View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by Institutional Knowledge at Singapore Management University

Singapore Management University

Institutional Knowledge at Singapore Management University

Research Collection Lee Kong Chian School Of Business

Lee Kong Chian School of Business

9-2020

Do short sellers use textual information? Evidence from annual reports

Hung Wan KOT

Frank Weikai LI
Singapore Management University, wkli@smu.edu.sg

Ming LIU

K.C. John WEI

Follow this and additional works at: https://ink.library.smu.edu.sg/lkcsb_research

Part of the Finance Commons, and the Finance and Financial Management Commons

Citation

KOT, Hung Wan; LI, Frank Weikai; LIU, Ming; and WEI, K.C. John. Do short sellers use textual information? Evidence from annual reports. (2020). 1-47. Research Collection Lee Kong Chian School Of Business.
Available at: https://ink.library.smu.edu.sg/lkcsb_research/6598

This Working Paper is brought to you for free and open access by the Lee Kong Chian School of Business at Institutional Knowledge at Singapore Management University. It has been accepted for inclusion in Research Collection Lee Kong Chian School Of Business by an authorized administrator of Institutional Knowledge at Singapore Management University. For more information, please email cherylds@smu.edu.sg.

Do Short Sellers Use Textual Information?
Evidence from Annual Reports[*]

Hung Wan Kot
Department of Finance and Business Economics, University of Macau
E-mail: hungwankot@um.edu.mo

Frank Weikai Li
Lee Kong Chian School of Business, Singapore Management University
E-mail: wkli@smu.edu.sg

Morris M. Liu
Department of Accounting and Information Management, University of Macau
E-mail: morrisliu@um.edu.mo

K.C. John Wei
School of Accounting and Finance, Hong Kong Polytechnic University
E-mail: johnwei@ust.hk

This version: August 11, 2020

Abstract

We examine short sellers' use of textual information in annual reports for shorting activities. We find that more uncertainty and negative words in annual reports are associated with greater abnormal shorting volume. Short selling motivated by textual information negatively predicts stock price reaction around the filing date of 10-K reports. We further provide some evidence that textual information used by short sellers are related to revisions of analysts' earnings forecasts, changes in firm fundamentals, and increasing crash risk subsequently. Our results suggest that textual information in annual reports forms an important part of short sellers' information advantage.

Keywords: Short selling; Annual reports; Textual analysis; Stock returns; Information environment

JEL Classification: G12; G14; G4; M41; M42

[*] We thank Monica Chen, Keith Lam, Xiaorong Ma, Xiumin Martin, Susan Ren, Lewis Tam, Vijay Singal, Feng Tang, and seminar participants at University of Macau for helpful comments and suggestions. This work was performed in part at the high performance computing cluster (HPCC) which is supported by Information and Communication Technology Office (ICTO) of the University of Macau. All errors are our own.

Corresponding author: K.C. John Wei, School of Accounting and Finance, Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong. E-mail: johnwei@ust.hk; Tel.: 852-2766-4953; Fax: 852-2330-9845.

Electronic copy available at: https://ssrn.com/abstract=3683086

Do Short Sellers Use Textual Information?
Evidence from Annual Reports

Abstract

We examine short sellers' use of textual information in annual reports for shorting activities. We find that more uncertainty and negative words in annual reports are associated with greater abnormal shorting volume. Short selling motivated by textual information negatively predicts stock price reaction around the filing date of 10-K reports. We further provide some evidence that textual information used by short sellers are related to revisions of analysts' earnings forecasts, changes in firm fundamentals, and increasing crash risk subsequently. Our results suggest that textual information in annual reports forms an important part of short sellers' information advantage.

Keywords: Short selling; Annual reports; Textual analysis; Stock returns; Information environment

JEL Classification: G12; G14; G4; M41; M42

Electronic copy available at: https://ssrn.com/abstract=3683086

1. Introduction

Information discovery and transmission in the financial market is at the core of finance and accounting research. Theories on public news release show that public news events can lead to differential interpretations by traders (Kandel and Pearson, 1995; Kim and Verrecchia, 1997; Hong and Stein, 1999). Moreover, greater and better public disclosures reduce informational asymmetry and cost of capital and enhance firms' liquidity (Diamond, 1985; Diamond and Verrecchia, 1991). Empirical studies also find that disclosure quality primarily affects information asymmetry by reducing the likelihood that investors trade on private information, and higher quality disclosures could improve aggregate shareholder welfare by reducing search costs (e.g., Brown and Hillegeist, 2007). Promoting high quality public information environments is always on top of regulators' agenda (Goldstein and Yang, 2019).

Short sellers, as a group of sophisticated investors, earn abnormal returns especially from heavily shorted stocks (Aitken et al., 1998; Desai et al., 2002; Arnold et al., 2005; Boehme et al., 2008; Diether et al., 2009; Rapach et al., 2016; Gargano et al., 2018; etc.). Researchers are interested in knowing the sources of short sellers' trading profitability and what kind of information motivates short selling activity. Some studies focus on private information while others examine public information like firm fundamentals and news announcements. For example, Engelberg et al. (2012) and von Beschwitz et al. (2017) show that short sellers are skilled processors of public information. Recently, Wang et al. (2019) document that short sellers appear to have shifted trading on short-term private information to trading on long-term public information that is gradually incorporated into stock prices. However, the process by which short sellers interpret public information is understudied in the literature.

Electronic copy available at: https://ssrn.com/abstract=3683086

Most existing studies on the relation between short selling and fundamentals focus on financial statement data but not the entire annual reports (Liberti and Petersen, 2019). In this study, we fill this gap by adopting the textual analysis approach to examine whether short sellers use qualitative information in annual reports. Particularly, we are interested in the type of annual reports that attract short sellers' attention. We adopt textual analysis to quantify the soft information as prior studies show that qualitative information in annual reports has predicative power for firm fundamentals and stock prices (Tetlock, 2007; Tetlock et al., 2008; Li, 2008; Jegadeesh and Wu, 2013; Loughran and MacDonald, 2011, 2014, 2016; Buehlmaier and Whited, 2018; Huang et al., 2019).

We focus on the annual reports for two reasons. First, annual reports are publicly available to all interested parties, especially in the Internet era (Drake et al., 2017). Drake et al. (2015) document that 10-Ks are the most commonly requested filings in SEC's EDGAR system, and count for 21% of all requests. As a result, financial disclosures such as 10-Ks make up a critical component of the information set available to investors. Second, annual reports provide an excellent setting because it contains the most comprehensive information about the firm over the past fiscal year. Psychology studies find that limited attention is a necessary consequence of the vast amount of information available in the environment, such that limited memory, attention, and processing capacities force investors to focus on a subset of available information (Hirshleifer, 2001; Hirshleifer and Teoh, 2003). On average the number of words in the annual reports is more than 50,000, which is long enough for short sellers to reveal their preferences on the type of information in the annual reports.[1] Short sellers might directly search for negative

[1] For example, apart from the details of financial statements, the section of Management's Discussion and Analysis (MD&A) provides the views of managements about the firm's future. In addition, Crane et al. (2018) provide evidence on hedge funds, one group of short sellers, download annual reports via EDGAR.

Electronic copy available at: https://ssrn.com/abstract=3683086

information which has not been incorporated into the stock price yet, or they may assess whether the stock is overvalued relative to fundamentals.[2]

Using textual data from annual reports during 2009 to 2015 and daily shorting volume data from NYSE/Amex/Nasdaq, we find that textual variables predict the abnormal shorting volume in the 4-day event window [0, 3]. Short sellers are more willing to short the stock if the firm's annual report is lengthy and contains more uncertainty and negative words. Using hedge funds' searching for 10-Ks from EDGAR as a proxy for their interest in a firm, we show that there is a strongly positive relation between hedge funds' requests of 10-Ks and abnormal shorting volume on the filing dates of 10-Ks, suggesting that information contained in 10-Ks is useful to short sellers.

We next investigate whether abnormal shorting volume and textual information can predict future stock returns. The literature documents that short sellers are skilled information processors.[3] If short sellers are able to extract negative information from ambiguous writings, we should observe a strong negative relation between abnormal shorting volume and abnormal stock returns. However, short sellers may have behavioral bias just as other investors.[4] For example, Ditto and Lopez (1992) find that the information consistent with a preferred conclusion will be examined less critically than information inconsistent with a preferred conclusion, and

[2] Hunton and McEwen (1997) use experiments and find that more accurate analysts employ a directive information search strategy, whereas less accurate analysts employ a sequential search strategy. In addition, post-experiment survey results show the linkage between specific accounting information used by analysts and the accuracy of their forecasts.

[3] See, for example, Desai et al. (2006), Karpoff and Lou, (2010), Drake et al. (2011), Fang et al. (2016), Engelberg et al. (2012), and von Beschwitz et al. (2017).

[4] See a survey paper by Hirshleifer (2001). For financial analysts, another group of important market participants, the literature also well documents the bias they have made. For example, Easterwood and Nutt (1999) find that analysts overreact to good news but underreact to bad news. Experimental studies find that given equivalent information disclosure about a firm, different presentation ways affect the valuations and trades of investors even experienced financial analysts (see Hirshleifer and Teoh, 2003).

Electronic copy available at: https://ssrn.com/abstract=3683086

consequently, less information is required to reach the former case than the latter case.[5] Recently, Huang et al. (2019) examine institutional trading surrounding corporate news, and find that institutions mainly trade on the tone of news directly after the earliest news release. If short sellers simply base their trading on the writing style of annual reports (like pessimistic tone with more negative words) and without further analyses, abnormal shorting volume may not lead to trading profits.

The empirical results on stock return predictability are summarized as follows. First, textual variables predict abnormal returns in the 4-day event window [0, 3]. Secondly, short selling is informative about stock returns when short sellers use textual information on modal weak words. Third, textual variables also predict 1-week to 12-week ahead abnormal returns. Using fitted abnormal shorting volume (i.e., shorting volume driven by textual information), we find a significantly negative relation between abnormal shorting volume and abnormal stock returns from [0, 3] days to 2-week ahead. The results suggest that short sellers are skilled processors of both qualitative and quantitative information, and they are able to discover negative information from complex annual reports.

We further investigate the source of the return predictability of short selling and textual information by examining the revisions of analysts' earnings forecasts, and changes in firm fundamentals subsequently. Our results show that the revisions of analysts' earnings forecasts around filing months are related to the ratio of negative and modal weak words, but unrelated to abnormal shorting volume. It suggests that analysts and short sellers do not influence each other

[5] Using experiments, Hales (2007) also finds that consistent with the theories of motivated reasoning, directional preferences affect how information is processed. Hales finds that investors are motivated to agree unthinkingly with information that suggests that they might make money on their investments, but disagree with information that suggests they might lose money.

Electronic copy available at: https://ssrn.com/abstract=3683086

around 10-Ks filing dates. The ratio of negative words also predicts changes in returns on assets and asset turnover in the next fiscal year.

Finally, motivated by Hutton et al. (2009), Callen and Fang (2015), Kim, Wang, and Zhang (2019), and Deng et al. (2020), we investigate whether abnormal shorting volume and textual information can predict future crash risk. We find some evidence that the ratio of uncertainty words is negatively related to crash risk proxies. However, the interaction of abnormal shorting volume with uncertainty words is positively related to crash risk. This suggests that short sellers are informative about a firm's future crash risk, which is consistent with findings of Karpoff and Lou (2010).

Our study contributes to the literature in the following ways. First, we contribute to the growing literature on how short sellers process information, by focusing on the textual information in annual reports. As far as we know, this is the first study to examine the relation between annual reports and short selling activities. The way for short sellers to make profit is through future stock price depreciation, which could come from current stock prices not reflecting the true future prospects of the firms.[6] Studies find that short sellers utilize fundamental analysis when targeting overvalued companies[7] and they are superior in analyzing public information and taking advantage from noise traders.[8] Another branch of studies focuses on how short sellers use private information to form their trading strategies (Christophe et al., 2004; Khan and Lu, 2013; Shi et al., 2017; Berkman et al., 2017; Berkman and Eugster, 2017, and Purnanandam and Seyhun, 2018). Our study complements Engelberg et al. (2012) and von

[6] This overvaluation could also be purely from negative information that has not been incorporated into share prices due to short-sale constraints and a large difference in investors' opinions (Miller, 1977; Chen et al., 2002; Nagel, 2005; Boehme et al., 2006; Berkman et al., 2009).
[7] For instance, Dechow (2001), Curtis and Fargher (2014), Deshmukh et al. (2015), and Drake et al. (2015).
[8] See, Desai et al. (2006), Karpoff and Lou (2010), Drake et al. (2011), Fang et al. (2016), Engelberg et al. (2012), and von Beschwitz et al. (2017).

Electronic copy available at: https://ssrn.com/abstract=3683086

Beschwitz et al. (2017) by further studying short sellers' use of textual information in annual reports.

Our study also contributes to the emerging literature on investor's information acquisition, especially information acquisition from EDGAR. Drake et al. (2015) analyze the determinants of users' access of SEC filings through EDGAR, and find that EDGAR search activity is positively related to firm events and information environments. Crane et al. (2018) examine whether hedge funds profit from publicly available SEC filings at EDGAR, and conclude that hedge funds use public information to complement their private signals. Li and Sun (2019) find that the abnormal number of IPs searching for firms' financial statements strongly predicts future stock returns and firm fundamentals. Our paper complements Drake et al. (2015, 2017), Crane et al. (2018), and Li and Sun (2019) and find that there is a strongly positive relation between the number of EDGAR requests and abnormal shorting volume, and these activities are informative about future stock returns.

Finally, we contribute to the literature on textual analysis. Existing finance and accounting literature on textual analysis mainly focus on annual reports. Those studies either focus on the relation between different attributes of annual reports and firm performance (Li, 2008; Dyer et al., 2017, and Buehlmaier and Whited, 2018) or develop new methods for textual analysis.[9] Textual analysis also helps to explain the underpricing of IPOs (Hanley and Hoberg, 2010; Jegadeesh and Wu, 2013) and are useful for predicting future stock market returns (Jiang et al., 2019). We add to the literature by studying how the textual information can be used by short sellers in their trading.

[9] See, Brown and Tucker (2011), Jegadeesh and Wu (2013), Loughran and MacDonald (2011, 2014, 2016), and Ke et al. (2019).

Electronic copy available at: https://ssrn.com/abstract=3683086

The rest of paper is organized as follows. Section 2 describes the data and variables construction. Section 3 provides empirical evidence from daily shorting volume and textual information. Section 4 discusses empirical results on abnormal stock returns and textual information. Section 5 investigates the relation with fundamentals. Section 6 examines whether related to crash risk. Section 7 concludes the paper.

2. Data and variables construction

2.1 Data

Our main analysis relies on the daily shorting volume from September 2009 to December 2015, which is available to download from the FINRA website.[10] We compute the daily shorting volume from the Regulation SHO monthly short sale transaction file from NYSE and Nasdaq which report independently. We then aggregate them at the stock level. We choose the monthly transaction file to construct daily shorting volume rather than directly use the daily short sale volume data which are also available at the FINRA website. The reason is that "some offsetting buying activity related to reported short selling would not be reflected in the Daily File" as reported in the FINRA website. In other words, using the daily file will underestimate shorting volume.[11] We include only common stocks (share code 10 and 11 in CRSP) listed on NYSE/Amex and Nasdaq.

[10] See the website: http://www.finra.org/. The sample of our study ends in December 2015, because our EDGAR download data end in December 2015.

[11] For a particular day (March 18, 2019) that we check, 83% of the securities have the same shorting volume from the monthly short sale transaction file and the daily short sale volume file; while the remaining 17% have higher shorting volume from the monthly file than the daily file. Overall, shorting volume from the monthly file is 0.44% higher than it from the daily file.

Electronic copy available at: https://ssrn.com/abstract=3683086

We obtained the initial 10-Ks textual data from the website of The Notre Dame Software Repository for Accounting and Finance.[12] We use the Loughran-McDonald 10X File summaries file, a file containing sentiment counts, file size, and other measures for all 10-X filings for all years.

The hedge funds' downloads of 10Ks from EDGAR are merged and computed from three databases. First, we obtain a list of hedge funds that are similar to Jiang (2019).[13] We next manually search each hedge fund's IP address from a commercial IP address database (https://db-ip.com/) and finally match each hedge fund's download of 10Ks from the Securities and Exchanges Commission's (SEC) EDGAR log file database at the daily level.[14] We further merge this data with other databases by the CIK code.

Monthly stock returns, prices, and the number of shares outstanding are obtained from the CRSP. Annual accounting data are obtained from Compustat. Analysts' earnings forecasts data are obtained from the I/B/E/S database. Institutional ownership data are obtained from the Thomson Institutional Ownership database (13-F). Daily shorting volume data are merged with the CRSP with the stock symbol trading on the exchanges. Textual data and EDGAR download data are merged with the CRSP with the CIK code by the filing date. Institutional ownership data, analysts' earnings forecasts data, and accounting data are merged either by PERMNO, Cusip, or stock trading symbol. Table 1 provides details of the sample construction.

[Insert Table 1 here]

2.2 Short sale variables

---

[12] See the website: https://sraf.nd.edu/. Up to our download date, the textual data are from 1986 to 2018, but only from 1996 to 2017 have the complete data for common stocks. We acknowledge Bill McDonald for maintaining this website and provide the data freely.

[13] We acknowledge Wenxi Jiang for sharing his hedge fund list. For the detail of the hedge fund list construction, please refer to Section III.A of Jiang (2019).

[14] For the detailed description of EDGAR log file database, please refer to Section 3.1 of Li and Sun (2019).

Electronic copy available at: https://ssrn.com/abstract=3683086

We use cumulative abnormal shorting volume for the event-day window [0, 3] ($CAShort$) as our main measure of shorting volume, where day 0 is the filing day of the 10-K annual report available on SEC's EDGAR. We first normalize shorting volume for firm i on day d ($Short_{i,d}$) as the ratio of daily shorting volume ($ShortVolume_{i,d}$) divided by total share outstanding ($SharesOutstanding_i$). We use the average of normalized daily shorting volume ($Short_{i,d}$) in the past 100 days skipping the first 20 days (i.e., from day -120 to day -21) to proxy for normal daily shorting volume ($AvgShort_{i,d}$). Abnormal shorting volume ($AbnShort_{i,d}$) for firm i on day d is therefore defined as follows:

$$AbnShort_{i,d} = Short_{i,d} - AvgShort_{i,d} = \frac{ShortingVolume_{i,d}}{SharesOutstanding_i} - \frac{\sum_{d=-21}^{-120} Short_{i,d}}{100}, \quad (1)$$

Finally, the four-day cumulative shorting volume for firm i ($CumShort_i$) is defined as:

$$CAShort_i = CAShort[0,3]_i = \sum_{d=0}^{3} AbnShort_{i,d}, \quad (2)$$

Hedge funds are the major type of short sellers. We use hedge funds' 10Ks download from EDGAR as a proxy for the short sellers' download of annual reports and define $Download$ as the number of hedge funds' downloads of 10Ks from EDGAR at the daily level.

2.3 Textual variables

We use the following six variables from Loughran-McDonald's 10X File summaries file: (i) n_words, which is the count of all words, where a word is any token appearing in the Master Dictionary; (ii) $n\_uncertainty$, which is the number of words related to uncertainty; (iii) $n\_modalweak$, which is the number of words related to modal weak; (iv) $n\_modalstrong$, which is the number of words related to modal strong; (v) $n\_negative$, which is the number of words related to negative; and (vi) $n\_positive$, which is the number of words related to positive.

Electronic copy available at: https://ssrn.com/abstract=3683086

We construct three ratios related to information from textual. Specifically, $R\_uncertainty$ is defined as the ratio of $n\_uncertainty$ divided by n_words; $R\_modalweak$ is defined as the ratio of $n\_modalweak$ divided by the sum of $n\_modalwek$ and $n\_modalstrong$; and $R\_negative$ is defined as the ratio of $n\_negative$ divided by the sum of $n\_negative$ and $n\_positive$. For other variables used in the regression, we take the natural logarithm. Finally, we calculate n_filing as the number of 10Ks filings per day.

2.4 Abnormal stock returns

We define the buy-and-hold abnormal return for stock *i* from day j to day k ($BHAR[j,k]_i$) as follows.

$$BHAR[j,k]_{i,} = \prod_{d=j}^{k} Ret_{i,d} - \prod_{d=j}^{k} Mkt_d, \tag{3}$$

where $Ret_{i,d}$ is the daily stock return for firm *i* on day d and Mkt_d is the daily value-weighted CRSP market index return on day *d*. We construct $BHAR$ during the 4-day event window [0, 3] and denote it as $BHAR4d$.[15] For the weekly frequency, similar to Eq. (3), we construct the cumulative abnormal weekly returns after the filing date in the 1-week, 2-week, 4-week, 12-week, 24-week, or 52-week period, which is denoted as $BHAR1w$, $BHAR2w$, $BHAR4w$, $BHAR12w$, $BHAR24w$, or $BHAR52w$.

2.5 Analysts' earnings forecasts

[15] We follow Loughran and McDonald (2011) to use the 4-day cumulative return of [0, 3]. Furthermore, Choi et al. (2017) report that hedge funds short sales covering five trading days are highly profitable, but not for more than five days.

Electronic copy available at: https://ssrn.com/abstract=3683086

We construct the revision of analysts' earnings forecasts for fiscal year 1 ($\Delta FEPS1_{i,t}$) as follows:

$$\Delta FEPS1_{i,t} = \frac{FEPS1_{i,t+1} - FEPS1_{i,t-1}}{StockPrice_{i,t-1}}, \quad (4)$$

where $FEPS1_{i,t-1}$ denotes analysts' earnings forecasts per share for fiscal year 1 at month *t*-1. We similarly calculate the revision of analysts' earnings forecasts for fiscal year 2 ($\Delta FEPS2_{i,t}$).

2.6 Fundamental variables

We adopt the following three accounting-related variables from annual financial statements to measure the changes in a firm's fundamental: (i) the change in return on assets from fiscal year *y* to *y*+1 ($\Delta ROA_{i,y}$), (ii) the change in asset turnover from fiscal year *y* to *y*+1 ($\Delta AssetTurn_{i,y}$); and (iii) the change in operating profit margin before depreciation from fiscal year *y* to *y*+1 ($\Delta OPM_{i,y}$).

2.7 Measures of crash risk

We follow Chen, Hong, and Stein (2002) and Bae, Lim, and Wei similarly to construct three measures of the crash risk in fiscal year *y*+1. The first crash risk measure is the negative of the third central moment of firm-specific weekly returns scaled by the variance of firm-specific weekly returns raised to the power of 3/2 ($NSkew_{i,t}$) using data from the past 52 weeks:

$$NSkew_{i,y} = -\frac{(n(n-1))^{3/2}\sum_{w=1}^{n}(Ret_{i,w,y} - \overline{Ret}_{i,y})^3}{(n-1)(n-2)(\sum_{w=1}^{n}(Ret_{i,w,y} - \overline{Ret}_{i,y})^2)^{3/2}}, \quad (5)$$

where $w_{i,w,y}$ is the firm-specific weekly stock return for week *w* in year *y*, and $\overline{w}_{i,y}$ is the mean firm-specific weekly stock return for year *y* and *n* is the number of weeks in year *y*. We put a

Electronic copy available at: https://ssrn.com/abstract=3683086

negative sign in front of the skewness so that a higher $NSkew$ value corresponds to a more negative-skewed stock return distribution, namely, higher crash risk.

The second crash risk measure is the "down-to-up" volatility ratio ($DUVolR_{i,y}$), which is calculated as follows:

$$DUVolR_{i,y} = Ln\left\{\frac{n_{up}(\sum_{w\in Down}(Ret_{i,w,y} - \overline{Ret}_{i,y})^2)}{n_{down}(\sum_{w\in Up}(Ret_{i,w,y} - \overline{Ret}_{i,y})^2)}\right\}, \quad (6)$$

where n_{down} (n_{up}) is the number of up (down) weeks, $Ret_{i,w,y}$ the return for firm *i* in *w* during year y and $\overline{Ret}_{i,y}$ is the mean of the weekly returns. A down (up) week is defined as a week when the firm-specific weekly return is above (below) the mean weekly return over fiscal year *y*. Since $DUVolR_{i,y}$ does not involve the third moment, it is therefore less likely to be affected by a small number of extremely weekly returns.

The last measure of crash risk is the difference in the frequencies between extreme negative returns and extreme positive returns ($n\_Crash_{i,y}$), which is defined as follows:

$$n\_Crash_{i,y} = n_{negative} - n_{positive}. \quad (7)$$

This measure is based on the number of firm-specific weekly returns exceeding 3.09 standard deviations below ($n_{negative}$) and above ($n_{positive}$) the mean firm-specific weekly return over the fiscal year. The value 3.09 is chosen to generate the frequency of 0.1% in the normal distribution. A higher value of n_Crash indicates a higher frequency of crashes.

2.8 Control variables

We include the following common control variables (i) firm size (SZ), which is measured as market capitalization in million dollars; (ii) book-to-market equity ratio (B/M); (iii) Amihud's

Electronic copy available at: https://ssrn.com/abstract=3683086

illiquidity measure ($Illiq$); (iv) institutional ownership scaled by the number of shares outstanding ($IOwner$); (v) cumulated past 1-year stock returns ($Pr1y$) to proxy for momentum strategy; and (vi) idiosyncratic volatility ($IVol$), which is the mean squared error of residuals of daily stock returns from the last three months estimated from the Fama-French three-factor model augmented by the Carhart momentum factor. The detailed definitions of all these variables are described in Appendix A.

2.9 Summary statistics

Table 2 reports the summary statistics of all variables, which are winsorized at the 1% level. Since our main study focuses on the daily shorting volume which is only available from September 2009, we also report the summary statistics from September 2009 to December 2015. Table 2 shows the mean $CAShort$ is 0.11%, which means that the daily shorting volume during the 4-day window [0, 3] is higher than average normal trading days. The mean $BHAR[0, 3]$ is -0.04% during the event window [0, 3], which suggests that the stocks of firms around their 10Ks filing dates [0, 3] on average underperform the market.

[Insert Table 2 here]

For the number of words in the annual reports, the average of total words (n_words) is 52,763. The average number of uncertainty words ($n\_uncertainty$) is 732. The average number of negative words is more than positive words in the annual reports, with the averages of $n\_negative$ and $n\_positive$ being 970 and 378, respectively. This is consistent with the number of negative and positive words in the Loughran and McDonald Sentiment Word Lists (LM list), which are 2,355 and 354. The annual reports also prefer to use modal weak tone, as the number of modal weak words is much more than modal strong words (324 vs. 158), compared to the

Electronic copy available at: https://ssrn.com/abstract=3683086

number of modal weak versus modal strong words in the LM list of 27 versus 19. Those textual variables are well described in Loughran and McDonald (2011, 2014). For the ratio of textual variables, $R\_negative$ is slightly higher than $R\_modalweak$ (0.711 vs 0.677), while the average percentage of uncertainty words in annual reports is 1.43%. The average number of filing firms per day is 89, with variation from a minimum of 1 to a maximum of 315. The average $Download$ of 10Ks per firm by hedge funds on the filing date is 20.9.

Other variables describe the characteristics of the sample in the study. For example, the average firm size is $4,335 million with the average institutional ownership of 54.9%. The average of previous one-year cumulative returns ($Pr1y$) is 25.9%. This is also consistent with the fact that overall stock prices in the market increase significantly. For example, the S&P 500 index increases from 998 in September 1, 2009 to 2,043 on December 31, 2015. This also indicates that only very skillful short sellers can make profit in a bull market.

Table 3 reports the correlation coefficients of main variables. Cumulative abnormal shorting volume ($CAShort$) during the event window [0, 3] is positively correlated with the length of annul reports ($Ln(n\_words)$) and the number of 10K downloads by hedge funds ($Ln(Download)$). Among textual ratios, $R\_modalweak$ is highly positively correlated with $R\_uncertainty$ (ρ = 0.53). Among firm characteristics, $Ln(SZ)$ is highly positively correlated with Institutional ownership ($IOwner$) but negatively with illiquidity ($Illiq$) and idiosyncratic volatility ($IVol$).

[Insert Table 3 here]

3. Daily shorting and textual information

Electronic copy available at: https://ssrn.com/abstract=3683086

To investigate short sellers' reaction to the filing of 10-K reports, we run the following pooled OLS regression on cumulative abnormal shorting volume in during the event window [0, 3] ($CAShort$):

$$CAShort_{i,t} = \alpha + \sum_{j=1}^{3} \beta_j Textual_{i,j,t} + \beta_4 Download_{i,t} + \gamma_1 Ln(n\_words)_{i,t} + \gamma_2 Ln(n\_filing)_t + \sum_{j=1}^{3} \theta_j Textual_{i,j,t} \times Download_{i,t} + control_{i,t} + d_{m,t} + \varepsilon_{i,t}. \quad (8)$$

where $Textual$ includes $R\_uncertainty$, $R\_modalweak$, $R\_negative$. To absorb the time-invariant stock-specific effects and aggregate time trends, we include stock fixed-effects (f_i) and year-month dummy variables $(d_{m,t})$ in the regression model. We consider six different model specifications and their results are reported in Table 4. Standard errors are double clustered at the firm and year-month levels (Petersen, 2009; Prado, Saffi, and Sturgess, 2016). We apply the same methods in the regression models throughout the paper.

Table 4 reports the results. Models 1-4 reports the results without $Textual$ interacted with $Download$, and Models 5-6 with. Among all the textual ratios analyzed, the coefficients on $R\_uncertainty$ are all significantly positive, ranging from 0.040 to 0.047 and with t-statistics ranging from 1.77 to 2.22. The coefficient estimate is also economically significant. Take the example of Model 1, a one-standard-deviation increase leads to a 20.5% (= 0.047×0.0048/0.0011) increase in $CAShort$. The results suggest that sellers are more willing to take a short position if they discover the hidden negative information from uncertainty words in the annual reports. The coefficients on $R\_negative$ are all negative, ranging from -0.023 (t-stat = -1.41) in Model 1 to -0.042 (t-stat = 2.41) in Model 3. When firm size is not controlled for, half of them (Models 3, 4, and 6) are significant at the 5% level with the coefficient ranging from -0.040 to -0.042. Take the example of Model 3, a one-standard-deviation increase in $R\_negative$ leads an 18.3% (=

Electronic copy available at: https://ssrn.com/abstract=3683086

0.042×0.0048/0.0011) decrease in $CAShort$. It suggests that short sellers view more negative words in the annual reports as containing positive information and are less willing to take short positions. However, this significantly negative relation between negative words and shorting volume becomes insignificant in Models 1, 2, and 5, firm size is controlled for. Interestingly, the coefficients of the interaction terms are all insignificant.

[Insert Table 4 here]

For the two variables used to measure the length of annual reports, $Ln(n\_words)$ is positively related to shorting volume in most model specifications. For example, the coefficient on $Ln(n\_words)$ is 0.069 with a *t*-statistic of 3.18 in Model 1, suggesting that short sellers short aggressively on firms with lengthy annual reports. A one-standard-deviation increase in $Ln(n\_words)$ causes a 30.1% (= 0.069×0.0048/0.0011) increase in abnormal shorting volume. The result suggests that short sellers may have better skill to analyze a large amount of information in the annual reports as the average number of words in annual reports is 52,763.[16] The results are consistent with von Beschwitz et al. (2017) who find that short sellers trade more on the days with qualitative news.

Model 1 also shows that short sellers short less on a particular firm if there are more firms filing 10-K reports on the filing date. The coefficient on $Ln(n\_filing)$ is -0.063 (*t*-stat = 2.50). However, this effect is absorbed by the number of 10-K requests by hedge funds on the filing date as shown in Models 2-6. Once we include $Download$, the coefficients on $Ln(n\_filing)$ is no longer significant. Models 2-6 further show that $Download$ has the strongest effect on shorting volume. For example, the coefficient on $Download$ in Model 2 is 0.115 (t-stat = 5.06), indicating that a one-standard-deviation increase in $Download$ causes a 50.1% (=

[16] Although not reported, we also find that the number of unique words in annual reports and the gross and net file size of annual reports are also positively related to shorting volume.

Electronic copy available at: https://ssrn.com/abstract=3683086

0.115 $\times$0.0048/0.0011) increase in abnormal shorting volume. The result provides direct evidence on the positive relation between short sellers' use of annual reports and their abnormal shorting activity. Specifically, short sellers take more short positions because more short sellers download and read annual reports and find useful information.

In Model 5-6 of Table 4, textual variables are interacted with $Download$ to investigate which type of textual information interacted with $Download$ is favored by short sellers. However, the results show that the coefficients on interaction terms are insignificant.

Finally, we find that several firm-level control variables are related to abnormal shorting volume around the filing days. For example, $CAShort$ is positively related to past one-year stock returns ($Pr1y$) and idiosyncratic volatility ($IVol$), but negatively related to stock illiquidity ($Illiq$). These suggest that short sellers prefer to short firms with better liquidity (less transaction costs), good past one-year stock performance (contrarian strategies), and a higher difference of investors' opinion. In general, those findings are consistent with the studies related to short selling activity (e.g., Negal, 2005; Arnold et al., 2005; Kot, 2007; Chen et al., 2013; Beneish et al., 2015, Cheung et al., 2019, etc.).

4. Abnormal stock returns and textual information

4.1 Abnormal stock returns during the event window [0, 3]

In this sub-section, we investigate whether textual variables are able to predict abnormal stock returns and the role of short sellers play If abnormal shorting volume contains negative information conditional on textual variables, we expect the coefficient on the interaction terms between them to be significant. We use the following pooled OLS regression to test our hypothesis:

Electronic copy available at: https://ssrn.com/abstract=3683086

$$BHAR4d_{i,t} = \alpha + \beta_0 CAShort_{i,t} + \sum_{j=1}^{3} \beta_j Textual_{i,j,t} + \beta_4 Download_{i,t}$$
$$+\beta_5 Ln(n\_words)_{i,t} + \beta_6 Ln(n\_filing)_t + \sum_{j=1}^{3} \theta_j Texxtual_{i,j,t} \times CAShort_{i,t} \quad (9)$$
$$+\theta_4 CAShort_{i,t} \times Download_{i,j,t} + control_{i,t} + d_{m,t} + \varepsilon_{i,t}.$$

where all variables are defined previously. Table 5 reports the results. We find that all three textual variables are significantly associated with future abnormal returns. First, a higher $R\_uncertainty$ implies a lower abnormal return, as indicated by the negative coefficients on $R\_uncertainty$ in all models. For example, the coefficient on $R\_uncertainty$ is -0.046 (*t*-stat = 2.14) in Model 1, suggesting that investors view the uncertainty words in 10-Ks as negative information, consistent with short sellers taking more short positions as shown in Table 4.

Second, a higher $R\_negative$ implies a higher abnormal return. For example, the coefficient on $R\_negative$ is 0.037 (*t*-stat = 2.01) in Model 2. This result is a bit counter-intuitive as negative words should represent negative information. We further explore the relation of $R\_negative$ with firm characteristics and find that it highly correlated with book-to-market (B/M) (ρ = 0.22), but not with the previous 1-year return ($Pr1y$) and changes in $R\_negative$ from the previous year. When we sort $R\_negative$ into quintiles in each year, the average value of B/M increases monotonically. Combining the negative relation between $R\_negative$ and $CAShort$ from Table 4, it suggests that short sellers are less likely to target value stocks with more negative words, probably because the price of value stocks already overreact to the negative prospects.

Third, for the two variables used to measure the length of the annual reports, we find that $Ln(n\_words)$ is significantly and negatively related to $BHAR4d$ in all models except Model 1. For example, the coefficient on $Ln(n\_words)$ in Model 2 is –0.078 (*t*-stat = -2.71), indicating that investors view longer annual reports more negatively, potentially because for such firms managers provide a lot of useless information to investors.

Electronic copy available at: https://ssrn.com/abstract=3683086

Fourth, the number of filings per day is also negatively related to abnormal stock returns. The coefficients on $Ln(n\_filing)$ is significantly negative in all six models in Table 5, even after controlling for $Download$ in Models 5-6. For example, the coefficient on $Ln(n\_filing)$ is -0.067 (*t*-stat = -2.82) in Model 1, which is consistent with the finding by Hirshleifer, Lim, and Teoh (2009) that investors are distracted if there are many 10-K filings in a day. It is also consistent with the proposition that managers of firms with bad news tend to time the filing day with many 10-K filings to divert investors' attention. Finally, $Download$ has a negative relation with $BHAR4d$, which is consistent with the result in Table 4 that $Download$ is positively associated with abnormal shorting volume.

[Table 5 here]

Next, the results from the interactions of $CAShort$ with textual variables indicate that short selling is more informative when the annual report contains more modal weak words. For example, the coefficients on $CAShort \times R\_modalweak$ are both significantly negative (for example, coeff = -0.035 with *t*-stat = -2.21 in Model 4) mean that short sellers taking more aggressive short positions when the annual reports contain more modal weak words lead to more negative abnormal returns. The informativeness of short selling is further confirmed by the highly significant coefficients on $Download$ and $CAShort \times Download$. In contrast, the effect of abnormal shorting volume on abnormal stock returns is not related to more negative or uncertainty words in the 10-K reports.

Finally, the coefficient on $CAShort$ is significantly positive in all models, suggesting that short selling activity itself is unable to predict negative future abnormal returns on average. This finding is consistent with prior studies including Choi et al. (2017), Crane et al. (2018), and Gargano et al. (2018). For example, Gargano et al. (2018) find that short sellers experience

Electronic copy available at: https://ssrn.com/abstract=3683086

losses on average, consistent with the positive equity risk premium in their sample. Choi et al. (2017) find that hedge funds are not profitable if they hold positions more than five trading days, and institutional investors' short sales are not profitable.

4.2 Abnormal stock returns from 1-week to 52-week ahead

Table 5 shows a significant contemporaneous relation between abnormal shorting volume, textual variables, and abnormal stock returns around 10-K filing dates. A natural question is whether abnormal shorting volume can predict stock returns in the long-run. We replace $BHAR4d$ with $BHAR1w$ to $BHAR52w$ and re-run Eq. (9). Table 6 reports the result. We find that the relation between future stock returns and abnormal shorting volume remains highly significant positive in all models with abnormal returns from 1-week to 24-week ahead, but not for 52-week ahead. For example, the coefficient on $CAShort$ is 0.126 (*t*-stat = 5.81) in Model 1 with $BHAR1w$ and is 0.039 (*t*-stat = 2.79) in Model 9 with $BHAR24w$. The evidence further supports the view that there is a strongly positive relation between abnormal shorting volume around 10-K filings and future stock returns up to 24 weeks. However, short sellers underperform if they take short positions.

In contrast, we find that the coefficient on $R\_uncertainty$ is significantly negative for $BHAR1w$, $BHAR12w$, and $BHAR24w$, $R\_modalweak$ has a significantly positive coefficient for $BHAR1w$, $BHAR2w$, and $BHAR52w$, and $R\_negative$ has a significantly negative coefficient for all holding horizons except for 12 and 24 weeks. However, the coefficients on all interaction terms are insignificant except for $CAShort \times R\_uncertainty$ (coeff = -0.028; t-stat = 2.05), suggesting that short sellers are quite efficient in incorporating textual information into stock prices during the event window [0, 3], and therefore there is no further return drift after the

Electronic copy available at: https://ssrn.com/abstract=3683086

textual information becomes publicly available.[17] Meanwhile, the relation between $Download$ and future stock returns is negatively significant for $BHAR2w$, $BHAR24w$, and $BHAR52w$, with the coefficients ranging from -0.021 (t-stat = -1.73) in Model 3 to -0.058 (t-stat = -3.80) in Model 12. Moreover, we find that the negative effect of $Download$ on future stock returns is amplified by abnormal shorting volume for the returns measured by $BHAR1w$, $BHAR4w$, and $BHAR24w$, with the coefficients on the interaction term ranging from -0.021 (t-stat = -1.78) in Model 9 to -0.029 (t-stat = -2.30) in Model 2. These results suggest that hedge funds' downloads in general can negatively predict future returns up to 52 weeks and there is some evidence that abnormal shorting volume can amplify this negative download effect.

[Table 6 here]

4.3 Robustness checks with fitted shorting volume

In this section, we decompose shorting volume into two parts: the fitted component and the residual component. The fitted shorting volume is driven or motivated by textual information contained in annual reports. As a result, if our story is true, it should have a stronger predictability for future returns than the residual shorting volume, which is not motivated by textual information in annual reports. We first run the following regression to obtain the fitted $CAShort_{i,t}$ ($\widehat{CAShort}_{i,t}$):

$$CAShort_{i,t} = \alpha + \sum_{j=1}^{3} \beta_j Textual_{i,j,t} + \varepsilon_{i,t}. \tag{10}$$

We then run the following regression,

$$\begin{aligned} BHAR_{i,t} = \alpha + \beta_1 \widehat{CAShort}_{i,t} + \beta_2 \operatorname{Ln}(n\_filing)_t + \beta_3 Download_{i,t} \\ + \beta_4 Ln(n\_words)_{i,t} + control_{i,t} + d_{m,t} + \varepsilon_{i,t}. \end{aligned} \tag{11}$$

[17] In unreported regression analyses, when replace $CAShort[0,3]$ by $CAShort[-3,-1]$, we find no relation between $BHAR4d$ and $CAShort[-3,-1]$.

Electronic copy available at: https://ssrn.com/abstract=3683086

If $\widehat{CAShort}$ contains useful textual information related to future stock returns, we would expect that the coefficient on $\widehat{CAShort}$ to be significantly negative.

Table 7 reports the results. The coefficients on $\widehat{CAShort}$ are indeed significantly negative when the dependent variable is $BHAR4d$, $BHAR1w$, and $BHAR2w$. The corresponding coefficients are -0.042 (*t*-stat = 2.92), -0.035 (*t*-stat = 2.84), and -0.028 (*t*-stat = 2.13), respectively. More interestingly, the coefficient of $\widehat{CAShort}$ is also significant when the dependent variable is $BHAR52w$. These results show that short sellers are indeed informative in predicting future poor stock performance, when their aggressive short selling volume is driven by textual information.

[Table 7 here]

5. Textual information and firm fundamentals

In Sections 3 and 4, we find significant relationship between abnormal shorting, textual variables, and abnormal stock returns. However, the type of information contained in the textual variables is still not clear. It is important to know whether textual variables capture information from financial statements, as prior studies show that such financial information are used by short sellers in identifying overvaluation (Dechow, 2001; Curtis and Fargher, 2014; Deshmukh et al., 2015; and Drake et al., 2015). In this section, we conduct tests relating abnormal shorting with revisions of analysts' earnings forecasts and change in firm fundamentals.

5.1 Analysts' earnings forecast revision

To investigate whether textual information and short selling predicts analyst forecast revision, we perform the following regression.

Electronic copy available at: https://ssrn.com/abstract=3683086

$$\Delta FEPS1_{i,t}(\Delta FEPS2_{i,t}) = \alpha + \beta_0 CAShort_{i,t} + \sum_{j=1}^{3} \beta_j Textual_{i,j,t} + \beta_4 Download_{i,t}$$
$$+\beta_5 Ln(n\_words)_{i,t} + \beta_6 Ln(n\_filing)_t + \sum_{j=1}^{3} \theta_j Texxtual_{i,j,t} \times CAShort_{i,t} \quad (12)$$
$$+\theta_4 CAShort_{i,t} \times Download_{i,j,t} + control_{i,t} + d_{m,t} + \varepsilon_{i,t}.$$

We focus on forecast revisions from month t-1 to month t+1 because the filing dates are randomly distributed within a month, and most analysts revisions are issued over the 10 days following the filing date (Celment et al., 2011).

Table 8 reports the results. Among all the textual variables, $R\_negative$ is positively and $R\_modalweak$ is negatively related to analysts' earnings forecast revisions from month t-1 to month t+1. The corresponding coefficients are 0.061 (t-stat = 2.82) and -0.026 (t-stat = 1.88) in Model 2 for $\Delta FEPS1$s and 0.044 (t-stat = 2.31) and -0.034 (t-stat = 2.31) in Model 4 for $\Delta FEPS2$. The results suggest that more negative (modal weak) words in annual reports are associated with upward (downward) revisions of analysts' forecasts for both fiscal year 1 and year 2 earnings. The results are consistent with Drake et al. (2015), who find that short selling strengthens the relation between current returns and future earnings, especially in the setting where short sellers are likely to possess an information advantage.

The significant coefficient on $CAShort \times Download$ shows that analysts' forecast revisions are related to short selling activity, i.e., higher shorting volume with higher 10-Ks download by short sellers is negatively related to downward revisions of analysts' forecasts. The coefficients on $CAShort \times Download$ are statistically significant in all models. For example, in model 4 the coefficient is -0.036 (t-stat = 2.99).

Moreover, the predictability of analysts' revision is stronger for fiscal year 2. The coefficients on $CAShort$ and $CAShort \times R\_negative$ are significantly negative in models 3 and 4. The corresponding coefficients are -0.026 (t-stat = 2.03) and -0.037 (t-stat = 2.12),

Electronic copy available at: https://ssrn.com/abstract=3683086

respectively. The results suggest that the revisions of analysts' forecasts for fiscal year 2 is directly captured by the short selling activity with negative words in the annual reports.

[Table 8 here]

5.2 Fundamental ratios

Prior studies have found that short sellers base on their shorting decisions on fundamental analysis (Dechow, 2001; Curtis and Fargher, 2014; Deshmukh et al., 2015). As a result, we investigate whether textual information used by short sellers is related to changes in firm fundamentals ($\Delta Fundamental_{i,y}$) from the current fiscal year y to year y+1 by replacing $\Delta FEPS1_{i,t}$ with $\Delta Fundamental_{i,y}$ in Eq. (12). Our $\Delta Fundamental$ measures include ΔROA, $\Delta AssetTurn$, and ΔOPM.

Table 9 reports the results. Among all the textual variables, only $R\_negative$ is positively related to changes in return-on-assets (ΔROA) and in asset turnover ($\Delta AssetTurn$). The coefficients on ΔROA and $\Delta AssetTurn$ are 0.047 (*t*-stat = 2.65) and 0.097 (*t*-stat = 4.34) in Models 1 and 3, respectively. The results are consistent with those in Sections 3 and 4. It suggests that more negative words in annual reports actually indicate that the firm's fundamentals will improve over the next fiscal year. The coefficients on $Download$ are negative for all measures of $\Delta Fundamental$, suggesting that the firm's fundamentals are less likely to improve over the next fiscal year if there are more requests of 10-Ks by hedge funds. This suggests that short sellers are able to identify firms with deteriorating fundamentals when they engage more in information acquisition activities on such firms.

[Table 9 here]

6. Crash risk

Electronic copy available at: https://ssrn.com/abstract=3683086

Callen and Fang (2015) find that short interest is positively related to one-year ahead stock price crash risk, and this relation is due to bad news hoarding by firm managers. Using Regulation SHO as a natural experiment, Deng et al. (2020) find that the lifting of short-sale constraints leads to a significant decrease in stock price crash. In addition, using earnings management as a proxy for opacity, Hutton et al. (2009) find that opaque firms are more prone to stock price crashes. Kim, Wang, and Zhang (2019) find that less readable 10-K reports are related to higher stock price crash risk. They argue that managers can successfully hide adverse information by writing complex financial reports, which leads to stock price crashes when the hidden bad news accumulates and reaches a tipping point. Motivated by these studies, we conjecture that short sellers may extract textual information from annual reports that can help them predict a firm's future crash risk ($CrashRisk_{i,y}$). To test this hypothesis, we replace $\Delta FEPS1_{i,t}$ with $CashRisk_{i,y}$ in Eq. (12). Our $CrashRisk$ measures include $NSKEW$, $DUVolR$, and n_Crash.

Table 10 presents the evidence of the predictability of crash risk using abnormal shorting volume and textual information. Model 1 ($DUVolR$) shows that the coefficient on $R\_uncertainty$ is -0.048 (*t*-stat = 1.92) and $CAShort \times R\_uncertainty$ is 0.021 (*t*-stat = 2.00), whereas Model 3 ($NSkew$) reveals that the corresponding coefficients are -0.046 (*t*-stat = 1.78) and is 0.018 (*t*-stat = 1.73), respectively. Interestingly, $R\_uncertainty$ is marginally negatively related to crash risk, which means that fewer uncertainty words in annual reports are associated with higher crash risk in the coming year. This finding is consistent with the literature that crash risk is caused by bad news hoarding (Callen and Fang, 2015; Kim et al., 2019). The positive coefficients on $CAShort \times R\_uncertainty$ suggest that short sellers can potentially identify

Electronic copy available at: https://ssrn.com/abstract=3683086

firms with increasing crash risk through focusing on the frequency of uncertainty words in the annual reports.

[Table 10 here]

7. Conclusion

Using textual data from annual reports and daily shorting volume data from NYSE/Amex/Nasdaq over 2009-2015, we find that more uncertainty and negative words in annual reports are associated with greater abnormal shorting volume. Short selling motivated by textual information negatively predicts stock price reaction around the filing date of 10-Ks. Further analysis shows that textual information used by short sellers are related to the revisions of analysts' earnings forecasts, changes in firm fundamentals, as well as increasing crash risk subsequently. Our results suggest that textual information in annual reports forms an important part of short sellers' information advantage.

Electronic copy available at: https://ssrn.com/abstract=3683086

References

Aitken, Michael J., Alex Frino, Michael S. McCorry, and Peter L. Swan, 1998, Short sales are almost instantaneously bad news: Evidence from the Australian Stock Exchange, Journal of Finance 53, 2205-2223.

Arnold, Tom, Alexander W. Butler, Timothy F. Crack, and Yan Zhang, 2005, The information content of short interest: A natural experiment, Journal of Business 78, 1307-1335.

Bae, Kee-Hong, Chanwoo Lim, and K.C. John Wei, 2006, Corporate governance and conditional skewness in the world's stock markets, Journal of Business 79, 2999-3028.

Beneish, M.D., C. M.C. Lee, and D. C. Nichols, 2015, In short supply: Short-sellers and stock returns, Journal of Accounting Economics 60, 33-57.

Berkman, Henk, Valentin Dimitrov, Prem C. Jain, Paul D. Koch, and Sheri Tice, 2009, Sell on the news: Differences of opinions, short-sales constraints, and returns around earnings announcements, Journal of Financial Economics 92, 376-398.

Berkman, Henk, and Marco Eugster, 2017, Short on drugs: Short selling during the drug development process, Journal of Financial Markets 33, 102-123.

Berkman, Henk, Michael D. McKenzie, and Patrick Verwijmeren, 2017, Hole in the wall: Informed short selling ahead of private placements, Review of Finance 21(3), 1047-1091.

von Beschwitz, Bastian, Oleg Chuprinin, and Massimo Massa, 2017, Why do short sellers lke qualitative news? Journal of Financial and Quantitative Analysis 52, 645-675.

Boehme, Rodney D., Bartley R. Danielsen, and Sorin M. Sorescu, 2006, Short sales constraints, differences of opinion, and overvaluation, Journal of Financial and Quantitative Analysis 41,455-487.

Boehmer, Ekkehart, Charles M. Jones, and Xiaoyan Zhang, 2008, Which shorts are informed?, Journal of Finance 63, 491-528.

Boubaker, Sabri, Dimitrios Gounopoulos, and Hatem Rjiba, 2019, Annual report readability and stock liquidity, Financial Markets, Institutions, and Instruments 28, 159-186.

Brown, Stephen, and Stephen A. Hillegeeist, 2007, How disclosure quality affects the level of information asymmetry, Review of Accounting Studies 12, 443-477.

Brown, Stephen V., and Jennifer Wu Tucker, 2011, Large-sample evidence on firms' year-over-year MD&A modifications, Journal of Accounting Research 49 (2), 309-346.

Buehlmaier, Matthias M., and Toni M. Whited, 2018, Are financial constraints priced? Evidence from textual analysis, Review of Financial Studies 31 (7), 2693-2728.

Electronic copy available at: https://ssrn.com/abstract=3683086

Callen, Jeffrey L., and Xiaohua Fang, 2015, Short interest and stock price crash risk, Journal of Banking and Finance 60, 181-194.

Celment, Michael B., Jeffrey Hales, and Yanfeng Xue, 2011, Understanding analysts' use of stock returns and other analysts' revisions when forecasting earnings, Journal of Accounting and Economics 51, 279-299.

Chen, Honghui, Hemang Desai, and Srinivasan Krishnamurthy, 2013, A first look at mutual funds that use short sales, Journal of Financial and Quantitative Analysis 48, 761-787.

Chen, Joseph, Harrison Hong, and Jeremy C. Stein, 2002, Breadth of ownership and stock returns, Journal of Financial Economics 66, 171-205.

Cheung, Adrian W. K., Hung Wan Kot, Eric F. Y. Lam, and Harry K. M. Leung, 2019, Toward understanding short-selling activity: Demand and supply, Accounting and Finance, in press.

Choi, Jaewon, Ji Min Park, Neil D. Pearson, and Shastri Sandy, 2017, A first glimpse into the short side of hedge funds, Working paper, UIUC.

Christophe, Stephen E., Michael G. Ferri, and James J. Angel, 2004, Short-selling prior to earnings announcement, Journal of Finance 59, 1845-1875.

Crane, Alan, Kevin Crotty, and Tarik Umar, 2018, Do hedge funds profit from public information? Working paper, Rice University.

Curtis, Asher, and Neil Fargher, 2014, Does short-selling amplify price declines or align stocks with their fundamental values? Management Science 60, 2324-2340.

Dechow, Patricia M., Amy P. Hutton, Lisa Meulbroek, and Richard G. Sloan, 2001, Short-sellers, fundamental analysis, and stock returns, Journal of Financial Economics 61, 77-106.

Deng, Xiaohu, Lei Gao, and Jeong-Bon Kim, 2020, Short-sale constraints and stock price crash risk: Causal evidence from a natural experiment, Journal of Corporate Finance, in press.

Desai, Hemang, K. Ramesh, S. Ramu Thiagarajan, and Bala V. Balachandran, 2002, An investigation of the informational role of short interest in the Nasdaq market, Journal of Finance 57, 2263-2287.

Desai, Hemang, Srinivasan Krishnamurthy, and Kumar Venkataraman, 2006, Do short sellers target firms with poor earnings quality? Evidence from earnings restatements, Review of Accounting Studies 11, 71-90.

Deshmukh, Sanjay, Keith J. Jamble, and Keith M. Howe, 2015, Short selling and firm operating performance, Financial Management 44, 217-236.

Diamond, Douglas W., 1985, Optimal release of information by firms, Journal of Finance 40 (4), 1071-1094.

Electronic copy available at: https://ssrn.com/abstract=3683086

Diamond, Douglas W., and Robert E. Verrecchia, 1991, Disclosure, liquidity, and the cost of capital, Journal of Finance 46 (4), 1325-1359.

Diether, Karl B., Kuan-Hui Lee, and Ingrid M. Werner, 2009, Short-sale strategies and return predictability, Review of Financial Studies 22, 575-607.

Ditto, Peter H., and David F. Lopez, 1992, Motivated skepticism: Use of differential decision criteria for preferred and nonpreferred conclusions, Journal of Personality and Social Psychology 63 (4), 568-584.

Drake, Michael S., Lynn Rees, and Edward P. Swanson, 2011, Should investors follow the prophets or the bears? Evidence on the use of public information by analysts and short sellers, The Accounting Review 86 (1), 101-130.

Drake, Michael S., James N. Myers, Linda A. Myers, and Michael D. Stuart, 2015, Short sellers and the informativeness of stock prices with respect to future earnings, Review of Accounting Studies 20, 747-774.

Drake, Michael S., Darren T. Roulstone, and Jacob R. Thornock, 2015, The determinants and consequences of information acquisition via EDGAR, Contemporary Accounting Research 32 (3), 1128-1161.

Drake, Michael S., Phillip J. Quinn, and Jacob R. Thornock, 2017, Who uses financial statements? A demographic analysis of financial statement downloads from EDGAR, Accounting Review 31 (3), 55-68.

Dyer, Travis, Mark Lang, and Lorien Stice-Lawrence, 2017, The evolution of 10-K textual disclosure: Evidence from Latent Dirichlet Allocation, Journal of Accounting and Economics 64, 221-245.

Engelbert, Joseph E., Adam V. Reed, and Matthew C. Ringgenberg, 2012, How are shorts informed? Short sellers, news, and information processing, Journal of Financial Economics 105, 260-278.

Fang, Vivian W., Allen H. Huang, and Jonathan Karpoff, 2016, Short selling and earnings management: A controlled experiment, Journal of Finance 71, 1251-1294.

Gargano, Antonio, Juan Sotes-Paladino, Partick Verwijmeren, 2018, The profits and losses of short sellers, Working paper, University of Melbourne.

Goldstein, Itay, and Liyan Yang, 2019, Good disclosure, bad disclosure, Journal of Financial Economics 131, 118-138.

Hanley, Kathleen Weiss, and Gerard Hoberg, 2010, The information content of IPPO prospectuses, Review of Financial Studies 23 (7), 2821-2864.

Hales, Jeffrey, 2007, Directional preferences, information processing, and investors' forecasts of earnings. Journal of Accounting Research 45 (3), 607-628.

Electronic copy available at: https://ssrn.com/abstract=3683086

Hirshleifer, David, 2001, Investor psychology and asset pricing, Journal of Finance 64(4), 1533-1597.

Hirshleifer, David, and Siew Hong Teoh, 2003, Limited attention, information disclosure, and financial reporting, Journal of Accounting and Economics 36, 337-386.

Hoberg, Gerard, and Gordon Phillips, 2010, Product market synergies and competition in mergers and acquisitions: A text-based analysis, Review of Financial Studies 23 (10), 3773-3811.

Hong, Harrison, and Jeremy C. Stein, 2007, Disagreement and the stock market, Journal of Economic Perspectives 21, 109–128.

Huang, Alan Guoming, Hongping Tan, and Russ Wermers, 2019, Institutional trading around corporate news: Evidence from textual analysis, Review of Financial Studies, forthcoming.

Hunton, James E., and Ruth Ann McEwen, 1997, An assessment of the relation between analysts' earnings forecast accuracy, motivational incentives and cognitive information search strategy, Accounting Review 72(4), 497-515.

Hutton, Amy P., Alan J. Marcus, and Hassan Tehranian, 2009, Opaque financial reports, R^2, and crash risk, Journal of Financial Economics 94, 67-86.

Jegadeesh, Narasimhan, and Di Wu, 2013, Word power: A new approach for content analysis, Journal of Financial Economics 110, 712-729.

Jiang, Fuwei, Joshua Lee, Xiumin Martin, and Guofu Zhou, 2019, Manager sentiment and stock returns, Journal of Financial Economics, in press.

Jiang, Wenxi, 2019, Leveraged speculators and asset prices, Working paper, Chinese University of Hong Kong.

Kandel, Eugene, and Neil D. Pearson, 1995, Differential interpretation of public signals and trade in speculative markets, Journal of Political Economy 103(4), 831-872.

Karpoff, Jonathan M., and Xiaoxia Lou, 2010, Short sellers and financial misconduct, Journal of Finance 65, 1879-1913.

Ke, Zheng Tracy, Bryan Kelly, and Dacheng Xiu, 2019, Predicting returns with text data, Working paper, Yale University.

Khan, M. and H. Lu, 2013, Do Short Sellers Front-Run Insider Sales? Accounting Review 88(5): 1743–1768.

Kim, Oliver, and Robert E. Verrecchia, 1997, Pre-announcement and event-period private information, Journal of Accounting and Economics 24, 395-419.

Electronic copy available at: https://ssrn.com/abstract=3683086

Kim, Chansog (Francis), Ke Wang, and Liandong Zhang, 2019, Readability of 10-K reports and stock price crash risk, Contemporary Accounting Research 36, 1184-1216.

Kot, Hung Wan, 2007, What determines the level of short-selling activity?, Financial Management 36 (4), 123-141.

Lang, Mark, and Lorien Stice-Lawrence, 2015, Textual analysis and international financial reporting: Large sample evidence, Journal of Accounting and Economics 60, 110-135.

Li, Frank Weikai, and Chengzhu Sun, 2019, Information acquisition and expected returns: Evidence from EDGAR search traffic, Working Paper, Singapore Management University and HKUST.

Li, Feng, 2008, Annual report readability, current earnings, and earnings persistence, Journal of Accounting and Economics 45, 221-247.

Liberti, Jose Maria, and Mitchell A. Petersen, 2019, Information: hard and soft, Review of Financial Studies 29, 1-41.

Loughran, Tim, and Bill McDonald, 2011, When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks, Journal of Finance 66(1), 35-65.

Loughran, Tim, and Bill McDonald, 2014, Measuring readability in financial disclosures, Journal of Finance 69(4), 1643-1671.

Loughran, Tim, and Bill McDonald, 2016, Textual analysis in accounting and finance: A survey, Journal of Accounting Research 54(4), 1187-1230.

Miller, Edward M., 1977, Risk, uncertainty, and divergence of opinion, Journal of Finance 32, 1151-1168.

Nagel, Stefan, 2005, Short sales, institutional investors, and the cross-section of stock returns, Journal of Financial Economics 78, 277-309.

Petersen, M. A., 2009, Estimating standard errors in finance panel data sets: Comparing approaches, Review of Financial Studies 22, 435–480.

Prado, M. P., P. A. C. Saffi, and J. Sturgess, 2016, Ownership structure, limits to arbitrage, and stock returns: Evidence from equity lending markets, Review of Financial Studies 29, 3211-3244.

Purnanandam, Amiyatosh, and Nejat Seyhun, 2018, Do short sellers trade on private information or false information?, Journal of Financial and Quantitative Analysis 53, 997-1023.

Rapach, David E., Matthew C. Ringgenberg, and Guofu Zhou, 2016, Short interest and aggregate stock returns, Journal of Financial Economics 121, 46-65.

Electronic copy available at: https://ssrn.com/abstract=3683086

Shi, Jian, Junbo Wang, and Ting Zhang, 2017, Are short sellers informed? Evidence from credit rating agency announcements, Journal of Financial Research 40(2), 179-221.

Tetlock, Paul C., 2007, Giving content to investor sentiment: The role of media in the stock market, Journal of Finance 62(3), 1139-1168.

Tetlock, Paul C., Maytal Saar-Tsechansky, and Sofus MacSkassy, 2008, More than words: Quantifying language to measure firms' fundamentals, Journal of Finance 63(3), 1437-1467.

Wang, Xue, Xuemin (Sterling) Yan, and Lingling Zheng, 2019, Shorting flows, public disclosure, and market efficiency, Journal of Financial Economics, in press.

Yan, Yumeng, Xiong Xiong, J. Ginger Meng, and Gaofeng Zou, 2019, Uncertainty and IPO initial returns: Evidence from the tone analysis of China's IPO prospectuses, Pacific-Basin Finance Journal, in press.

Electronic copy available at: https://ssrn.com/abstract=3683086

Appendix: Variable definition

Short sale variable	
CAShort	Cumulative abnormal daily shorting volume ratio during the 4-day event window [0, 3]. Daily shorting volume ratio = daily shorting volume/shares outstanding. Abnormal daily shorting volume ratio = Daily shorting volume ratio in day *t* minus the average of the ratio during the event window [-120, -21].
Download	Number of 10K downloads in EDGAR by hedge funds on the filing date.
Textual variables	
n_words	The count of all words, where a word is any token appearing in the Master Dictionary.
n_uncertainty	The number of words related to uncertainty.
n_modalweak	The number of words related to modal weak.
n_modal_strong	The number of words related to modal strong.
n_negative	The number of words related to negative.
n_positive	The number of words related to positive.
n_filing	The number of 10Ks filings per day.
R_uncertainty	n_uncertainty/n_words.
R_modalweak	n_modalweak/(n_modalweak + n_modal_strong).
R_negative	n_negative/(n_negative + n_positive).
Stock returns	
BHAR	Buy and hold abnormal return: which is the cumulative buy and hold stock returns minus the corresponding value-weighted CRSP returns in various event windows. For daily returns BHAR4d, the event window is [0, 3]. For weekly returns from 1 week to 52 weeks after the filing date, it is denoted as BHAR1w to BHAR52w.
Analysts' earnings forecasts	
$\Delta FEPS1$	Change in analysts' earnings forecasts per share for fiscal year y+1 from month *t*-1 to *t*+1, measured as $\Delta FEPS1_{i,t} = \frac{FEPS1_{i,t+1} - FEPS1_{i,t-1}}{StockPrice_{i,t-1}}$|.
$\Delta FEPS2$	Change in analysts' earnings forecasts per share for fiscal year y+2 from month *t*-1 to *t*+1, measured as $\Delta FEPS2_{i,t} = \frac{FEPS2_{i,t+1} - FEPS2_{i,t-1}}{StockPrice_{i,t-1}}$.
Fundamental variables	
ΔROA	Change in return of assets (ROA) from fiscal year *t* to fiscal year *t*+1.
$\Delta AssetTurn$	Change in asset turnover from fiscal year *t* to fiscal year *t*+1, where asset turnover is sales divided by assets.
ΔOPM	Change in operating profit margin before depreciation measured from fiscal year *t* to fiscal year *t*+1.

Electronic copy available at: https://ssrn.com/abstract=3683086

Crash risk	
NSkew	The negative of the third central moment of firm-specific weekly return divided by the variance of firm-specific weekly returns raised to the power of 3/2. A higher NSkew corresponds to a more negative-skewed stock return distribution and higher crash risk.
DUVolR	Down-to-up return volatility ratio and is measured as $DUVolR_{i,y} = Ln\left\{\frac{n_{up}(\sum_{w\in Down}(Ret_{i,w,y} - \overline{Ret}_{i,y})^2)}{n_{down}(\sum_{w\in Up}(Ret_{i,w,y} - \overline{Ret}_{i,y})^2)}\right\}$. An up (down) week is defined as a week when the firm-specific weekly return is above (below) the annual mean. A higher value of DUVolR indicates a higher crash risks.
n_Crash	The difference in the frequencies between extreme negative returns and extreme positive returns based on the number of firm-specific weekly returns exceeding 3.09 standard deviations above and below the mean firm-specific weekly return over the fiscal year. A higher value of n_Crash corresponds to a higher frequency of crashes.
Control variables	
B/M	Book-to-market equity ratio.
SZ	Market capitalization in million dollars.
IOwner	Institutional ownership scaled by the number of outstanding shares.
Illiq	Ahumid's illiquidity measure.
Pr1y	Cumulated stock returns in the previous one year.
IVol	Idiosyncratic volatility, which is the mean squared error of residuals of daily stock returns from the Fama-French three-factor model augmented by the Carhart momentum based on return data from the past three months.

Electronic copy available at: https://ssrn.com/abstract=3683086

Table 1: Sample construction

This table reports the details of the sample construction from the initial 10-Ks sample. CIK is the Central Index Key assigned by the SEC. PERMNO is the permanent issue identification number assigned by the CRSP. We obtain the initial 10-Ks textual data from the website of The Notre Dame Software Repository for Accounting and Finance, daily shorting volume data from the FINRA website, and stock price, returns, shares outstanding, trading volume, and accounting data from the CRSP/Compustat.

Source/Filter	Sample size	Observations removed
Original textual data downloaded from Loughran-McDonald 10X File summaries file. Fiscal year ended is from 1988.12.31 – 2018.12.06. Filing date is from 1993.11.29 – 2018.12.30.	1,028,674	
Keep form types for 10-K, 10-K405, 10KSB, and 10KSB40 only	242,180	786,494
Number of words in 10-Ks >= 2,000	235,531	6,649
Exclude if fiscal year end is missed	234,349	1,182
Drop the duplicated firms' fiscal year end or filing date	234,266	83
Drop if the current filing date and previous filing date is < 180	231,565	2,701
Drop if filing date is same as fiscal year end	230,325	1.240
Merge with monthly stock returns and control variables by PERMNO and YYYYMM for 1996.1 – 2017.12	94,896	
Daily shorting volume from 2009.8 – 2018.12	8,928,481	
Merge with monthly file by PERMNO and filing date, merged sample period is 2009.8 – 2015.12	19,645	

Electronic copy available at: https://ssrn.com/abstract=3683086

Table 2: Summary statistics

This table reports the summary statistics of variables used in the study. The sample period is from September 2009 to December 2015. All variables are defined in the Appendix. We obtain the initial 10-Ks textual data from the website of The Notre Dame Software Repository for Accounting and Finance, daily shorting volume data from the FINRA website, the number of 10-Ks requests by hedge funds from the EDGAR system, and stock price, returns, shares outstanding, trading volume, and accounting data from the CRSP/Compustat.

	N	Mean	Std Dev	Min	p5	p50	p95	Max
CAShort	19,081	0.0011	0.0048	-0.0384	-0.0033	0	0.0096	0.0593
BHAR4d	19,544	-0.0004	0.0562	-0.2605	-0.0972	-0.0014	0.094	0.2891
n_words	19,645	53,246	32,979	3,033	23,019	45,833	108,301	1,034,542
n_uncertainty	19,645	732	359	9	306	672	1,322	5,407
n_modalweak	19,645	324	186	1	104	287	670	3,286
n_modalstrong	19,645	158	132	7	49	127	354	5,354
n_negative	19,645	970	658	27	320	823	2,086	15,892
n_positive	19,645	378	230	9	140	328	780	3,665
n_filing	19,645	89	80	1	3	64	232	315
R_uncertainty	19,645	0.0143	0.0028	0.0027	0.0096	0.0143	0.0187	0.0256
R_modalweak	19,645	0.6774	0.0769	0.0556	0.5406	0.6862	0.7879	0.8878
R_negtative	19,645	0.7106	0.0614	0.3995	0.6025	0.7169	0.7994	0.9607
Download	18,183	20.9	17.6	0	2	16	57	100
Firm size (SZ)	19,630	4,335	17,628	2	21	518	17,242	666,252
B/M	17,909	0.8638	1.1992	0.001	0.1186	0.6407	2.31	75.9019
IOwner	18,618	0.5227	0.3270	0	0.0021	0.5880	0.9495	0.9993
Illiq	19,643	0.9593	3.5421	0	0.0001	0.0053	5.6806	19.7147
Pr1y	18,864	0.2594	0.7124	-0.9649	-0.4862	0.1434	1.3183	8.125
IVol	19,611	0.0239	0.0171	0.0019	0.0078	0.0192	0.0556	0.5437

Electronic copy available at: https://ssrn.com/abstract=3683086

Table 3: Correlation matrix

This table reports the Pearson correlation coefficients of variables used in the study. The sample period is from September 2009 to December 2015. All variables are defined in the Appendix. We obtain the initial 10-Ks textual data from the website of The Notre Dame Software Repository for Accounting and Finance, daily shorting volume data from the FINRA website, the number of 10-Ks requests by hedge funds from the EDGAR system, and stock price, returns, shares outstanding, trading volume, and accounting data from the CRSP/Compustat. * indicates *p*-value < 0.05.

		(1) CAShort	(2)	(3)	(4)	(5)	(5)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
(2)	BHAR4d	0.12*												
(3)	R_uncertainty	0.00	0.01											
(4)	R_modalweak	0.01	0.02*	0.53*										
(5)	R_negative	0.01	0.01	0.06*	0.11*									
(6)	Ln(n_words)	0.04*	0.00	-0.25*	-0.11*	0.21*								
(7)	Ln(n_filing)	-0.01	-0.02*	0.03*	-0.02*	0.09*	0.21*							
(8)	Ln(Download)	0.06*	0.02*	-0.03*	0.01	0.02*	0.13*	-0.32*						
(9)	Ln(SZ)	-0.01	0.06*	0	0.12*	-0.09*	0.41*	0.12*	0.27*					
(10)	Ln(B/M)	-0.03*	0.01	-0.07*	-0.09*	0.22*	-0.01	0.02*	-0.02*	-0.32*				
(11)	IOwner	0.03*	0.07*	0.07*	0.19*	-0.07*	0.21*	0.05*	0.12*	0.61*	-0.21*			
(12)	Illiq	-0.05*	-0.02*	-0.06*	-0.12*	0.06*	-0.20*	-0.10*	-0.07*	-0.41*	0.22*	-0.34*		
(13)	Pr1y	0.05*	0.05*	-0.02*	-0.01	-0.02*	0.00	0.05*	0.06*	0.05*	0.14*	0.04*	-0.07*	
(14)	IVol	0.22*	-0.02*	0.01	-0.08*	0.02	-0.11*	-0.06*	-0.07*	-0.48*	0.08*	-0.37*	0.27*	0.05*

Electronic copy available at: https://ssrn.com/abstract=3683086

Table 4: Determinants of abnormal shorting volume

This table reports the results of the following abnormal shorting volume regression:

$$CAShort_{i,t} = \alpha + \sum_{j=1}^{3} \beta_j Textual_{i,j,t} + \beta_4 Download_{i,t} + \sum_{j=1}^{3} \theta_j Download_{i,t} \times Textual_{i,j,t}$$
$$+\gamma_1 Ln(n\_words)_{i,t} + \gamma_2 Ln(n\_filing)_t + control_{i,t} + d_{m,t} + \varepsilon_{i,t},$$

where $CAShort_{i,t}$ is the 4-day cumulative abnormal shorting volume during the event window [0, 3]. *Textual* includes *R_uncertainty*, *R_modalweak*, and *R_negative*. All variables are defined in the Appendix and are standardized to have the mean of 0 and the standard deviation of 1. We obtain the initial 10-Ks textual data from the website of The Notre Dame Software Repository for Accounting and Finance, daily shorting volume data from the FINRA website, the number of 10-Ks requests by hedge funds from the EDGAR system, and stock price, returns, shares outstanding, trading volume, and accounting data from the CRSP/Compustat. All models include stock fixed-effects and year-month dummy variables. Standard errors are double clustered at the firm and year-month levels. *, **, *** indicate significance at the 10%, 5%, and 1% levels, respectively. The sample period is from September 2009 to December 2015.

	(1) CAShort	(2) CAShort	(3) CAShort	(4) CAShort	(5) CAShort	(6) CAShort
R_uncertainty	0.047**	0.046*	0.040*		0.046*	
	(2.22)	(1.92)	(1.77)		(1.90)	
R_negative	-0.023	-0.026	-0.042**	-0.040**	-0.026	-0.040**
	(-1.41)	(-1.50)	(-2.41)	(-2.39)	(-1.49)	(-2.40)
R_modalweak				0.003		0.003
				(0.19)		(0.19)
Download		0.115***	0.098***	0.098***	0.114***	0.098***
		(5.06)	(4.64)	(4.65)	(5.13)	(4.76)
Download ×R_uncertainty					-0.004	
					(-0.34)	
Download ×R_negative					0.002	0.002
					(0.15)	(0.13)
Download ×R_modalweak						-0.000
						(-0.01)
Ln(n_words)	0.069***	0.062**	0.048**	0.015	0.062**	0.015
	(3.18)	(2.63)	(2.29)	(1.20)	(2.59)	(1.21)
Ln(n_filing)	-0.063**	-0.008	-0.005	-0.005	-0.008	-0.005
	(-2.50)	(-0.28)	(-0.18)	(-0.17)	(-0.28)	(-0.18)
Ln(SZ)	0.026	-0.017			-0.017	
	(0.33)	(-0.21)			(-0.20)	
Ln(B/M)	-0.013	-0.007	-0.031	-0.030	-0.007	-0.030
	(-0.50)	(-0.28)	(-1.50)	(-1.46)	(-0.28)	(-1.45)
Pr1y	0.046***	0.049***	0.038***	0.038***	0.049***	0.038***
	(3.05)	(3.07)	(2.75)	(2.74)	(3.07)	(2.75)
IOwner			0.042	0.043		0.043
			(1.61)	(1.66)		(1.66)
Illiq			-0.041***	-0.041***		-0.041***
			(-4.07)	(-4.06)		(-4.05)
IVol			0.333***	0.333***		0.333***
			(6.43)	(6.44)		(6.44)
Intercept	-0.011***	-0.014**	0.013***	0.011***	-0.015**	0.011***
	(-3.53)	(-2.22)	(3.23)	(2.82)	(-2.21)	(2.81)
Observations	17,290	16,006	15,995	15,995	16,006	15,995
Adjusted R^2	0.17	0.18	0.22	0.22	0.18	0.22

Electronic copy available at: https://ssrn.com/abstract=3683086

Table 5: Return predictability of abnormal shorting volume and textual information around filing dates

This table reports the results of the following cumulative abnormal return regression:

$$BHAR4d_{i,t} = \alpha + \beta_0 CAShort_{i,t} + \sum_{j=1}^{3} \beta_j Textual_{i,j,t} + \beta_4 Download_{i,t} + \beta_5 Ln(n\_words)_t$$
$$+\beta_6 Ln(n\_filing)_{i,t} + \sum_{j=1}^{3} \theta_j CAShort \times Textual_{i,j,t}$$
$$+\theta_4 CAShort \times Download_{i,j,t} + control_{i,t} + d_{m,t} + \varepsilon_{i,t},$$

where $BHAR4d_{i,t}$ ($CAShort_{i,t}$) is the 4-day cumulative abnormal return (shorting volume) during the event window [0, 3]. *Textual* includes *R_uncertainty*, *R_modalweak*, and *R_negative*. All variables are defined in the Appendix and are standardized to have the mean of 0 and the standard deviation of 1. We obtain the initial 10-Ks textual data from the website of The Notre Dame Software Repository for Accounting and Finance, daily shorting volume data from the FINRA website, the number of 10-Ks requests by hedge funds from the EDGAR system, and stock price, returns, shares outstanding, trading volume, and accounting data from the CRSP/Compustat. All models include stock fixed-effects and year-month dummy variables. Standard errors are double clustered at the firm and year-month levels. *, **, *** indicate significance at the 10%, 5%, and 1% levels, respectively. The sample period is from September 2009 to December 2015.

	(1) BHAR4d	(2) BHAR4d	(3) BHAR4d	(4) BHAR4d	(5) BHAR4d	(6) BHAR4d
CAShort	0.154***		0.155***	0.152***	0.170***	0.167***
	(5.99)		(6.20)	(6.51)	(6.18)	(6.56)
R_uncertainty		-0.046**	-0.054**		-0.044*	
		(-2.14)	(-2.46)		(-1.80)	
R_negative		0.037**	0.042**	0.043**	0.050**	0.051**
		(2.01)	(2.16)	(2.18)	(2.57)	(2.60)
R_modalweak				-0.039***		-0.037***
				(-3.48)		(-3.03)
Download					-0.054***	-0.054***
					(-3.86)	(-3.82)
CAShort×R_uncertainty			-0.012		-0.013	
			(-0.73)		(-0.75)	
CAShort×R_negative			-0.002	0.001	0.007	0.011
			(-0.14)	(0.07)	(0.42)	(0.68)
CAShort×R_modalweak				-0.035**		-0.042**
				(-2.21)		(-2.31)
CAShort×Download					-0.046**	-0.048***
					(-2.59)	(-2.77)
Ln(n_words)	-0.030	-0.078***	-0.090***	-0.060***	-0.078***	-0.057**
	(-1.66)	(-2.71)	(-3.09)	(-2.79)	(-2.65)	(-2.63)
Ln(n_filing)	-0.067***	-0.074***	-0.066***	-0.066***	-0.090***	-0.090***
	(-2.82)	(-3.15)	(-2.79)	(-2.76)	(-3.37)	(-3.36)
Ln(SZ)	0.460***	0.479***	0.477***	0.479***	0.475***	0.477***
	(6.20)	(6.34)	(6.42)	(6.54)	(7.04)	(7.18)
Ln(B/M)	0.152***	0.149***	0.153***	0.150***	0.152***	0.150***
	(6.08)	(5.83)	(5.99)	(5.95)	(5.26)	(5.23)
Pr1y	-0.037*	-0.031	-0.039*	-0.038**	-0.038**	-0.038**
	(-1.92)	(-1.57)	(-1.98)	(-2.00)	(-2.00)	(-2.02)
Intercept	-0.011***	-0.016***	-0.015***	-0.013***	-0.023***	-0.021***
	(-4.12)	(-5.84)	(-4.93)	(-4.88)	(-4.58)	(-4.45)
Observations	17,288	17,322	17,288	17,288	16,006	16,006
Adjusted R^2	0.09	0.07	0.09	0.09	0.09	0.09

Electronic copy available at: https://ssrn.com/abstract=3683086

Table 6: Return predictability of abnormal shorting volume and textual information in 1 to 52-weeks ahead

This table reports the results of following cumulative abnormal return regression in 1 50 52 weeks ahead:

$$BHAR1w \ldots 52w_{i,t} = \alpha + \beta_0 CAShort_{i,t} + \sum_{j=1}^{3} \beta_j Textual_{i,j,t} + \beta_4 Download_{i,t}$$
$$+\beta_5 Ln(n\_words)_t + \beta_6 Ln(n\_filing)_{i,t} + \sum_{j=1}^{3} \theta_j CAShort \times Textual_{i,j,t}$$
$$+\theta_4 CAShort \times Download_{i,j,t} + control_{i,t} + d_{m,t} + \varepsilon_{i,t},$$

where $BHAR1w \ldots 52w_{i,t}$ ($CAShort_{i,t}$) is the 1-week…52-week cumulative abnormal return (the cumulative shorting volume during the event window [0, 3]). *Textual* includes *R_uncertainty*, *R_modalweak*, and *R_negative*. All variables are defined in the Appendix and are standardized to have the mean of 0 and the standard deviation of 1. We obtain the initial 10-Ks textual data from the website of The Notre Dame Software Repository for Accounting and Finance, daily shorting volume data from the FINRA website, the number of 10-Ks requests by hedge funds from the EDGAR system, and stock price, returns, shares outstanding, trading volume, and accounting data from the CRSP/Compustat. All models include stock fixed-effects and year-month dummy variables. Standard errors are double clustered at the firm and year-month levels. *, **, *** indicate significance at the 10%, 5%, and 1% levels, respectively. The sample period is from September 2009 to December 2015.

	(1)	(2)	(3)	(4)	(5)	(6)
	BHAR1w	BHAR1w	BHAR2w	BHAR2w	BHAR4w	BHAR4w
CAShort	0.126***	0.125***	0.097***	0.096***	0.073***	0.072***
	(5.81)	(5.99)	(5.17)	(5.37)	(4.90)	(5.09)
R_uncertainty	-0.033**		-0.014		-0.028	
	(-2.02)		(-0.60)		(-1.49)	
R_negative	0.047**	0.048**	0.039*	0.040*	0.035*	0.035*
	(2.52)	(2.60)	(1.89)	(1.97)	(1.79)	(1.82)
R_modalweak		-0.032**		-0.027**		-0.020
		(-2.50)		(-2.13)		(-1.60)
Download	-0.021	-0.021	-0.021*	-0.021*	-0.021	-0.021
	(-1.47)	(-1.50)	(-1.73)	(-1.75)	(-1.37)	(-1.39)
CAShort×R_uncertainty	0.003		-0.006		-0.011	
	(0.20)		(-0.33)		(-0.83)	
CAShort×R_negative	-0.007	-0.006	-0.007	-0.005	-0.010	-0.008
	(-0.40)	(-0.33)	(-0.39)	(-0.29)	(-0.63)	(-0.48)
CAShort×R_modalweak		-0.017		-0.019		-0.021
		(-0.81)		(-1.09)		(-1.47)
CAShort×Download	-0.027**	-0.029**	-0.019	-0.020	-0.020*	-0.021*
	(-2.06)	(-2.30)	(-1.31)	(-1.46)	(-1.77)	(-1.90)
Ln(n_words)	-0.065***	-0.050**	-0.058**	-0.057***	-0.057**	-0.041**
	(-2.69)	(-2.36)	(-2.09)	(-3.31)	(-2.36)	(-2.47)
Ln(n_filing)	-0.061**	-0.060**	-0.077***	-0.077***	-0.094***	-0.093***
	(-2.11)	(-2.11)	(-2.80)	(-2.82)	(-3.91)	(-3.94)
Ln(SZ)	0.394***	0.396***	0.331***	0.334***	0.111	0.112
	(4.74)	(4.76)	(2.90)	(2.95)	(0.84)	(0.85)
Ln(B/M)	0.172***	0.171***	0.227***	0.226***	0.199***	0.197***
	(6.00)	(6.02)	(7.46)	(7.55)	(7.04)	(7.05)
Pr1y	-0.061***	-0.061***	-0.062**	-0.062**	-0.054**	-0.054**
	(-3.02)	(-3.03)	(-2.55)	(-2.56)	(-2.22)	(-2.21)
Intercept	-0.017***	-0.016***	-0.014*	-0.013	0.004	0.005
	(-3.02)	(-2.80)	(-1.67)	(-1.66)	(0.38)	(0.51)
Observations	16,004	16,004	16,004	16,004	16,001	16,001
Adjusted R^2	0.06	0.07	0.07	0.07	0.08	0.08

Electronic copy available at: https://ssrn.com/abstract=3683086

	(7) BHAR12w	(8) BHAR12w	(9) BHAR24w	(10) BHAR24w	(11) BHAR52w	(12) BHAR52w
CAShort	0.041***	0.040***	0.039***	0.038***	-0.004	-0.005
	(2.89)	(2.80)	(2.79)	(2.74)	(-0.50)	(-0.58)
R_uncertainty	-0.051**		-0.041*		-0.016	
	(-2.04)		(-1.92)		(-0.83)	
R_negative	0.025	0.023	-0.001	-0.002	-0.034*	-0.033*
	(1.61)	(1.48)	(-0.05)	(-0.13)	(-1.76)	(-1.73)
R_modalweak		0.001		-0.005		-0.024*
		(0.07)		(-0.36)		(-1.80)
Download	-0.013	-0.013	-0.042***	-0.043***	-0.057***	-0.058***
	(-0.82)	(-0.83)	(-2.89)	(-2.91)	(-3.77)	(-3.80)
CAShort×R_uncertainty	-0.028**		-0.022		-0.017	
	(-2.05)		(-1.58)		(-1.28)	
CAShort×R_negative	-0.007	-0.004	-0.009	-0.006	-0.015	-0.013
	(-0.46)	(-0.25)	(-0.61)	(-0.45)	(-1.15)	(-1.07)
CAShort×R_modalweak		-0.020		-0.015		-0.009
		(-1.34)		(-1.17)		(-0.78)
CAShort×Download	-0.009	-0.008	-0.021*	-0.019	-0.005	-0.003
	(-0.66)	(-0.56)	(-1.78)	(-1.66)	(-0.43)	(-0.31)
Ln(n_words)	-0.098***	-0.054***	-0.063***	-0.029**	-0.027	-0.023
	(-3.63)	(-4.37)	(-2.96)	(-2.31)	(-1.22)	(-1.40)
Ln(n_filing)	-0.074***	-0.074***	-0.052**	-0.052**	-0.028	-0.028
	(-3.32)	(-3.34)	(-2.59)	(-2.59)	(-1.42)	(-1.40)
Ln(SZ)	-0.572***	-0.578***	-0.851***	-0.855***	-1.727***	-1.725***
	(-5.83)	(-5.98)	(-5.10)	(-5.15)	(-8.20)	(-8.17)
Ln(B/M)	0.147***	0.144***	0.156***	0.153***	0.133***	0.131***
	(4.45)	(4.38)	(5.07)	(5.01)	(3.58)	(3.53)
Pr1y	-0.034*	-0.033	-0.023	-0.022	0.002	0.002
	(-1.69)	(-1.66)	(-1.43)	(-1.38)	(0.12)	(0.13)
Intercept	0.044***	0.047***	0.085***	0.087***	0.182***	0.183***
	(5.84)	(6.51)	(6.74)	(6.92)	(10.22)	(10.15)
Observations	15,976	15,976	15,934	15,934	15,766	15,766
Adjusted R^2	0.11	0.10	0.13	0.13	0.22	0.22

Electronic copy available at: https://ssrn.com/abstract=3683086

Table 7: Return predictability using fitted abnormal shorting volume

This table reports the results from the following regression of cumulative abnormal returns on fitted cumulative abnormal shorting volume:

$$BHAR4d_{i,t}\ (BHAR1w \ldots 52w_{i,t}) = \alpha + \beta_1 \widehat{CAShort}_{i,t} + \beta_2 Ln(n\_words)_t + \beta_3 Ln(n\_filing)_t$$
$$+\beta_4 Download_{i,t} + control_{i,t} + d_{m,t} + \varepsilon_{i,t},$$

where $\widehat{CAShort}_{i,t}$ is the fitted cumulative abnormal shorting volume during the event window [0, 3] and is obtained from the following regression,

$$\widehat{CAShort}_{i,t} = \hat{\alpha} + \sum_{j=1}^{3} \hat{\beta}_j Textual_{i,j,t},$$

where $\hat{\alpha}$ and $\hat{\beta}_j$ are the estimates from the above equation. *Textual* includes *R_uncertainty*, *R_modalweak*, and *R_negative*. All variables are defined in the Appendix and are standardized to have the mean of 0 and the standard deviation of 1. We obtain the initial 10-Ks textual data from the website of The Notre Dame Software Repository for Accounting and Finance, daily shorting volume data from the FINRA website, the number of 10-Ks requests by hedge funds from the EDGAR system, and stock price, returns, shares outstanding, trading volume, and accounting data from the CRSP/Compustat. All models include stock fixed-effects and year-month dummy variables. Standard errors are double clustered at the firm and year-month levels. *, **, *** indicate significance at the 10%, 5%, and 1% levels, respectively. The sample period is from September 2009 to December 2015.

	(2) BHAR4d	(3) BHAR1w	(4) BHAR2w	(6) BHAR4w	(7) BHAR12w	(8) BHAR24w	(9) BHAR52w
$\widehat{CAShort}_{i,t}$	-0.042***	-0.035***	-0.028**	-0.024*	-0.006	-0.012	-0.026**
	(-2.92)	(-2.84)	(-2.13)	(-1.94)	(-0.43)	(-0.76)	(-2.07)
Ln(n_words)	-0.046**	-0.039**	-0.047***	-0.034**	-0.046***	-0.032**	-0.038**
	(-2.56)	(-2.11)	(-3.02)	(-2.00)	(-3.75)	(-2.35)	(-2.65)
Ln(n_filing)	-0.085***	-0.055**	-0.072***	-0.091***	-0.072***	-0.049**	-0.026
	(-3.35)	(-2.10)	(-2.77)	(-3.97)	(-3.38)	(-2.61)	(-1.36)
Download	-0.040***	-0.010	-0.013	-0.015	-0.011	-0.041***	-0.058***
	(-2.92)	(-0.74)	(-0.97)	(-1.05)	(-0.71)	(-2.79)	(-3.85)
Ln(SZ)	0.449***	0.361***	0.314***	0.104	-0.577***	-0.824***	-1.661***
	(6.61)	(4.46)	(2.77)	(0.78)	(-6.05)	(-5.09)	(-8.07)
Ln(B/M)	0.147***	0.163***	0.220***	0.195***	0.141***	0.148***	0.125***
	(5.35)	(6.11)	(7.57)	(6.95)	(4.40)	(5.04)	(3.46)
Pr1y	-0.027	-0.048***	-0.051**	-0.046**	-0.028	-0.019	0.002
	(-1.56)	(-2.72)	(-2.33)	(-2.04)	(-1.55)	(-1.32)	(0.13)
Intercept	-0.035***	-0.032***	-0.024***	-0.012**	0.010**	0.024***	0.086***
	(-9.42)	(-7.98)	(-4.63)	(-2.10)	(2.27)	(3.21)	(6.96)
Observations	16,034	16,032	16,032	16,029	16,004	15,962	15,792
Adjusted R^2	0.07	0.05	0.06	0.07	0.10	0.13	0.22

Electronic copy available at: https://ssrn.com/abstract=3683086

Table 8: Revisions of analysts' earnings forecasts around 10-K filing dates

This table reports the results of the following regression on the revisions of analysts' earnings forecasts:

$$\Delta FEPS1_{i,t}(\Delta FEPS2_{i,t}) = \alpha + \beta_0 CAShort_{i,t} + \sum_{j=1}^{3}\beta_j Textual_{i,j,t} + \beta_4 Download_{i,t}$$
$$+\beta_5 Ln(n\_words)_t + \beta_6 Ln(n\_filing)_{i,t} + \sum_{j=1}^{3}\theta_j CAShort \times Textual_{i,j,t}$$
$$+\theta_4 CAShort \times Download_{i,j,t} + control_{i,t} + d_{m,t} + \varepsilon_{i,t},$$

where $\Delta FEPS1$ ($\Delta FEPS2$) is the revision of analysts' consensus earnings forecasts per share for fiscal year 1 (2) earnings from month *t*-1 to month *t*+1, where *t* is the 10-K filing month. *Textual* includes *R_uncertainty*, *R_modalweak*, and *R_negative*. All variables are defined in the Appendix and are standardized to have the mean of 0 and the standard deviation of 1. We obtain the initial 10-Ks textual data from the website of The Notre Dame Software Repository for Accounting and Finance, daily shorting volume data from the FINRA website, the number of 10-Ks requests by hedge funds from the EDGAR system, and stock price, returns, shares outstanding, trading volume, and accounting data from the CRSP/Compustat. All models include stock fixed-effects and year-month dummy variables. Standard errors are double clustered at the firm and year-month levels. *, **, *** indicate significance at the 10%, 5%, and 1% levels, respectively. The sample period is from September 2009 to December 2015.

	(1) $\Delta FEPS1$	(2) $\Delta FEPS1$	(3) $\Delta FEPS2$	(4) $\Delta FEPS2$
CAShort	0.008	0.007	-0.026**	-0.026**
	(0.54)	(0.51)	(-2.04)	(-2.03)
R_uncertainty	-0.037		-0.019	
	(-1.33)		(-1.00)	
R_negative	0.061***	0.062***	0.043**	0.044**
	(2.82)	(2.84)	(2.23)	(2.31)
R_modalweak		-0.026*		-0.034**
		(-1.88)		(-2.31)
Download	-0.008	-0.009	0.002	0.001
	(-0.52)	(-0.59)	(0.11)	(0.05)
CAShort ×R_uncertainty	-0.019*		-0.010	
	(-1.82)		(-0.84)	
CAShortt ×R_negative	-0.025	-0.023	-0.038**	-0.037**
	(-1.26)	(-1.12)	(-2.17)	(-2.12)
CAShort ×R_modalweak		-0.003		0.003
		(-0.29)		(0.30)
CAShort ×Download	-0.028**	-0.027**	-0.037***	-0.036***
	(-2.36)	(-2.30)	(-3.13)	(-2.99)
Ln(n_words)	-0.027	-0.006	-0.009	-0.008
	(-0.96)	(-0.41)	(-0.44)	(-0.58)
Ln(n_filing)	-0.009	-0.009	-0.026	-0.026
	(-0.38)	(-0.38)	(-1.16)	(-1.16)
Ln(SZ)	0.037	0.035	0.045	0.044
	(0.25)	(0.23)	(0.41)	(0.40)
Ln(B/M)	0.169***	0.166***	0.125***	0.123***
	(4.46)	(4.45)	(3.37)	(3.36)
Pr1y	0.025*	0.025*	0.007	0.008
	(1.69)	(1.69)	(0.46)	(0.48)
Intercept	-0.041***	-0.040***	-0.025*	-0.024*
	(-2.87)	(-2.73)	(-1.84)	(-1.80)
Observations	11525	11525	11261	11261
Adjusted R^2	0.16	0.16	0.25	0.25

Electronic copy available at: https://ssrn.com/abstract=3683086

Table 9: Changes of fundamental ratios after filing dates

This table reports the results of the following fundamental change regression:

$$\Delta Fundamental_{i,t} = \alpha + \beta_0 CAShort_{i,t} + \sum_{j=1}^{3} \beta_j Textual_{i,j,t} + \beta_4 Download_{i,t}$$
$$+\beta_5 Ln(n\_words)_t + \beta_6 Ln(n\_filing)_{i,t} + \sum_{j=1}^{3} \theta_j CAShort \times Textual_{i,j,t}$$
$$+\theta_4 CAShort \times Download_{i,j,t} + control_{i,t} + d_{m,t} + \varepsilon_{i,t},$$

where $\Delta Fundamental$ is the change in firm fundamental ratios from fiscal year *t* to *t*+1. $\Delta Fundamental$ is ΔROA, $\Delta AssetTurn$, or ΔOPM. $Textual$ includes $R\_uncertainty$, $R\_modalweak$, and $R\_negative$. All variables are defined in the Appendix and are standardized to have the mean of 0 and the standard deviation of 1. We obtain the initial 10-Ks textual data from the website of The Notre Dame Software Repository for Accounting and Finance, daily shorting volume data from the FINRA website, the number of 10-Ks requests by hedge funds from the EDGAR system, and stock price, returns, shares outstanding, trading volume, and accounting data from the CRSP/Compustat. All models include stock fixed-effects and year-month dummy variables. Standard errors are double clustered at the firm and year-month levels. *, **, *** indicate significance at the 10%, 5%, and 1% levels, respectively. The sample period is from September 2009 to December 2015.

	(1) ΔROA	(2) ΔROA	(3) $\Delta AssetTurn$	(4) $\Delta AssetTurn$	(5) ΔOPM	(6) ΔOPM
CAShort	0.024	0.024	0.001	0.001	0.017	0.017
	(1.16)	(1.18)	(0.08)	(0.05)	(1.09)	(1.12)
R_uncertainty	0.021		0.004		-0.018	
	(0.85)		(0.16)		(-1.11)	
R_negative	0.047***	0.049***	0.097***	0.099***	-0.003	-0.003
	(2.65)	(2.78)	(4.34)	(4.35)	(-0.19)	(-0.21)
R_modalweak		-0.002		-0.014		-0.006
		(-0.16)		(-0.89)		(-0.67)
Download	-0.044***	-0.043***	-0.043***	-0.043***	-0.012*	-0.013*
	(-2.95)	(-2.89)	(-2.97)	(-2.96)	(-1.74)	(-1.76)
CAShort ×R_uncertainty	-0.031*		-0.001		-0.011	
	(-1.71)		(-0.09)		(-1.60)	
CAShort ×R_negative	-0.016	-0.013	-0.015	-0.014	-0.034*	-0.034
	(-1.04)	(-0.84)	(-0.91)	(-0.85)	(-1.68)	(-1.64)
CAShort ×R_modalweak		-0.021		-0.013		0.001
		(-1.21)		(-0.95)		(0.14)
CAShort ×Download	-0.029*	-0.027	-0.001	-0.002	-0.010*	-0.009*
	(-1.69)	(-1.60)	(-0.14)	(-0.26)	(-1.89)	(-1.68)
Ln(n_words)	-0.005	-0.025	0.015	0.006	-0.009	0.004
	(-0.21)	(-1.42)	(0.53)	(0.29)	(-0.64)	(0.42)
Ln(n_filing)	-0.032*	-0.031*	-0.045	-0.045	-0.000	-0.000
	(-1.99)	(-1.99)	(-1.62)	(-1.61)	(-0.03)	(-0.04)
Ln(SZ)	-0.314**	-0.311**	-0.157**	-0.153**	-0.049	-0.050
	(-2.48)	(-2.45)	(-2.26)	(-2.19)	(-0.70)	(-0.72)
Ln(B/M)	0.071	0.071	0.364***	0.365***	0.026	0.024
	(1.54)	(1.53)	(11.60)	(11.68)	(1.33)	(1.28)
Pr1y	0.052**	0.052**	-0.015	-0.015	-0.000	0.000
	(2.53)	(2.56)	(-0.89)	(-0.92)	(-0.00)	(0.05)
Intercept	0.015*	0.011	0.003	0.002	-0.002	0.000
	(1.83)	(1.46)	(0.50)	(0.34)	(-0.31)	(0.05)
Observations	15,670	15,670	15,450	15,450	15,519	15,519
Adjusted R^2	-0.02	-0.02	0.05	0.05	0.24	0.24

Electronic copy available at: https://ssrn.com/abstract=3683086

Table 10: Predictability of crash risk

This table reports the results from the following regression on firms' crash risk in fiscal year *t*+1:

$$CrashRisk_{i,t} = \alpha + \beta_0 CAShort_{i,t} + \sum_{j=1}^{3} \beta_j Textual_{i,j,t} + \beta_4 Download_{i,t}$$
$$+\beta_5 Ln(n\_words)_t + \beta_6 Ln(n\_filing)_{i,t} + \sum_{j=1}^{3} \theta_j CAShort \times Textual_{i,j,t}$$
$$+\theta_4 CAShort \times Download_{i,j,t} + control_{i,t} + d_{m,t} + \varepsilon_{i,t},$$

where $CrashRisk$ is a firm's crash risk in fiscal year t+1. Our $CrashRisk$ measures include $NSkew$, $DUVolR$, and n_Crash. *Textual* includes *R_uncertainty*, *R_modalweak*, and *R_negative*. All variables are defined in the Appendix and are standardized to have the mean of 0 and the standard deviation of 1. We obtain the initial 10-Ks textual data from the website of The Notre Dame Software Repository for Accounting and Finance, daily shorting volume data from the FINRA website, the number of 10-Ks requests by hedge funds from the EDGAR system, and stock price, returns, shares outstanding, trading volume, and accounting data from the CRSP/Compustat. All models include stock fixed-effects and year-month dummy variables. Standard errors are double clustered at the firm and year-month levels. *, **, *** indicate significance at the 10%, 5%, and 1% levels, respectively. The sample period is from September 2009 to December 2015.

	(1) $DUVolR$	(2) $DUVolR$	(3) $NSkew$	(4) $NSkew$	(5) n_Crash	(6) n_Crash
CAShort	0.009	0.010	0.008	0.008	-0.003	-0.003
	(0.81)	(0.84)	(0.78)	(0.78)	(-0.32)	(-0.28)
R_uncertainty	-0.048*		-0.046*		-0.039	
	(-1.92)		(-1.78)		(-1.51)	
R_negative	0.016	0.014	0.018	0.016	-0.001	-0.002
	(0.79)	(0.70)	(0.98)	(0.91)	(-0.06)	(-0.10)
R_modalweak		-0.008		-0.011		-0.017
		(-0.46)		(-0.55)		(-0.93)
Download	-0.010	-0.011	-0.012	-0.013	-0.014	-0.014
	(-0.74)	(-0.80)	(-0.89)	(-0.95)	(-1.15)	(-1.22)
CAShort ×R_uncertainty	0.021**		0.018*		0.009	
	(2.00)		(1.73)		(0.72)	
CAShort ×R_negative	-0.002	-0.003	0.000	-0.001	0.004	0.003
	(-0.18)	(-0.36)	(0.04)	(-0.11)	(0.42)	(0.33)
CAShort ×R_modalweak		0.016		0.011		0.012
		(1.56)		(1.15)		(1.10)
CAShort ×Download	-0.009	-0.010	-0.009	-0.010	-0.005	-0.005
	(-0.80)	(-0.90)	(-0.87)	(-0.98)	(-0.43)	(-0.45)
Ln(n_words)	-0.033	0.005	-0.031	0.004	-0.019	0.007
	(-1.04)	(0.20)	(-1.01)	(0.15)	(-0.65)	(0.30)
Ln(n_filing)	-0.009	-0.010	-0.009	-0.009	-0.007	-0.007
	(-0.42)	(-0.43)	(-0.40)	(-0.40)	(-0.31)	(-0.31)
Ln(SZ)	0.970***	0.966***	0.855***	0.851***	0.585***	0.583***
	(9.59)	(9.48)	(9.04)	(8.94)	(8.00)	(7.87)
Ln(B/M)	0.013	0.012	0.016	0.015	0.010	0.008
	(0.48)	(0.41)	(0.61)	(0.54)	(0.42)	(0.34)
Pr1y	-0.016**	-0.016**	-0.017***	-0.017***	-0.008	-0.008
	(-2.37)	(-2.31)	(-2.71)	(-2.65)	(-1.12)	(-1.07)
Intercept	-0.041***	-0.035***	-0.032***	-0.027***	-0.016***	-0.012**
	(-5.58)	(-4.78)	(-4.44)	(-3.85)	(-3.15)	(-2.34)
Observations	15,311	15,311	15,311	15,311	15,310	15,310
Adjusted R^2	0.06	0.06	0.06	0.05	0.03	0.03

Electronic copy available at: https://ssrn.com/abstract=3683086