0343)	(0.0373)	(0.0132)	(0.0009)	(0.0263)	(1.1575)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes
Industry × Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	21,212	21,212	18,492	21,342	21,353	17,676
R-squared	0.5707	0.5183	0.7064	0.2794	0.3576	0.1970

Panel C: ESV and accounting conservatism

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
				C-SCORE			
Variables	*ESV_RYANS*	*ESV_LM*	*ESV_DRT*	*ESV_FCOMP*	*ABN_ESV_RYANS*	*ABN_ESV_LM*	*ABN_ESV_DRT*
ESV	-0.0030***	-0.0022***	-0.0030***	-0.0018***	-0.0030***	-0.0022***	-0.0030***
	(0.0008)	(0.0008)	(0.0008)	(0.0005)	(0.0008)	(0.0008)	(0.0008)
Constant	0.2243***	0.2188***	0.2250***	0.2008***	0.2058***	0.2055***	0.2059***
	(0.0073)	(0.0070)	(0.0071)	(0.0054)	(0.0052)	(0.0052)	(0.0052)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Industry×Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	21,617	21,617	21,617	21,617	21,617	21,617	21,617
R-squared	0.4621	0.4619	0.4622	0.4621	0.4621	0.4619	0.4622

Notes: This table shows the relation between sophisticated active attention and accounting information quality as well as accounting conservatism. Panel A presents the impact of sophisticated attention on accrual-based and real earnings management. The first two columns show the effect of sophisticated attention on accrual-based earnings management. Dependent variables in Columns (1) and (2) are the absolute value of discretionary accruals based on modified Jones model and Kothari et al. (2006). The rest columns show the effect of sophisticated attention on real earnings management. Column (3) presents the results for the absolute value of real earnings management (*ABSRM*) based on Zang (2012). Columns (4)-(6) show the results for the absolute value of real earnings management on production (*ABSRM_PROD*), discretionary expenses (*ABSRM_DISC_EXP*), and operating cash flow (*ABSRM_OANCF*), respectively. Panel B presents the impact of sophisticated attention on accrual quality. Six measures of accrual quality are used in this analysis. *AQ_DD* is the standard deviation of firm-level residual from Dechow and Dichev's (2002) model. *AQ_MDD* is the standard deviation of firm-level residual from modified Dechow and Dichev's (2002) model proposed by McNichols (2002). *AQ_FLOS* is the five-year standard deviation of residuals from the model proposed by Francis et al. (2005). *WDA* is defined as 100 times the total write-down scaled by total assets. *SPI* is 100 times the special item scaled by total assets. $\triangle$*GDWL* denotes 100 times the increasing value of goodwill scaled by total assets. Column (1)-(6) show the results for *AQ_DD*, *AQ_MDD*, *AQ_FLOS*, *WDA*, *SPI*, and $\triangle$*GDWL*, respectively. Panel C presents the effect of sophisticated attention on accounting conservatism. We use Khan and Watt's (2009) firm-year accounting conservatism measure (*C-SCORE*) as the dependent variable. Panel C presents the effect of sophisticated attention on accrual quality. Control variables include firm size (*SIZE*), Tobin's Q(*TOBINQ*), cash flow (*CASH_FLOW*), capital expenditure (*CAPX*), growth of sales (*GSALES*), analyst coverage (*NAN*), and institutional ownership (*IO*). Columns (1)-(6) show the results for various measures of EDGAR searching volume. Firm-fixed effect and Industry ×year-fixed effect are used to control for the effect of firm-, year-, and industry-invariant factors. Industries are classified by Fama-French 48 Industries. Standard errors, shown in the parentheses, are adjusted for heteroskedasticity and clustered by firm. *, **, and *** indicate significance at the 10%, 5%, and 1% level, respectively.

TABLE 11
The role of executive compensation

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Variables	*DUVOL*	*DUVOL*	*DUVOL*	*DUVOL*	*NCSKEW*	*NCSKEW*	*NCSKEW*	*NCSKEW*
ESV_RYANS	0.0387**	0.0397**	0.0228	0.0255	0.1078***	0.1100***	0.0732**	0.0783**
	(0.0168)	(0.0172)	(0.0173)	(0.0175)	(0.0357)	(0.0363)	(0.0363)	(0.0369)
ESV_RYANS×Ln(1+Option)	-0.0038**				-0.0083**			
	(0.0019)				(0.0040)			
Ln(1+Option)	0.0320**				0.0705**			
	(0.0157)				(0.0332)			
ESV_RYANS×Ln(1+CEO_Option)		-0.0034*				-0.0077**		
		(0.0018)				(0.0038)		
Ln(1+CEO_Option)		0.0268*				0.0624**		
		(0.0143)				(0.0305)		
ESV_RYANS×Ln(1+Noneq)			0.0029*				0.0062*	
			(0.0017)				(0.0036)	
Ln(1+Noneq)			-0.0266*				-0.0570*	
			(0.0149)				(0.0313)	
ESV_RYANS×Ln(1+CEO_Noneq)				0.0026*				0.0058*
				(0.0015)				(0.0031)
Ln(1+CEO_Noneq)				-0.0228*				-0.0506*
				(0.0128)				(0.0272)
Constant	-1.2550***	-1.2565***	-1.1225***	-1.1394***	-2.6191***	-2.6199***	-2.3314***	-2.3539***
	(0.1522)	(0.1565)	(0.1524)	(0.1557)	(0.3300)	(0.3379)	(0.3246)	(0.3328)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Industry×Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	9,397	9,260	9,397	9,260	9,397	9,260	9,397	9,260
R-squared	0.2338	0.2351	0.2338	0.2351	0.2300	0.2315	0.2300	0.2315

Notes: This table shows the role of executive incentives on the relation between sophisticated active attention and stock price crash risk. The dependent variables are *DUVOL* and *NCSKEW*. The independent variable is the EDGAR searching volume measured as Ryans (2017). Columns (1)-(4) shows the results of *DUVOL*. Columns (5)-(8) show the results of *NCSKEW*. Executive compensation data are from Execucomp. Executive-average option compensation based on book value (*Ln(1+Option)*), Non-equity incentive plan compensation (*Ln(1+Noneq)*) ,CEO option compensation based on book value *Ln(1+CEO_Option)*, and CEO non-equity incentive plan compensation, *Ln(1+CEO_Noneq)* are used in this analysis. Control variables are the same as those in baseline regressions. Firm-fixed effect and Industry ×year-fixed effect are used to control for the effect of firm-, year-, and industry-invariant factors. Industries are classified by Fama-French 48 Industries. Standard errors, shown in the parentheses, are adjusted for heteroskedasticity and clustered by firm. *, **, and *** indicate significance at the 10%, 5%, and 1% level, respectively.

TABLE 12
Natural experiment: adoption of XBRL

	(1)	(2)	(3)	(4)	(5)	(6)
Variables	*DUVOL*	*DUVOL*	*DUVOL*	*NCSKEW*	*NCSKEW*	*NCSKEW*
TREAT×POST	0.0675***		0.0609***	0.1287***		0.1145***
	(0.0178)		(0.0192)	(0.0379)		(0.0412)
$TREAT \times BEFORE^{-1}$		0.0332			0.0465	
		(0.0254)			(0.0532)	
TREAT×CURRENT		0.0847***			0.1809***	
		(0.0279)			(0.0597)	
$TREAT \times AFTER^{1}$		0.0999***			0.1785***	
		(0.0291)			(0.0609)	
$TREAT \times AFTER^{2\&3}$		0.0641**			0.1022*	
		(0.0261)			(0.0566)	
Constant	-0.0146**	-0.0215***	-0.8645***	0.0580***	0.0487***	-1.5850***
	(0.0057)	(0.0082)	(0.1552)	(0.0122)	(0.0172)	(0.3513)
Controls	No	No	Yes	No	No	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	6,087	6,087	6,087	6,087	6,087	6,087
R-squared	0.2126	0.2130	0.2567	0.2036	0.2040	0.2424

Notes: This table presents the results of difference-in-differences tests examining the causal impact of sophisticated attention on stock price crash risk. We choose the Tier 3 adoption of eXtensible Business Reporting Language (XBRL) in 2011 as the exogenous shock. According to the literature on liquidity (e.g., Fang et al., 2014), we use a similar method to construct treatment and control groups. Firms that receive higher increased searching volume during the period 2010-2012 (top half) are included in the treatment group, and firms that receive lower increased searching volume during the same period (bottom half) are included in the control group. We apply propensity score matching between treatment and control group using the observations in the fiscal year 2010, matching for fundamental variables including the control variables used in regression analysis, the first principal component score *ESV_FCOMP* in the independent variables, and dependent variables *DUVOL* and *NCSKEW*. Each observation in the treatment group is matched to one observation in the control group by the nearest-neighbor criterion. Standard errors, shown in the parentheses, are adjusted for heteroskedasticity and clustered by firm. *, **, and *** indicate significance at the 10%, 5%, and 1% level, respectively.

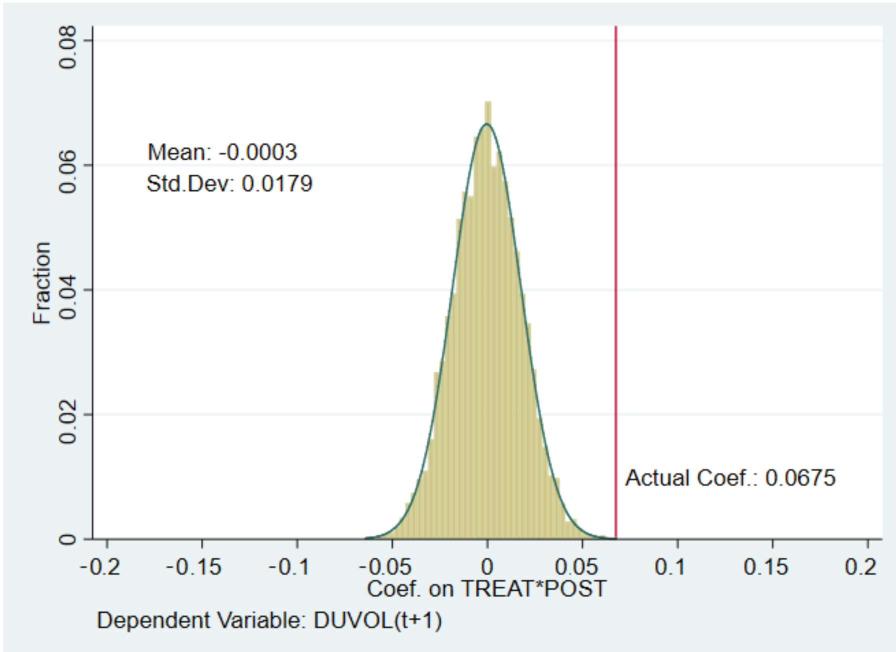
TABLE 13
Natural experiment: shareholder distraction

	(1)	(2)	(3)	(4)
Variables	*NCSKEW*	*NCSKEW*	*DUVOL*	*DUVOL*
DISTRACTION	-0.8309***	-1.1208***	-0.3929***	-0.4910***
	(0.2553)	(0.3400)	(0.1261)	(0.1675)
SIZE		0.1558***		0.0756***
		(0.0202)		(0.0095)
TOBINQ		0.0840***		0.0416***
		(0.0102)		(0.0047)
CASH_FLOW		0.0914		0.0768*
		(0.0881)		(0.0399)
BLEV		-0.1154		-0.0578*
		(0.0711)		(0.0330)
CAPX		0.1924		0.0945
		(0.1548)		(0.0767)
GSALE		0.0433*		0.0185
		(0.0255)		(0.0113)
NAN		0.0466**		0.0226**
		(0.0185)		(0.0088)
IO		0.0120		0.0294
		(0.0527)		(0.0245)
RET		22.4888		10.4721
		(14.2148)		(6.6761)
SIGMA		1.1529		0.4298
		(1.1229)		(0.5344)
DTURN		0.0228***		0.0102***
		(0.0077)		(0.0036)
NCSKEW		-0.1243***		-0.0567***
		(0.0100)		(0.0046)
OPAQUE		0.2411		0.0985
		(0.1786)		(0.0809)
Constant	0.2162***	-1.0962***	0.0524**	-0.5992***
	(0.0422)	(0.1679)	(0.0209)	(0.0788)
Firm FE	Yes	Yes	Yes	Yes
Industry×Year FE	Yes	Yes	Yes	Yes
Observations	15,939	15,939	15,939	15,939
R-squared	0.2229	0.2610	0.2217	0.2608

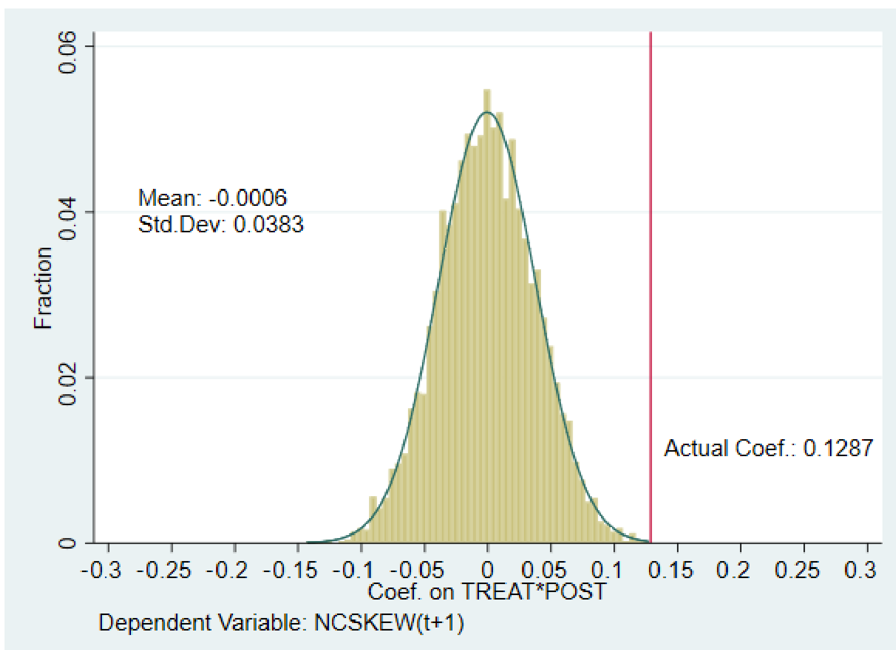
Notes: This table shows the results of the effect of shareholder distraction on firms' stock price crash risk in the natural experiment based on Kempf et al. (2017). The shareholder distraction measure is from Kempf et al. (2017). We use the year-average shareholder distraction (*DISTRACTION*) as the dependent variable. The dependent variables are future *DUVOL* and future *NCSKEW*. Control variables are the same as those in baseline regressions. Firm-fixed effect and Industry ×year-fixed effect are used to control for the effect of firm-, year-, and industry-invariant factors. Industries are classified by Fama-French 48 Industries. Columns (1) and (2) show results for future *DUVOL*. Columns (3) and (4) report results for future *NCSKEW*. Standard errors, shown in the parentheses, are adjusted for heteroskedasticity and clustered by firm. *, **, and *** indicate significance at the 10%, 5%, and 1% level, respectively.

FIGURE 1 Placebo tests

Panel A: *DUVOL*



Panel B: *NCSKEW*



Notes: This figure shows the histogram of the coefficients on *TREAT* ×*POST* from 5,000 bootstrap simulations of difference-in-differences analysis. Specifically, we first randomly select a group of observations as a pseudo treatment group and then use the same method to construct a pseudo control group. The group size of pseudo-treated and control groups should be the same as our real difference-in-differences analysis. We require that firms should not have missing value in 2010 to make sure that the experiment is processed in the same way as Table 12. Then, we replicate columns (1) and (4) of Table 12 using the pseudo samples for 5,000 times and record the coefficient on *TREAT* ×*POST*. Panel A of this figure shows the results when the dependent variable is future *DUVOL*. Panel B of this figure shows the results when the dependent variable is future *NCSKEW*. In each panel, the red line paralleled to the y-axis shows the actual results given in Table 12. The mean and standard deviation are shown in each panel.

APPENDIX TABLE A1
Variable definitions and data sources

Variable	Definition	Source
Dependent variables		
DUVOL	Down-to-up volatility, which calculates the logarithm of the ratio of standard deviation in weeks with negative returns to the standard deviation in weeks with positive returns in a year. $$DUVOL = ln\left[\left(\frac{1}{(n_d-1)}\sum_{Down} W_{it}^2\right)/\left(\frac{1}{(n_u-1)}\sum_{Up} W_{it}^2\right)\right]$$ where n_d and n_u are the number of weeks with negative returns and weeks with positive returns, respectively. $W_{i,t}$ is the firm-specific return under Equation (1).	CRSP
NCSKEW	The negative of the ratio of the third momentum of firm-specific return to its standard deviation raised to the third power, as shown below: $$NCSKEW = -\frac{n(n-1)^{3/2}\sum W_{i,t}^3}{(n-1)(n-2)(\sum W_{i,t}^2)^{3/2}}$$ where $W_{i,t}$ is the firm-specific return under Equation (1).	CRSP
COUNT	The balance of extremely negative and positive returns. Following Jin and Myers (2006), *COUNT* is calculated as the difference between the frequency of firm-specific returns falling 3 times of its standard deviation or more below the average return within the fiscal year and the frequency of firm-specific returns falling 3 times of its standard deviation or more above the average return within the fiscal year.	CRSP
CRASH	A dummy variable indicating the extreme losses in a firm-year. The value of *CRASH* equals to 1 if the firm experiences a firm-specific return falling 3 times or more of its standard deviation below the average return over the years, and equals to 0 otherwise.	CRSP
IV_SKEW	The difference between the implied volatility of the Out-of-the-Money put option and implied volatility of the At-the-Money call option.	Option Metrics
FREQ_BAD_MG	The number of bad-news management guidance in each firm-year, where bad-news is defined as management guidance lower than the most recent consensus analyst forecast.	I\B\E\S
Independent variables		
ESV_RYANS	The logarithm of 1 plus *ryans*, namely *ESV_RYANS*=ln(1+*ryans*), where *ryans* is the number of page views according to Ryans (2017) on year basis.	EDGAR Log File
ESV_DRT	The logarithm of 1 plus *drt*, namely *ESV_DRT*=ln(1+*drt*), where *drt* is the number of page views according to Drake et al. (2015) on year basis.	EDGAR Log File

ESV_LM	The logarithm of 1 plus *lm*, namely *ESV_LM*=ln(1+*lm*), where *lm* is the number of page views according to Loughran and McDonald (2014) on year basis.	EDGAR Log File
ESV_FCOMP	The first principal component score of *ESV_RYANS*, *ESV_DRT*, and *ESV_LM*.	EDGAR Log File
ABN_ESV_RYANS	Residuals of equation (2) when the independent variable is *ESV_RYANS*.	EDGAR Log File
ABN_ESV_LM	Residuals of equation (2) when the independent variable is *ESV_LM*.	EDGAR Log File
ABN_ESV_DRT	Residuals of equation (2) when the independent variable is *ESV_DRT*.	EDGAR Log File
Control variables		
SIZE	The logarithm of 1 plus the book value of total assets (*#AT*).	COMPUSTAT
TOBINQ	The ratio of market value to book value (*#AT*), where market value is defined as total assets (*#AT*) minus common equity (*#CEQ*) and deferred taxes (*#TXDB*) plus the market equity *(#PRCC_F* × *#CSHO*).	COMPUSTAT
CASH FLOW	The sum of income before extraordinary items (*#IB*) and depreciation and amortization (*#DP*) scaled by the book value of total assets (*#AT*).	COMPUSTAT
BLEV	Long-term debt (*#DLTT*) and debt in current liabilities (*#DLC*) scaled by the book value of total assets (*#AT*).	COMPUSTAT
CAPX	Capital Expenditures (*#CAPX*) scaled by the book value of total assets (*#AT*).	COMPUSTAT
GSALE	The increase of sales (*#SALE*) scaled by the book value of total assets (*#AT*).	COMPUSTAT
NAN	The logarithm of 1 plus the arithmetic mean of the 12 monthly numbers of analysts following a firm in a fiscal year.	I\B\E\S
IO	Institutional ownership measured by the percent of share held by institutional investors. (*INSTOWN_PERC* in WRDS Thomson Reuters Institutional (13f) Holdings Stock Ownership Summary File).	Thomson Reuters Institutional (13f) Holdings
RET	The mean of weekly firm-specific return over the years times 100.	CRSP
SIGMA	The standard deviation of weekly firm-specific returns over the years.	CRSP
DTURN	The average change in monthly share turnover over the years, where turnover is defined as the monthly trading volume scaled by the total number of shares outstanding.	CRSP

OPAQUE	The 3-year moving average in the prior 3 years of the absolute value of discretionary accruals based on modified Jones model.	COMPUSTAT
Other variables		
WW-Index	Financial constraint status is measured by Whited and Wu (2006).	COMPUSTAT
LN_NUMBER_TOTAL	The logarithm form of total number of filings.	EDGAR
LN_NUMBER_8K	The logarithm form of the number of 8-K filings identified as Lerman and Livnat (2010).	EDGAR
Ln(1+Option)	Executive-average option compensation based on book value.	Excucomp
Ln(1+CEO_Option)	CEO option compensation based on book value.	Excucomp
Ln(1+Noneq)	Executive-average non-equity incentive plan compensation.	Excucomp
Ln(1+CEO_Noneq)	CEO non-equity incentive plan compensation.	Excucomp
ABSDA_MJONES	Absolute value of discretionary accruals based on modified Jones model (Dechow et al. (1995)).	COMPUSTAT
ABSDA_KOTHARI	Absolute value of discretionary accruals based on performance matched model (Kothari et al. (2006)).	COMPUSTAT
ABSRM	Absolute value of real earnings management based on Zang (2012).	COMPUSTAT
ABSRM_PROD	Absolute value of real earnings management on production.	COMPUSTAT
ABSRM_DISC_EXP	Absolute value of real earnings management on discretionary expenses.	COMPUSTAT
ABSRM_OANCF	Absolute value of real earnings management on operating cash flow.	COMPUSTAT
C-SCORE	Khan and Watt's (2009) firm-year accounting conservatism measure.	COMPUSTAT /CRSP
AQ_DD	The standard deviation of firm-level residual from Dechow and Dichev's (2002) model.	COMPUSTAT
AQ_MDD	The standard deviation of firm-level residual from modified Dechow and Dichev's (2002) model proposed by McNichols (2002).	COMPUSTAT
AQ_FLOS	The five-year standard deviation of residuals from the model proposed by Francis et al. (2005).	COMPUSTAT
WDA	100 times the total write-down (*#WDA*) scaled by total assets (*#AT*).	COMPUSTAT
SPI	100 times the special item (*#SPI*) scaled by total assets (*#AT*).	COMPUSTAT
△GDWL	100 times the increasing value of goodwill (*#GDWL*) scaled by	COMPUSTAT

	total assets (*#AT*).	
Google	The average searching index of all months in a fiscal year, scaled by 100.	Google Searching Index

APPENDIX TABLE A2
Two-way distributions among measures of EDGAR searching volume

Panel A: *ESV_RYANS* and *ESV_LM*

		ESV_RYANS										
		Low	2	3	4	5	6	7	8	9	High	Total
ESV_LM	Low	1,869	281	6	0	0	0	0	0	0	0	2,156
	2	272	1,441	413	35	2	0	0	0	0	0	2,163
	3	9	390	1,220	485	57	2	0	0	0	0	2,163
	4	6	36	466	1,103	479	70	4	0	0	0	2,164
	5	0	6	35	470	1,091	492	60	5	1	0	2,160
	6	0	6	16	41	453	1,066	527	54	1	0	2,164
	7	0	3	5	18	55	467	1,143	452	22	0	2,165
	8	0	0	2	11	16	52	399	1,326	355	1	2,162
	9	0	0	0	1	6	15	31	322	1,582	207	2,164
	High	0	0	0	0	1	0	1	3	203	1,948	2,156
	Total	2,156	2,163	2,163	2,164	2,160	2,164	2,165	2,162	2,164	2,156	21,617

Panel B: *ESV_RYANS* and *ESV_DRT*

		ESV_RYANS										
		Low	2	3	4	5	6	7	8	9	High	Total
ESV_DRT	Low	2,007	148	1	0	0	0	0	0	0	0	2,156
	2	148	1,778	230	7	0	0	0	0	0	0	2,163
	3	1	235	1,628	281	15	3	0	0	0	0	2,163
	4	0	2	303	1,557	274	21	4	2	0	1	2,164
	5	0	0	1	312	1,553	274	17	2	0	1	2,160
	6	0	0	0	6	316	1,522	292	28	0	0	2,164
	7	0	0	0	0	1	337	1,560	250	15	2	2,165
	8	0	0	0	0	1	6	289	1,640	224	2	2,162
	9	0	0	0	1	0	1	3	240	1,784	135	2,164
	High	0	0	0	0	0	0	0	0	141	2,015	2,156
	Total	2,156	2,163	2,163	2,164	2,160	2,164	2,165	2,162	2,164	2,156	21,617

Panel C: *ESV_DRT* and *ESV_LM*

		ESV_LM										
		Low	2	3	4	5	6	7	8	9	High	Total
ESV_DRT	Low	1,894	236	12	4	2	5	2	1	0	0	2,156
	2	260	1,544	304	30	7	4	7	6	1	0	2,163
	3	2	364	1,384	355	22	10	8	14	4	0	2,163
	4	0	19	431	1,257	383	31	13	16	14	0	2,164
	5	0	0	31	482	1,213	352	36	22	21	3	2,160
	6	0	0	1	35	486	1,232	338	30	40	2	2,164
	7	0	0	0	0	45	490	1,291	284	43	12	2,165
	8	0	0	0	1	2	39	449	1,426	230	15	2,162
	9	0	0	0	0	0	1	20	363	1,608	172	2,164
	High	0	0	0	0	0	0	1	0	203	1,952	2,156
	Total	2,156	2,163	2,163	2,164	2,160	2,164	2,165	2,162	2,164	2,156	21,617

Notes: This table presents the two-way comparison among measures of EDGAR searching volume. There are 3 measures used in this study: (1) *lm* from Loughran and MacDonald (2017); (1) *drt* from Drake et al. (2015); (3) and *ryans* from Ryans (2017). Loughran and MacDonald (2017) identify non-robust page viewers under the assumption that human does not download more than 50 items in a day. Furthermore, Drake et al. (2015) require that human does not download more than five items per minute. Similar to these criteria, Ryans (2019) looses the two above-mentioned criteria to 500 items/day and 25 items/minute and introduces another restriction that human does not search more than 3 firms in a minute. In order to eliminate the skewness of counting variables, we use the logarithm form of these counting variables as our proxies of sophisticated attention. Specifically, *ESV_LM*, *ESV_DRT*, *ESV_RYANS* are defined as the logarithm of 1 plus *lm*, *drt*, and *ryans*, respectively. In this table, these variables are first sorted into ten groups according to ascending order over the years. We then calculate the number of observations in each two-way group. Panel A shows the distribution of *ESV_RYANS* and *ESV_LM*; Panel B shows the distribution of *ESV_RYANS* and *ESV_DRT*; Panel C shows the distribution of *ESV_LM* and *ESV_DRT*.

APPENDIX TABLE A3
Balance Tests

Variable	Matched	Mean (T)	Mean (C)	Diff.	p-value
SIZE	Unmatched	6.2426	6.5061	-0.2635***	0.0270
	Matched	6.2345	6.2769	-0.0424	0.7140
TOBINQ	Unmatched	2.0568	1.7630	0.2938***	0.0000
	Matched	2.0498	2.0301	0.0197	0.8010
CASH_FLOW	Unmatched	0.0653	0.0623	0.0030	0.7510
	Matched	0.0650	0.0725	-0.0074	0.4370
BLEV	Unmatched	0.1654	0.2074	-0.0419***	0.0000
	Matched	0.1658	0.1687	-0.0029	0.7930
CAPX	Unmatched	0.0483	0.0510	-0.0027	0.4370
	Matched	0.0482	0.0475	0.0007	0.8360
GSALE	Unmatched	0.1675	0.2233	-0.0557**	0.0250
	Matched	0.1665	0.1711	-0.0046	0.8190
NAN	Unmatched	1.5147	1.3685	0.1462**	0.0280
	Matched	1.5137	1.3903	0.1234*	0.0590
IO	Unmatched	0.5281	0.4633	0.0648***	0.0030
	Matched	0.5291	0.5122	0.0169	0.4250
RET	Unmatched	-0.1189	-0.1464	0.0275***	0.0090
	Matched	-0.1194	-0.1183	-0.0012	0.9130
SIGMA	Unmatched	0.0423	0.0471	-0.0049***	0.0020
	Matched	0.0424	0.0410	0.0014	0.3660
DTURN	Unmatched	-0.1882	-0.0944	-0.0937	0.1330
	Matched	-0.1853	-0.1407	-0.0446	0.4230
NCSKEW	Unmatched	0.0533	0.1580	-0.1047**	0.0160
	Matched	0.0538	0.0785	-0.0247	0.5390
OPAQUE	Unmatched	0.0753	0.0800	-0.0048	0.1930
	Matched	0.0753	0.0742	0.0010	0.7730

Notes: This table presents the univariate comparison of firms' characteristics between treated and control groups in the pre-match and post-match period and their corresponding p-value. Firms' characteristics include firm size (*SIZE)*, Tobin's Q (*TOBINQ)*, cash flow (*CASH_FLOW)*, book leverage (*BLEV)*, capital expenses *(CAPX)*, growth of sales (*GSALE)*, analyst coverage (*NAN)*, institutional ownership (*IO)*, mean of firm-specific weekly return (*RET)*, the standard deviation of firm-specific weekly return (*SIGMA)*, change of monthly turnover (*DTURN)*, negative conditional skewness of return (*NCSKEW)*, and opacity proxy based on modified Jones model (*OPAQUE)*. Mean (T) denotes the mean of the treatment group. Mean (C) denotes the mean of the control group.