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 19, 2021

Abstract

We examine short sellers' use of textual information in annual reports for shorting activities. We find that more uncertainty words in annual reports are associated with greater shorting volume. Short selling motivated by textual information negatively predicts stock price reaction after 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 An-Sing Chen, Monica Chen, Keith Lam, Xiaorong Ma, Xiumin Martin, Susan Ren, Lewis Tam, Vijay Singal, Feng Tang, Jason Xiao, Endong Yang, and participants of seminar at University of Macau and the 28th SFM conference in Kaohsiung 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.

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 words in annual reports are associated with greater shorting volume. Short selling motivated by textual information negatively predicts stock price reaction after 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

1. Introduction

Information discovery and transmission in the financial market is at the core of finance and accounting research. Theories 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, theories predict that better information disclosures mitigate informational asymmetry between insiders and outsiders, enhance stocks' liquidity, and reduce firms' cost of capital (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 can improve aggregate shareholder welfare by reducing search costs (e.g., Brown and Hillegeist, 2007). As a result, 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; Reed et al., 2020; Gargano et al., 2021). Researchers are interested in knowing the sources of short sellers' trading profitability and what kind of information motivates short selling activity.[1] Some studies focus on private information while others examine public information such as 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

[1] Other studies use the pilot program of Regulation SHO to examine the issues, from the *ex ante* perspective, related to (i) information and trading profit competition between short sellers and insiders (Massa, Qian, Xu, and Zhang, 2015) and (ii) the disciplinary role of short sellers under the threat of potential short selling (Wang, Wang, Wei, Zhang, and Zhou, 2021).

sellers interpret public information is understudied in the literature.

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. Annual reports contain both hard and soft information, while soft information dominate in terms of length. We are particularly interested in the writing style 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 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, which counts 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 summarizes the most relevant information about the firm over the past fiscal year.[2] Psychological 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

[2] Compare with other textual information, annual reports also have limitations. Blau et al. (2015) claim that formal documents and filings like 10Ks are static while the conference call information environment is dynamic. Outsiders have no opportunity to challenge the content in 10Ks and the conference call is an exchange of dialogue. Therefore, conference calls provide more additional pertinent information about the firm (Blau et al., 2015). In addition, conference calls provide more updated outlook of the firm performance compared to annual reports, although the information contained in annual reports is richer and more structured than information in conference calls.

capacities force investors to focus on a subset of available information (Hirshleifer, 2001; Hirshleifer and Teoh, 2003). On average the number of words in annual reports is more than 50,000, which is long enough for short sellers to reveal their preferences on the type of information in annual reports.[3] Short sellers might directly search for negative information that has not been incorporated into the stock price yet, or they may assess whether the stock is overvalued relative to fundamentals.[4]

Using textual data from annual reports during 2009 to 2015 and daily shorting volume data from NYSE/Amex/Nasdaq, we find that textual variables are correlated with the shorting volume on the 10-Ks filing dates. Short sellers are more willing to short the stock if the firm's annual report is lengthy and contains more uncertainty 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 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 shorting volume and textual information can predict future stock returns. The literature documents that short sellers are skilled information processors.[5] If short sellers are able to extract negative information from ambiguous writings, we should observe a strong negative relation between shorting volume and abnormal stock returns in a short period

[3] 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. (2020) provide evidence on hedge funds, one group of short sellers, that download annual reports via EDGAR.

[4] 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.

[5] 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).

after firms filed 10-Ks. However, short sellers may have behavioral bias just as other investors.[6] For example, Ditto and Lopez (1992) find that the information consistent with a preferred conclusion is examined less critically than information inconsistent with a preferred conclusion, and consequently, less information is required to reach the former case than the latter case.[7] 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, shorting volume may not lead to trading profits.

We summarize the empirical results on stock return predictability as follows. First, textual variables predict abnormal returns in the 3-day event window [1, 3] after the filing of 10-Ks (i.e., day 0). Specifically, more uncertainty words and more modal weak words predict lower abnormal returns, whereas more negative words predict higher abnormal returns. Second, shorting volume on the filing dates is unrelated to [1, 3] abnormal returns; however, short volume predicts abnormal returns when short sellers use textual information of negative words. Third, textual information used by short sellers also predict 1-week to 52-week ahead abnormal returns. Using fitted shorting volume (i.e., shorting volume driven by textual information), we find a significantly negative relation between shorting volume and abnormal stock returns from [0, 3] days to 4-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.

---

[6] 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 disclosures about a firm, different ways of presentations by even experienced financial analysts can affect the valuations and trades of investors (Hirshleifer and Teoh, 2003).

[7] 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 investment, but disagree with information that suggests they might lose money.

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. Our results show that the revisions of analysts' earnings forecasts around filing months are related to the ratio of negative and modal weak words. Short sellers do influence financial analysts after 10-Ks filing dates by incorporating the amount of uncertainty words in the short selling activity. Shorting volume, negative words, and modal weak words also predict the revision of earnings forecasts for the next fiscal year. We also find that the ratio of negative words is positively related to changes in firm fundamentals as measured by return-on-assets (ROA) and asset turnover in the next fiscal year. Shorting volume combined with textual information also plays some role in predicting future changes in firm fundamentals.

Finally, motivated by Hutton et al. (2009), Callen and Fang (2015), Kim, Wang, and Zhang (2019), and Deng et al. (2020), we investigate whether 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 probability. However, we also find some weak evidence that the interaction of 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 several ways. First, we show how short sellers process information, by focusing on the textual information in annual reports. As far as we know, our study is the first to examine the relation between annual reports and short selling activities. The way for short sellers to make profits is through future stock price depreciation, which could come from current stock prices not reflecting the true prospects of the firms.[8] Studies find that short sellers utilize fundamental analysis when targeting overvalued companies[9] and they are skilled in analyzing public information and taking advantage of noise traders.[10] Another branch of studies focuses on how short sellers use private information to form their trading strategies.[11] We complement Engelberg et al. (2012) and von Beschwitz et al. (2017) by further studying short sellers' use of textual information in annual reports.

Our study is also related to investor's information acquisition, especially information acquisition from EDGAR. Drake et al. (2015) analyze the determinants of users' access to SEC filings through EDGAR, and find that EDGAR search activity is positively related to firm events and information environments. Drake et al. (2017) conduct a demographic analysis of financial statements downloads via EDGAR. Recent studies also find that EDGAR searching by institutional investors including hedge funds can predict future firm stock returns and fundamentals.[12] Our paper complements those studies and finds that there is a strongly positive relation between the number of EDGAR requests and 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

---

[8] This overvaluation can also be purely from negative information that has not been incorporated into share prices due to short-sale constraints and differences in investor opinions (Miller, 1977; Chen et al., 2002; Nagel, 2005; Boehme et al., 2006; Berkman et al., 2009).

[9] Dechow (2001), Curtis and Fargher (2014), Deshmukh et al. (2015), and Drake et al. (2015).

[10] Desai et al. (2006), Karpoff and Lou (2010), Drake et al. (2011), Fang et al. (2016), Engelberg et al. (2012), von Beschwitz et al. (2017), and Reed et al. (2020).

[11] Christophe et al. (2004), Khan and Lu (2013), Shi et al. (2017), Berkman et al. (2017), Berkman and Eugster (2017), Purnanandam and Seyhun (2018), and Choi et al. (2020).

[12] See, for instance, Li and Sun (2019), Chen et al. (2020), Chen, Kelly, and Wu (2020), Cohen et al. (2020), Crane et al. (2020), and Drake et al. (2020). Gibbons et al. (2021) study the financial analysts' information acquisition via EDGAR and find that analysts' attention to public information is driven by the demand, analysts' incentive and career concerns. Information acquisition via EDGAR is related to a significant reduction in analysts' forecasting error, and their recommendation updates are associated with significant abnormal returns.

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.[13] Textual analysis also helps to explain the underpricing of IPOs (Hanley and Hoberg, 2010; Jegadeesh and Wu, 2013) and are useful for predicting future market returns (Jiang et al., 2019). Drake et al. (2020) also find that the tone of 10-Ks has a negative moderating effect for the positive relation between EDGAR searching and future ownership by sophisticated investors. We add to this literature by studying how the textual information is used by short sellers in their trading.

The rest of paper is organized as follows. Section 2 describes the data and variables constructions. 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 between shorting volume, textual information, and firm fundamentals. Section 6 examines whether shorting volume and textual information are related to crash risk. Finally, 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 from the FINRA website.[14] 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

[13] See, Brown and Tucker (2011), Jegadeesh and Wu (2013), Loughran and MacDonald (2011, 2014, 2016), and Ke et al. (2019). In addition, Liberti and Petersen (2019) discusses the advantages and disadvantages of hard and soft information, with the impact from technology changes. Hardening of information, which is defined as how much valuable information is lost in the process, is a fundamental challenge (Liberti and Petersen, 2019).

[14] 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.

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 would underestimate shorting volume.[15] We include only common stocks (share codes 10 and 11) listed on NYSE/Amex and Nasdaq.

We obtained the initial textual data of 10-Ks from the website of The Notre Dame Software Repository for Accounting and Finance.[16] 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 10-Ks from EDGAR are merged and computed from three databases. First, we obtain a list of hedge funds that are similar to Jiang (2019).[17] 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 10-Ks from the Securities and Exchanges Commission's (SEC) EDGAR log file database at the daily level.[18] 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 come from Compustat. Analysts' earnings forecasts data are from the I/B/E/S database. Institutional ownership data are obtained from the Thomson Institutional

[15] 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.

[16] 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.

[17] 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).

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

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

We use shorting volume ($Short$) on the filing date of the 10-K annual reports available on SEC's EDGAR as our main measure of shorting volume. $Short$ is defined as daily shorting volume divided by total share outstanding.[19] Hedge funds are the major group of short sellers. We use hedge funds' download of 10-Ks from EDGAR as a proxy for short sellers' download of annual reports and define $Download$ as the number of hedge funds' downloads of 10-Ks 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 tone; and (vi) $n\_positive$, which is the number of words related to positive tone.

[19] We are motivated by Wang et al. (2020) who find that short-term shorting flows predict future returns, but abnormal short-term shorting flows cannot predict future returns.

We construct three ratios related to textual information. 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 10-Ks 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.

$$BHAR[j,k]_{i,} = \prod_{d=j}^{k}(1 + Ret_{i,d}) - \prod_{d=j}^{k}(1 + Mkt_d), \tag{1}$$

where $Ret_{i,d}$ is the daily stock return for stock *i* on day *d* and Mkt_d is the daily value-weighted CRSP market index return on day d. We construct $BHAR$ during the 3-day event window [1, 3] and denote it as $BHAR3d$.[20] Similar to Eq. (1), we also 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

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

$$\Delta FEPS1_{i,t} = \frac{FEPS1_{i,t+1} - FEPS1_{i,t-1}}{StockPrice_{i,t-1}}, \tag{2}$$

[20] We follow Loughran and MDonald (2011) to choose the event window [0, 3] but excluding the event day [0].

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

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 (ROA) 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 (2006) to construct three measures of 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:

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

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

$$DUVolR_{i,y} = Ln\left\{\frac{n_{up}(\sum_{w\in Down}(Ret_{i,w,y} - \overline{Ret}_{i,y})^2)}{n_{down}(\sum_{w\in Up}(Ret_{i,w,y} - \overline{Ret}_{i,y})^2)}\right\}, \tag{4}$$

where n_{down} (n_{up}) is the number of down (up) weeks. A down (up) week is defined as a week when the firm-specific weekly return is below (above) 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 negative 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:

$$n\_Crash_{i,y} = n_{negative} - n_{positive}. \tag{5}$$

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 illiquidity measure ($Illiq$); (iv) institutional holdings scaled by the number of shares outstanding ($IOwner$); (v) past 1-year cumulative 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 with the Carhart momentum factor; and (vii) firm's standardized unexplained earnings (SUE). 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 analysis uses the daily shorting volume, which is only available from September 2009, we also report the summary statistics over the period of September 2009 to December 2015. Table 2 shows that the mean $Short$ is 0.13%, which means that the daily shorting volume on the filing date is not high at the absolute level. The mean $BHAR3d$ is -0.07% during the event window [1, 3], which suggests that stocks of firms around their 10-Ks filing dates [1, 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 53,249. The average number of uncertainty words ($n\_uncertainty$) is 731. Negative words on average far outnumber positive words in the annual reports, with the averages of $n\_negative$ and $n\_positive$ being 970 and 377, 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, respectively. The annual reports also prefer to use modal weak tone, as the number of modal weak words is much more than the number of modal strong words (324 vs. 158), compared to the 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 ratios of textual variables, $R\_negative$ is slightly higher than $R\_modalweak$ (0.7106 vs 0.6774), while the average percentage of uncertainty words is 1.43%. The average number of filing firms per day is 89.4, with a median value of 64. The average $Download$ of 10-Ks per firm by hedge funds on the filing date is 20.9.[21]

Other variables describe the characteristics of the sample in our study. For example, the average firm size is $4,334 million and the average institutional ownership is 52.27%. The average previous one-year cumulative returns ($Pr1y$) is 25.94%. This is also consistent with the fact that overall stock market performed well during our sample period. For example, the S&P 500 index increases from 998 on September 1, 2009, to 2,043 on December 31, 2015. This also means that only very skilled short sellers can make profit in a bull market.

Table 3 reports the pairwise correlation among main variables. Shorting volume ($Short$) on the filing date is negatively correlated with abnormal stock returns in next three days (BHAR3d), indicating that short selling on the filing date is informative and can predict future stock returns. *Short* is positively correlated with the length of annul reports ($Ln(n\_words)$) and the number of 10-K 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 correlated with illiquidity ($Illiq$) and idiosyncratic volatility ($IVol$).

[Insert Table 3 here]

3. Daily shorting volume and textual information

To investigate short sellers' reaction to the filing of 10-K reports, we run the following pooled OLS regression with shorting volume on the filing date ($Short$) as the dependent variable:

[21] Laughran and McDonald (2017) find that the average download of a firm's annual report is 28.4 times, after excluding robot requests.

$$Short_{i,t} = \alpha + \sum_{j=1}^{3} \beta_j Textual_{i,j,t} + \beta_4 Download_{i,t} \; + \gamma_1 Ln(n\_words)_{i,t}$$
$$+\gamma_2 Ln(n\_filing)_t + \sum_{j=1}^{3} \theta_j Textual_{i,j,t} \times Download_{i,t} \qquad (6)$$
$$+control_{i,t} + d_{m,t} + \varepsilon_{i,t}.$$

where $Textual$ includes $R\_uncertainty$, $R\_modalweak$, and $R\_negative$. To absorb the time-invariant stock-specific effects and aggregate time trends, we include stock fixed-effects (f_i) and year-month dummy variables $\left(d_{m,t}\right)$ in the regression model. We consider six different model specifications. 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 report the results without the interaction terms between $Textual$ and $Download$, and Models 5-6 with the interaction terms. Among all the textual ratios analyzed, the coefficients on $R\_uncertainty$ are all significantly positive, ranging from 0.057 to 0.067, and t-statistics show significant at the 1% level. The coefficient estimate is also economically significant. Take the estimate of Model 1 as an example, a one-standard-deviation increase in $R\_uncertainty$ leads to a 11.3% (= 0.067×0.0022/0.0013) standard-deviation increase in $Short$. The results suggest that short sellers are more likely to take a short position if they discover negative information from uncertainty words in the annual reports. This result is consistent with the prediction of Miller (1977), who predict that uncertainty-induced investor disagreement can lead to overvaluation. However, the coefficients on $R\_negative$ and $R\_modalweak$ are all insignificant. The coefficients on the interaction terms are also insignificant across all model specifications.

[Insert Table 4 here]

For the variable used to measure the length of annual reports, $Ln(n\_words)$ is positively related to shorting volume in all model specifications. For example, the coefficient on $Ln(n\_words)$ is 0.085 with a t-statistic of 4.28 in Model 1, suggesting that short sellers take more aggressive short positions on firms with lengthier annual reports. Economically, a one-standard-deviation increase in $Ln(n\_words)$ leads to a 14.3% (= 0.085×0.0022/0.0013) standard-deviation increase in shorting volume. The result suggests that short sellers find that lengthier annual reports contain more negative information probably because firms try to use more spaces to explain adverse information.[22] The results are consistent with von Beschwitz et al. (2017) who find that short sellers trade more on the days with qualitative news.

However, Model 1 shows that short sellers short less on a particular firm if there are more firms filing 10-K reports on the same filing date. The coefficient on $Ln(n\_filing)$ is -0.052 (t-stat = -2.79), suggesting the possibility of short sellers' limited attention. 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 second strongest effect on shorting volume, behind $IVol$. For example, the coefficient on $Download$ in Model 2 is 0.126 (t-stat = 5.39), indicating that a one-standard-deviation increase in $Download$ is associated with a 21.3% (= 0.126×0.0022/0.0013) standard deviation increase in shorting volume. The result provides direct evidence on the positive relation between short sellers' use of annual reports and their shorting activity. Specifically, short sellers take more short positions at the same time when they download annual reports from EDGAR more intensively, suggesting that annual reports contain useful information to them.

[22] 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.

In Models 5-6 of Table 4, we interact textual variables with $Download$ to investigate which types of textual information are favored by short sellers. However, the results show that the coefficients on interaction terms are all insignificant. Finally, we find that several firm-level control variables are significantly related to shorting volume around the filing days. For example, $Short$ is positively associated with past one-year stock returns ($Pr1y$) and idiosyncratic volatility ($IVol$), but negatively correlated with stock illiquidity ($Illiq$). These results indicate that short sellers prefer to short stocks of firms with better liquidity (lower transaction costs), higher past one-year performance (contrarian strategies), and a greater difference in investors' opinion. In general, those findings are consistent with prior studies on short selling strategies (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 [1, 3]

We next investigate the return predictability of textual variables and the role played by short sellers in driving the predictability. If shorting volume contains negative information conditional on textual variables, we expect the coefficient on the interaction terms between shorting volume and textual variables to be significant. We use the following pooled OLS regression to test our hypothesis:

$$\begin{aligned} BHAR3d_{i,t} = \alpha + \beta_0 Short_{i,t} + \textstyle\sum_{j=1}^{3} \beta_j Textual_{i,j,t} + \beta_4 Download_{i,t} \\ + \beta_5 Ln(n\_words)_{i,t} + \beta_6 Ln(n\_filing)_t + \textstyle\sum_{j=1}^{3} \theta_j Texxtual_{i,j,t} \times Short_{i,t} \\ + \theta_4 Short_{i,t} \times Download_{i,j,t} + control_{i,t} + d_{m,t} + \varepsilon_{i,t}. \end{aligned} \tag{7}$$

where all variables are defined previously. Table 5 reports the results. We find that all three textual variables are significantly or marginally significantly associated with future abnormal returns. First, higher $R\_uncertainty$ predicts lower abnormal returns, as indicated by the negative coefficients on $R\_uncertainty$ in models 2 and 3. For example, the coefficient on $R\_uncertainty$ is -0.043 (t-stat = -2.00) in Model 2, suggesting that investors view uncertainty words in 10-Ks as negative information, consistent with short sellers taking more short positions as shown in Table 4.

[Table 5 here]

Second, higher $R\_negative$ predicts higher abnormal returns. For example, the coefficient on $R\_negative$ is 0.037 (t-stat = 2.01) in Model 3. This result is a bit puzzling and counter-intuitive as negative words should represent negative information. One potential explanation is that managers might try their best to explain the past negative news that had impacted their stock price in detail so that investors might interpret that all the bad news has been out. Another explanation is that $R\_negative$ is highly correlated with other firm characteristics that predict positive returns. We find that $R\_negative$ is highly correlated with book-to-market (B/M), but not with the previous 1-year return ($Pr1y$) or the change 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. It suggests that short sellers are less likely to target value stocks with more negative words, probably because the price of value stocks have already overreacted to the negative prospects.

Third, higher $R\_modalweak$ predicts lower abnormal returns in next three days. For example, the coefficient on $R\_modalweak$ is -0.038 (t-stat = -2.75) in Model 4, indicating that a one-standard-deviation increase in $R\_modalweak$ predicts a decrease in $BHAR3d$ by -0.19% (-0.038×0.0522). The results indicate that investors view modal weak words in 10-Ks as negative information. Taken together, our results suggest that investors react to the textual information in 10-Ks as indicated by the significant coefficients on three textual variables.

Fourth, $Ln(n\_words)$ is significantly and negatively related to $BHAR3d$ in all models except Model 1. For example, the coefficient on $Ln(n\_words)$ in Model 2 is -0.071 (t-stat = -2.29), indicating that investors view lengthier annual reports more negatively, potentially because managers try to use more spaces to explain potential negative information to investors.

Fifth, 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, even after controlling for $Download$ in Models 5-6. For example, the coefficient on $Ln(n\_filing)$ is -0.067 (t-stat = -2.49) in Model 1, which is consistent with the finding by Hirshleifer, Lim, and Teoh (2009) who show that market underreaction is more severe when more firms announce earnings on the same date because investors are distracted if there are many 10-K filings in a day. It could be explained by studies showing that managers of firms with bad news tend to time their earnings announcements and 10-K filing days to divert investors' attention to mitigate a large negative stock price reactions on the announcement or filing date and litigation risk (Bowen et al., 1992; Donelson et al., 2012).

Finally, $Download$ has a negative relation with $BHAR3d$, which is consistent with the result in Table 4 that $Download$ is positively associated with shorting volume. The marginally significant coefficient on *Download* also reveals that the number of hedge funds on the filing date is more related to short selling on the same day, but less related to future stock returns.

Next, the coefficient on $Short$ is insignificant in all models, suggesting that short selling activity itself is unable to predict future abnormal returns in a short time period on average. The results from the interactions of $Short$ with textual variables indicate that short selling is more informative about stock returns when the annual reports contain more negative words. For example, the coefficients on $Short \times R\_negative$ are significantly positive in all models (for example, coeff = 0.046 with *t*-stat = 2.80 in Model 3). In contrast, the return predictability of shorting volume is not related to more uncertainty or modal weak words in the 10-K reports.

To further understand the informativeness of textual information used by short sellers, we include the filing date return to calculate the abnormal returns ($BHAR4d$) in the [0, 3] event period. We find that short sellers who use textual information to form trading strategies are more informative. Appendix Table 1 shows that the coefficients on $Short \times R\_uncertainty$, $Short \times R\_negative$, and $Short \times R\_modalweak$ are all statistically significant. For example, the coefficient on $Short \times R\_uncertainty$ is -0.060 (*t*-stat = -2.22) and that on $Short \times R\_negative$ is 0.047 (*t*-stat = 2.00) in Model 3, and the coefficient on $Short \times R\_modalweak$ is -0.080 (*t*-stat = -2.32) in Model 4. In addition, the coefficient on *Short* is significantly positive in all models. It means that controlling for other variables, average investors short on the 10-Ks filing dates will lose money. However, short on stocks with more uncertainty and modal weak words, and less negative words are profitable. The results further suggest that 10-Ks contain rich information that is used by short sellers to form their trading strategies.

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

Table 5 shows a significant relation between shorting volume, textual variables, and future short-term abnormal returns around 10-K filing dates. A natural question is whether abnormal shorting volume predicts stock returns in the longer run. We replace $BHAR3d$ with $BHAR1w$ to $BHAR52w$ and re-run Eq. (7). Table 6 reports the result. We find that the relation between shorting volume and future abnormal stock returns remains insignificant from 1-week to 12-week ahead, but highly negative and significant for 24- and 52-week ahead. For example, the coefficient on $Short$ is -0.032 (*t*-stat = -2.50) in Model 9 with $BHAR24w$ and is -0.074 (*t*-stat = -5.30) in Model

11 with $BHAR52w$. The evidence supports the view that short selling is informative and can predict future long-term stock returns, which is largely consistent with prior studies (Desai et al., 2002; Arnold et al., 2005; Boehme et al., 2008; Reed et al., 2020; Gargano et al., 2021).

[Table 6 here]

We also find that the coefficient on $R\_uncertainty$ is (marginally) significantly negative for $BHAR12w$ and $BHAR24w$, the coefficient on $R\_modalweak$ is significantly negative for $BHAR1w$ and $BHAR2w$, and the coefficient on $R\_negative$ is significantly negative for $BHAR1w$ and marginal negatively significant for $BHAR2w$ and $BAHR52w$. In addition, the coefficients on some interaction terms are also significant. For example, the coefficient on $Short \times R\_uncertainty$ is significantly negative for all future returns except for $BHA1w$. The coefficient on $Short \times R\_negative$ is significantly positive for $BHAR1w$, $BHAR2w$, and $BHAR4w$. The coefficient on $Short \times R\_modalweak$ is significantly negative for $BHAR4w$ and $BAHR52w$ and marginal significantly negative for $BHAR2w$. The results suggest that short sellers are informative in incorporating textual information, and that stock prices during the event window [1, 3] do not fully reflect such information, and therefore there is a further return drift after the textual information becomes publicly available.

Meanwhile, the relation between $Download$ and future stock returns is negatively significant for $BHAR24w$ and $BHAR52w$, with the coefficients ranging from -0.035 (*t*-stat = -2.45) in Model 9 to -0.046 (*t*-stat = -3.00) in Model 11. The results suggest that the number of hedge fund downloads can predict long-term abnormal stock returns, so can shorting volume (*Short*). However, the interaction term $Short \times Downlaod$ is insignificant in all models.

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 derived from 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 explained by textual information in the annual reports. We first run the following regression to obtain the fitted component of $Short_{i,t}$ ($\widehat{Short}_{i,t}$):

$$Short_{i,t} = \alpha + \sum_{j=1}^{3} \beta_j Textual_{i,j,t} + \varepsilon_{i,t}. \tag{8}$$

We then run the following regression,

$$\begin{aligned} BHAR_{i,t} = \alpha + \beta_1 \widehat{Short}_{i,t} + \beta_2 \,\mathrm{Ln}(n\_filing)_t + \beta_3 Download_{i,t} \\ +\beta_4 Ln(n\_words)_{i,t} + control_{i,t} + d_{m,t} + \varepsilon_{i,t}. \end{aligned} \tag{9}$$

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

Table 7 Panel A reports the results. The coefficients on $\widehat{Short}$ are indeed significantly negative when the dependent variable is $BHAR3d$, $BHAR1w$, $BHAR2w$ and marginal significantly negative for $BHAR4w$. The corresponding coefficients are -0.041 (*t*-stat = -2.40), -0.035 (*t*-stat = -2.40), -0.031 (*t*-stat = -2.10), and -0.026 (*t*-stat = -1.96), respectively. These results show that short sellers are indeed informative in predicting future poor stock performance up to four weeks after filing 10-Ks, when their aggressive short selling volume is driven by textual information.

[Table 7 here]

Panel B of Table 7 reports the results from the residual component (Res_short) from Eq. (8). The results show that the coefficient on Res_short is insignificant for $BHAR$s up to 12 weeks. However, the coefficient on Res_short is statistically negative for $BHAR24w$ and $BHAR52w$.

The results suggestion once we decompose $Short$ information into textual related ($\widehat{Short}$) and non-textual related (Res_short), the textual-related component predicts the short-horizon stock returns while the non-textual component predicts the long-horizon stock returns. The results provide further evidence for the predictability of $Short$ on $BHAR24w$ and $BHAR52w$ reported in Table 6 are mainly from non-textual related information.

5. Textual information and firm fundamentals

In Sections 3 and 4, we find significant relation between 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 shorting with revisions of analysts' earnings forecasts and changes in firm fundamentals.

5.1 Analysts' earnings forecast revisions

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

$$\begin{aligned}\Delta FEPS1_{i,t}\left(\Delta FEPS2_{i,t}\right) = \alpha &+ \beta_0 Short_{i,t} + \textstyle\sum_{j=1}^{3}\beta_j Textual_{i,j,t} + \beta_4 Download_{i,t}\\ &+\beta_5 Ln(n\_words)_{i,t} + \beta_6 Ln(n\_filing)_t + \textstyle\sum_{j=1}^{3}\theta_j Texxtual_{i,j,t} \times Short_{i,t} \\ &+\theta_4 Short_{i,t} \times Download_{i,j,t} + control_{i,t} + d_{m,t} + \varepsilon_{i,t}.\end{aligned} \quad (10)$$

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 results collaborate well with the short-term return predictability of $R\_negative$ and $R\_modalweak$ reported in Tables 5 and 6. The coefficients on $R\_negative$ are 0.060 (t-stat = 2.76) in Model 2 for $\Delta FEPS1$ and 0.040 (t-stat = 2.01) in Model 4 for $\Delta FEPS2$. The corresponding coefficients on $R\_modalweak$ are -0.030 (t-stat = -1.95) and -0.035 (t-stat = -2.25). The results suggest that more negative words (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 coefficients on $Short \times R\_uncertainty$ are significantly negative in Model 1 for $\Delta FEPS1$ (coeff = -0.037; t-stat = -2.20) and Model 3 for $\Delta FEPS2$ (coeff = -0.031; t-stat = -2.29). The results indicate that analysts' forecast revisions are negatively related to short selling activity with more uncertainty words. The coefficients on $Short$ are also significantly negative for $\Delta FEPS2$ as shown in Model 3 (coeff = -0.038; t-stat = -2.33) and Model 4 (coeff = -0.039; t-stat = -2.40) but insignificant for $\Delta FEPS1$. Combining with the significant coefficients on $R\_negative$, $R\_modalweak$, and $Short \times R\_uncertainty$, the results suggest that the revisions of analysts' forecasts for fiscal year 2 is directly captured by the short selling activity with uncertainty words and other textual variables in the annual reports.

[Table 8 here]

5.2 Changes in firm fundamentals

Prior studies have found that short sellers' shorting decisions are based 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. (10). 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 ($\Delta AssetTurn$). For example, the coefficient on ΔROA is 0.053 (*t*-stat = 2.64) in Models 1 and 0.094 (*t*-stat = 4.31) on $\Delta AssetTurn$ in Model 3. It suggests that more negative words in annual reports actually indicate that firm fundamentals will improve over the next fiscal year. The results are consistent with those on the short-term return predictability and analysts' earnings revisions of $R\_negative$ reported in Tables 5-7. The coefficients on $Download$ are significantly negative for ΔROA and $\Delta AssetTurn$, suggesting that firm 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. Finally, the coefficients on $Short \times R\_uncertainty$ and $Short \times R\_modalweak$ are significantly negative in some models. For example, the coefficient on $Short \times R\_uncertainty$ is -0.025 (*t*-stat = -2.53) for ΔOPM in Model 5. The result suggests that the interaction of shorting volume with textual variables also have some predictive power and informative content on the changes of firm fundamentals.

[Table 9 here]

6. Crash risk

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 crashes. 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. (10). Our $CrashRisk$ measures include $NSKEW$, $DUVolR$, and n_Crash.

Table 10 presents the evidence on the predictability of crash risk using shorting volume and textual information. Model 1 ($DUVolR$) shows that the coefficient on $R\_uncertainty$ is -0.054 (*t*-stat = -2.15) and on $Short \times R\_uncertainty$ is 0.020 (*t*-stat = 1.84), whereas Model 3 ($NSkew$) reveals that the corresponding coefficients are -0.051 (*t*-stat = -1.78) and 0.018 (*t*-stat = 1.65), respectively. Interestingly, $R\_uncertainty$ is 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). In contrast, the coefficients on $Short \times R\_uncertainty$ are all positive and it is significant in Model 1, suggesting that short sellers can potentially identify firms with increasing crash risk through focusing on the shorting activities with those firms with a higher frequency of uncertainty words in their 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 words in annual reports are associated with greater 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. Overall, our results suggest that textual information in annual reports forms an important part of short sellers' information advantage.

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.

Bowen, Robert M., Marilyn F. Johnson, Terry Shevlin, and D. Shores, 1992, Determinants of the timing of quarterly earnings announcements, Journal of Accounting, Auditing and Finance 7(4), 395-422.

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.

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, Huaizhi, Lauren Cohen, Umit Gurun, Dong Lou, and Christopher Malloy, 2020, IQ from IP: Simplifying search in portfolio choice, Journal of Financial Economics 138, 118–137.

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

Chen, Yong, Bryan Kelly, and Wei Wu, 2020, Sophisticated investors and market efficiency: Evidence from a natural experiment, Journal of Financial Economics 138, 316–341.

Cheung, Adrian W. K., Hung Wan Kot, Eric F. Y. Lam, and Harry K. M. Leung, 2020, Toward understanding short-selling activity: Demand and supply, Accounting and Finance 60(3), 2203-2230.

Choi, Jaewon, Ji Min Park, Neil D. Pearson, and Shastri Sandy, 2020, Profitability of hedge fund short sales: Evidence from opening and closing transactions, 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.

Cohen, Lauren, Christopher Malloy, and Quoc Nguyen, 2020, Lazy prices, Journal of Finance 75(3), 1371–1415.

Crane, Alan, Kevin Crotty, and Tarik Umar, 2020, The role of public and private information in fund performance complements or substitutes? 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 60, 101498.

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.

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.

Donelson, Dain C., John M. McInnis, Richard D. Mergenthaler, and Yong Yu, 2012, The timeliness of bad earnings news and litigation risk, The Accounting Review, 87(6), 1967-1991.

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, The Accounting Review 31 (3), 55-68.

Drake, Michael S., Bret A. Johnson, Darren T. Roulstone, and Jacob R. Thornock, 2020, Is there information content in information acquisition? The Accounting Review 95(2), 113–139.

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, 2021, Short of capital: Stock market implications of short sellers' losses, Working paper, University of Houston.

Gibbons, Brian, Peter Lliev, Jonathan Kalodimos, 2021, Analyst information acquisition via EDGAR, Management Science 67(2), 769–793.

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.

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, 2020, Institutional trading around corporate news: Evidence from textual analysis, Review of Financial Studies 33(10), 4627-4675.

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, The 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 132(1), 126-149.

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? The 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.

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, José María, and Mitchell A Petersen, 2019, Information: Hard and soft, Review of Corporate Finance Studies 8(1), 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.

Massa, Massimo, Wenlan Qian, Weibiao Xu, and Hong Zhang, 2015, Competition of the informed: Does the presence of short sellers affect insider selling? Journal of Financial Economics 118, 268-288.

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.

Reed, Adam V., Mehrdad Samadi, and Jonathan S. Sokobin, 2020, Shorting in broad daylight: Short sales and venue choice, Journal of Financial and Quantitative Analysis 55(7), 2246–2269.

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, Rencheng, Kemin Wang, K.C. John Wei, Bohui Zhang, and Yi Zhou, 2021, Insider sales under the threat of short sellers: New theory and new tests, The Accounting Review, Forthcoming.

Wang, Xue, Xuemin (Sterling) Yan, and Lingling Zheng, 2020, Shorting flows, public disclosure, and market efficiency, Journal of Financial Economics 135, 191-212.

Yan, Yumeng, 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 57, 101075.

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	

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, earnings from IBES, and stock price, returns, shares outstanding, trading volume, and accounting data from the CRSP/Compustat.

	N	Mean	Std. Dev.	p5	Median	p95
Short	19,549	0.0013	0.0022	0	0.0006	0.0058
BHAR3d	19,540	-0.0007	0.0522	-0.0779	-0.0013	0.0739
n words	19,641	53,249	32,981	23,027	45,836	108,301
n uncertainty	19,641	731	359	307	673	1322
n modalweak	19,641	324	186	104	287	670
n modalstrong	19,641	158	131	49	128	354
n negative	19,641	970	658	320	823	2,086
n positive	19,641	377	229	140	328	780
n filing	19,641	89.4	79.7	3	64	232
R_uncertainty	19,641	0.0143	0.0028	0.0096	0.0143	0.0187
R_modalweak	19,641	0.6774	0.0769	0.5406	0.6862	0.7879
R_negative	19,641	0.7106	0.0614	0.6025	0.7169	0.7994
Download	18,179	20.9	17.6	2	16	57
Firm Size (SZ)	19,626	4334	17629	21	517	17242
B/M	17,901	0.8551	0.8912	0.1186	0.6405	2.3052
IOwner	18,615	0.5227	0.327	0.0021	0.588	0.9495
Illiq	19,639	0.9595	3.5425	0.0001	0.0053	5.69
Pr1y	18,860	0.2594	0.7124	-0.4862	0.1434	1.3184
IVol	19,607	0.0239	0.0171	0.0078	0.0192	0.0556
SUE	17,745	0.0133	0.1286	-0.1013	0.0018	0.2101

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, earnings from IBES, and stock price, returns, shares outstanding, trading volume, and accounting data from the CRSP/Compustat.

Variables	(1) Short	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
(2) BHAR3d	-0.01													
(3) R_uncertainty	0.03	0.02												
(4) R_modalweak	0.04	0.02	0.53											
(5) R_negative	-0.00	0.00	0.06	0.11										
(6) Ln(n_words)	0.12	0.00	-0.23	-0.10	0.20									
(7) Ln(n_filing)	0.00	-0.00	0.02	-0.02	0.08	0.21								
(8) Ln(Download)	0.14	0.01	-0.03	0.00	0.02	0.13	-0.31							
(9) Ln(SZ)	0.12	0.05	0.00	0.11	-0.09	0.42	0.13	0.26						
(10) Ln (B/M)	-0.11	0.02	-0.06	-0.07	0.22	-0.00	0.02	-0.01	-0.31					
(11) IOwner	0.11	0.07	0.07	0.18	-0.07	0.21	0.05	0.10	0.61	-0.18				
(12) Illiq	-0.13	-0.02	-0.06	-0.12	0.06	-0.20	-0.10	-0.06	-0.40	0.22	-0.34			
(13) Pr1y	0.11	0.04	-0.02	-0.00	-0.01	-0.00	0.05	0.07	0.06	0.14	0.06	-0.08		
(14) IVol	0.21	-0.03	0.01	-0.08	0.00	-0.13	-0.08	-0.06	-0.53	0.07	-0.42	0.30	0.02	
(15) SUE	0.02	0.055	-0.01	0.00	0.03	0.00	0.01	0.01	-0.02	0.07	-0.00	-0.00	0.28	0.04

Table 4: Determinants of shorting volume

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

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

where $Short_{i,t}$ is the shorting volume ratio on the filing dates. *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, earnings from IBES, 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) Short	(2) Short	(3) Short	(4) Short	(5) Short	(6) Short
R_uncertainty	0.067***	0.063***	0.057***		0.062***	
	(3.52)	(3.20)	(2.93)		(3.18)	
R_negative	0.014	0.011	-0.001	0.001	0.011	0.001
	(0.77)	(0.60)	(-0.03)	(0.06)	(0.62)	(0.06)
R_modalweak				0.009		0.009
				(0.65)		(0.65)
Download		0.126***	0.110***	0.111***	0.125***	0.109***
		(5.39)	(5.18)	(5.19)	(5.39)	(5.19)
Download×R_uncertainty					0.000	
					(0.04)	
Download×R_negative					0.011	0.011
					(1.36)	(1.32)
Download×R_modalweak						0.000
						(0.01)
Ln(n_words)	0.085***	0.082***	0.068***	0.024**	0.082***	0.024**
	(4.28)	(4.22)	(3.70)	(2.28)	(4.16)	(2.29)
Ln(n_filing)	-0.052***	-0.002	0.002	0.002	-0.002	0.002
	(-2.79)	(-0.08)	(0.09)	(0.11)	(-0.09)	(0.09)
Ln(SZ)	0.021	-0.003			-0.002	
	(0.32)	(-0.04)			(-0.03)	
Ln(B/M)	-0.068***	-0.070***	-0.089***	-0.087***	-0.070***	-0.087***
	(-2.96)	(-3.03)	(-4.50)	(-4.45)	(-3.03)	(-4.46)
Pr1y	0.077***	0.079***	0.076***	0.076***	0.079***	0.076***
	(3.76)	(3.77)	(4.15)	(4.09)	(3.79)	(4.11)
IOwner			-0.032*	-0.032*		-0.032*
			(-1.94)	(-1.95)		(-1.93)
Illiq			-0.046***	-0.047***		-0.046***
			(-5.13)	(-5.04)		(-5.03)
IVol			0.275***	0.276***		0.275***
			(5.35)	(5.36)		(5.36)
SUE	-0.006	-0.007	-0.010	-0.010	-0.007	-0.010
	(-0.68)	(-0.75)	(-1.24)	(-1.19)	(-0.74)	(-1.19)
Intercept	-0.021***	-0.022***	0.003	0.001	-0.022***	0.001
	(-9.20)	(-4.54)	(0.56)	(0.23)	(-4.51)	(0.20)
Observations	16,548	15,304	15,293	15,293	15,304	15,293
Adjusted R^2	0.42	0.43	0.46	0.46	0.43	0.46

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

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

$$BHAR3d_{i,t} = \alpha + \beta_0 Short_{i,t} + \sum_{j=1}^{3} \beta_j Textual_{i,j,t} + \beta_4 Download_{i,t} + \beta_5 Ln(n\_words)_t$$

$$+\beta_6 Ln(n\_filing)_{i,t} + \sum_{j=1}^{3} \theta_j Short \times Textual_{i,j,t}$$

$$+\theta_4 Short \times Download_{i,j,t} + control_{i,t} + d_{m,t} + \varepsilon_{i,t},$$

where $BHAR3d_{i,t}$ is the 3-day cumulative abnormal return (shorting volume) during the event window [1, 3]. $Short_{i,t}$ is the shorting volume ratio on the filing dates. *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, earnings from IBES, 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) BHAR3d	(2) BHAR3d	(3) BHAR3d	(4) BHAR3d	(5) BHAR3d	(6) BHAR3d
Short	-0.005		-0.007	-0.008	-0.003	-0.004
	(-0.27)		(-0.39)	(-0.50)	(-0.16)	(-0.24)
R_uncertainty		-0.043**	-0.040*		-0.032	
		(-2.00)	(-1.88)		(-1.46)	
R_negative		0.036*	0.037**	0.039**	0.049**	0.050**
		(1.93)	(2.01)	(2.06)	(2.56)	(2.60)
R_modalweak				-0.038***		-0.034**
				(-2.75)		(-2.25)
Download					-0.022*	-0.022*
					(-1.67)	(-1.74)
Short×R_uncertainty			-0.028		-0.023	
			(-1.47)		(-1.11)	
Short×R_negative			0.046***	0.048***	0.036**	0.038**
			(2.80)	(2.84)	(2.25)	(2.30)
Short×R_modalweak				-0.023		-0.019
				(-0.90)		(-0.68)
Short×Download					0.004	0.004
					(0.31)	(0.36)
Ln(n_words)	-0.023	-0.071**	-0.072**	-0.052**	-0.063**	-0.048**
	(-1.31)	(-2.29)	(-2.39)	(-2.64)	(-2.11)	(-2.51)
Ln(n_filing)	-0.067**	-0.066**	-0.067**	-0.067**	-0.072**	-0.072**
	(-2.49)	(-2.42)	(-2.51)	(-2.50)	(-2.47)	(-2.48)
Ln(SZ)	0.362***	0.375***	0.387***	0.387***	0.334***	0.334***
	(4.98)	(5.17)	(5.44)	(5.42)	(4.62)	(4.64)
Ln(B/M)	0.181***	0.181***	0.180***	0.178***	0.175***	0.173***
	(6.39)	(6.56)	(6.47)	(6.32)	(6.45)	(6.33)
Pr1y	-0.058***	-0.060***	-0.061***	-0.060***	-0.058***	-0.058***
	(-3.64)	(-3.70)	(-3.72)	(-3.68)	(-3.45)	(-3.40)
SUE	0.054***	0.054***	0.054***	0.054***	0.056***	0.056***
	(3.44)	(3.44)	(3.50)	(3.51)	(3.55)	(3.55)
Intercept	-0.005***	-0.008***	-0.007***	-0.007***	-0.013***	-0.012***
	(-2.82)	(-3.96)	(-3.59)	(-3.08)	(-2.93)	(-2.71)
Observations	16,542	16,542	16,542	16,542	15,303	15,303
Adjusted R^2	0.08	0.08	0.08	0.08	0.08	0.08

Table 6: Return predictability of 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:

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

where $BHAR1w \ldots 52w_{i,t}$ is the 1-week…52-week cumulative abnormal return. $Short_{i,t}$ is the shorting volume ratio on the filing dates. *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, earnings from IBES, 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) BHAR1w	(2) BHAR1w	(3) BHAR2w	(4) BHAR2w	(5) BHAR4w	(6) BHAR4w
Short	0.006	0.005	0.008	0.007	-0.010	-0.011
	(0.29)	(0.24)	(0.31)	(0.26)	(-0.44)	(-0.53)
R_uncertainty	-0.021		-0.005		-0.018	
	(-1.17)		(-0.25)		(-1.05)	
R_negative	0.042**	0.043**	0.038*	0.040*	0.032	0.033
	(2.01)	(2.07)	(1.81)	(1.90)	(1.59)	(1.63)
R_modalweak		-0.030**		-0.030**		-0.022
		(-2.22)		(-2.30)		(-1.66)
Download	-0.010	-0.010	-0.018	-0.018	-0.017	-0.017
	(-0.70)	(-0.73)	(-1.40)	(-1.39)	(-1.16)	(-1.16)
Short×R_uncertainty	-0.028		-0.053**		-0.043**	
	(-1.56)		(-2.38)		(-2.53)	
Short×R_negative	0.033*	0.035**	0.061**	0.064**	0.039**	0.042**
	(1.97)	(2.07)	(2.51)	(2.57)	(2.13)	(2.26)
Short×R_modalweak		-0.022		-0.044*		-0.041**
		(-1.03)		(-1.79)		(-2.56)
Short×Download	0.005	0.005	0.002	0.003	0.006	0.007
	(0.41)	(0.48)	(0.19)	(0.34)	(0.53)	(0.66)
Ln(n_words)	-0.050*	-0.043**	-0.045	-0.050***	-0.047*	-0.039**
	(-1.98)	(-2.10)	(-1.65)	(-2.85)	(-1.88)	(-2.39)
Ln(n_filing)	-0.053*	-0.053*	-0.065**	-0.065**	-0.082***	-0.081***
	(-1.82)	(-1.82)	(-2.39)	(-2.39)	(-3.41)	(-3.43)
Ln(SZ)	0.401***	0.402***	0.411***	0.414***	0.133	0.134
	(4.46)	(4.48)	(3.34)	(3.39)	(0.98)	(1.00)
Ln(B/M)	0.188***	0.187***	0.247***	0.246***	0.201***	0.200***
	(5.89)	(5.86)	(7.85)	(7.93)	(6.69)	(6.68)
Pr1y	-0.080***	-0.080***	-0.096***	-0.096***	-0.069**	-0.069**
	(-3.38)	(-3.35)	(-3.49)	(-3.47)	(-2.61)	(-2.62)
SUE	0.059***	0.059***	0.056***	0.056***	0.067***	0.067***
	(3.46)	(3.46)	(3.03)	(3.05)	(3.59)	(3.57)
Intercept	-0.018***	-0.018***	-0.019**	-0.019**	-0.003	-0.002
	(-3.47)	(-3.45)	(-2.40)	(-2.56)	(-0.29)	(-0.22)
Observations	15,301	15,301	15,301	15,301	15,298	15,298
Adjusted R^2	0.07	0.07	0.06	0.06	0.08	0.08

	(7) BHAR12w	(8) BHAR12w	(9) BHAR24w	(10) BHAR24w	(11) BHAR52w	(12) BHAR52w
Short	-0.019	-0.021	-0.032**	-0.034**	-0.074***	-0.074***
	(-0.90)	(-1.00)	(-2.50)	(-2.61)	(-5.30)	(-5.18)
R_uncertainty	-0.049*		-0.047*		0.002	
	(-1.91)		(-1.74)		(0.10)	
R_negative	0.022	0.020	-0.005	-0.006	-0.037*	-0.036*
	(1.21)	(1.11)	(-0.26)	(-0.35)	(-1.85)	(-1.83)
R_modalweak		-0.005		-0.009		-0.021
		(-0.33)		(-0.57)		(-1.67)
Download	-0.013	-0.014	-0.035**	-0.036**	-0.046***	-0.046***
	(-0.91)	(-0.93)	(-2.45)	(-2.48)	(-3.00)	(-3.01)
Short×R_uncertainty	-0.047***		-0.035**		-0.030**	
	(-3.27)		(-2.50)		(-2.38)	
Short×R_negative	-0.007	-0.004	-0.007	-0.005	-0.018	-0.016
	(-0.35)	(-0.19)	(-0.43)	(-0.28)	(-1.47)	(-1.34)
Short×R_modalweak		-0.023		-0.019		-0.024**
		(-1.62)		(-1.43)		(-1.99)
Short×Download	0.009	0.010	-0.005	-0.004	-0.001	-0.001
	(0.60)	(0.66)	(-0.42)	(-0.36)	(-0.13)	(-0.05)
Ln(n_words)	-0.100***	-0.057***	-0.068**	-0.030**	-0.013	-0.022
	(-3.69)	(-4.79)	(-2.59)	(-2.17)	(-0.56)	(-1.39)
Ln(n_filing)	-0.065***	-0.065***	-0.043**	-0.043**	-0.036*	-0.036*
	(-3.28)	(-3.29)	(-2.18)	(-2.18)	(-1.73)	(-1.70)
Ln(SZ)	-0.566***	-0.575***	-0.824***	-0.830***	-1.732***	-1.729***
	(-5.44)	(-5.61)	(-4.95)	(-5.00)	(-8.01)	(-7.96)
Ln(B/M)	0.140***	0.136***	0.148***	0.145***	0.108***	0.108***
	(4.12)	(4.05)	(4.82)	(4.74)	(3.04)	(3.01)
Pr1y	-0.043*	-0.040*	-0.037*	-0.035*	0.003	0.003
	(-1.96)	(-1.89)	(-1.82)	(-1.76)	(0.16)	(0.15)
SUE	0.032***	0.032***	0.022*	0.022*	0.027	0.027
	(2.75)	(2.68)	(1.97)	(1.93)	(1.57)	(1.56)
Intercept	0.033***	0.035***	0.070***	0.072***	0.167***	0.167***
	(4.42)	(4.82)	(6.23)	(6.38)	(10.08)	(9.93)
Observations	15,273	15,273	15,230	15,230	15,067	15,067
Adjusted R^2	0.09	0.09	0.12	0.12	0.22	0.22

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:

$$BHAR3d_{i,t}\ (BHAR1w \ldots 52w_{i,t}) = \alpha + \beta_1 \widehat{Short}_{i,t} + \beta_2 Ln(n\_words)_t + \beta_3 Ln(n\_filing)_t$$
$$+\beta_4 Download_{i,t} + control_{i,t} + d_{m,t} + \varepsilon_{i,t},$$

where $\widehat{Short}_{i,t}$ is the fitted shorting volume ratio on filing dates and is obtained from the following regression,

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

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, earnings from IBES, 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.

Panel A: Predicted component

	(1) BHAR3d	(2) BHAR1w	(3) BHAR2w	(4) BHAR4w	(5) BHAR12w	(6) BHAR24w	(7) BHAR52w
$\widehat{Short}_{i,t}$	-0.041**	-0.035**	-0.031**	-0.026*	-0.015	-0.019	-0.021
	(-2.40)	(-2.40)	(-2.10)	(-1.96)	(-0.95)	(-1.01)	(-1.58)
Ln(n_words)	-0.044**	-0.038*	-0.042**	-0.034*	-0.058***	-0.040**	-0.039**
	(-2.04)	(-1.83)	(-2.37)	(-1.88)	(-4.31)	(-2.47)	(-2.40)
Ln(n_filing)	-0.071**	-0.052*	-0.063**	-0.080***	-0.065***	-0.043**	-0.036
	(-2.44)	(-1.75)	(-2.24)	(-3.29)	(-3.20)	(-2.13)	(-1.65)
Download	-0.020	-0.007	-0.014	-0.016	-0.014	-0.042***	-0.056***
	(-1.50)	(-0.45)	(-1.04)	(-1.06)	(-0.92)	(-2.68)	(-3.57)
Ln(SZ)	0.314***	0.383***	0.386***	0.113	-0.581***	-0.827***	-1.720***
	(4.31)	(4.32)	(3.16)	(0.83)	(-5.71)	(-4.92)	(-7.84)
Ln(B/M)	0.177***	0.189***	0.248***	0.202***	0.138***	0.147***	0.110***
	(6.51)	(6.00)	(8.00)	(6.69)	(4.11)	(4.78)	(3.02)
Pr1y	-0.057***	-0.077***	-0.092***	-0.067**	-0.040*	-0.037*	-0.002
	(-3.38)	(-3.30)	(-3.30)	(-2.47)	(-1.89)	(-1.85)	(-0.08)
SUE	0.057***	0.059***	0.056***	0.067***	0.032***	0.022*	0.027
	(3.56)	(3.46)	(3.01)	(3.56)	(2.71)	(1.96)	(1.56)
Intercept	-0.010**	-0.016***	-0.018**	-0.000	0.036***	0.070***	0.166***
	(-2.41)	(-3.08)	(-2.25)	(-0.06)	(5.55)	(6.11)	(9.67)
Observations	15,303	15,301	15,301	15,298	15,273	15,230	15,067
Adjusted R^2	0.07	0.07	0.06	0.08	0.09	0.12	0.21

Panel B: Residual component

	(1) BHAR3d	(2) BHAR1w	(3) BHAR2w	(4) BHAR4w	(5) BHAR12w	(6) BHAR24w	(7) BHAR52w
Res_short	0.002	0.011	0.014	-0.005	-0.017	-0.036***	-0.076***
	(0.11)	(0.51)	(0.53)	(-0.21)	(-0.87)	(-2.67)	(-4.91)
Ln(n_words)	-0.018	-0.017	-0.023	-0.017	-0.047***	-0.026**	-0.021
	(-1.06)	(-0.88)	(-1.43)	(-1.01)	(-4.19)	(-2.28)	(-1.56)
Ln(n_filing)	-0.072**	-0.052*	-0.063**	-0.080***	-0.065***	-0.043**	-0.036*
	(-2.45)	(-1.77)	(-2.25)	(-3.31)	(-3.22)	(-2.18)	(-1.73)
Download	-0.020	-0.008	-0.016	-0.015	-0.012	-0.037**	-0.047***
	(-1.55)	(-0.56)	(-1.21)	(-1.00)	(-0.75)	(-2.48)	(-3.09)
Ln(SZ)	0.307***	0.377***	0.381***	0.108	-0.584***	-0.830***	-1.724***
	(4.19)	(4.24)	(3.12)	(0.79)	(-5.73)	(-4.96)	(-7.94)
Ln(B/M)	0.176***	0.189***	0.249***	0.202***	0.136***	0.144***	0.104***
	(6.44)	(5.96)	(7.98)	(6.67)	(4.06)	(4.70)	(2.93)
Pr1y	-0.056***	-0.078***	-0.092***	-0.066**	-0.038*	-0.034*	0.005
	(-3.39)	(-3.36)	(-3.43)	(-2.51)	(-1.81)	(-1.70)	(0.23)
SUE	0.057***	0.059***	0.056***	0.066***	0.032***	0.022*	0.026
	(3.55)	(3.46)	(3.01)	(3.54)	(2.68)	(1.92)	(1.54)
Intercept	-0.009**	-0.015***	-0.017**	0.000	0.036***	0.070***	0.164***
	(-2.12)	(-2.87)	(-2.12)	(0.01)	(5.61)	(6.10)	(9.72)
Observations	15,303	15,301	15,301	15,298	15,273	15,230	15,067
Adjusted R^2	0.07	0.07	0.06	0.07	0.09	0.12	0.22

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:

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

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, earnings from IBES, 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)
	$\Delta FEPS1$	$\Delta FEPS1$	$\Delta FEPS2$	$\Delta FEPS2$
Short	-0.022	-0.023	-0.038**	-0.039**
	(-1.31)	(-1.33)	(-2.33)	(-2.40)
R_uncertainty	-0.035		-0.026	
	(-1.19)		(-1.29)	
R_negative	0.060***	0.060***	0.039*	0.040**
	(2.74)	(2.76)	(1.95)	(2.01)
R_modalweak		-0.030*		-0.035**
		(-1.95)		(-2.25)
Download	-0.009	-0.010	0.001	-0.000
	(-0.68)	(-0.74)	(0.05)	(-0.02)
Short×R_uncertainty	-0.037**		-0.031**	
	(-2.20)		(-2.29)	
Short×R_negative	0.001	0.004	-0.007	-0.004
	(0.06)	(0.16)	(-0.39)	(-0.26)
Short×R_modalweak		-0.019		-0.006
		(-1.20)		(-0.56)
Short×Download	-0.003	-0.003	-0.017	-0.017
	(-0.21)	(-0.18)	(-1.50)	(-1.40)
Ln(n_words)	-0.026	-0.006	-0.019	-0.010
	(-0.83)	(-0.35)	(-0.78)	(-0.65)
Ln(n_filing)	-0.007	-0.006	-0.024	-0.024
	(-0.28)	(-0.26)	(-1.01)	(-0.98)
Ln(SZ)	0.063	0.058	0.063	0.057
	(0.35)	(0.33)	(0.47)	(0.43)
Ln(B/M)	0.178***	0.174***	0.129***	0.126***
	(3.99)	(3.90)	(2.98)	(2.94)
Pr1y	0.010	0.011	-0.003	-0.001
	(0.53)	(0.58)	(-0.15)	(-0.06)
SUE	0.117***	0.117***	0.083***	0.083***
	(3.19)	(3.18)	(3.78)	(3.81)
Intercept	-0.031	-0.031	-0.017	-0.015
	(-0.44)	(-0.45)	(-0.34)	(-0.29)
Observations	11,051	11,051	10,798	10,798
Adjusted R^2	0.17	0.17	0.25	0.25

Table 9: Changes of fundamental ratios after filing dates

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

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

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, earnings from IBES, 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
Short	-0.004	-0.004	-0.026*	-0.025*	-0.012	-0.013
	(-0.18)	(-0.19)	(-1.77)	(-1.73)	(-1.10)	(-1.16)
R_uncertainty	0.026		0.008		-0.016	
	(1.02)		(0.34)		(-0.79)	
R_negative	0.053**	0.054***	0.094***	0.095***	0.002	0.002
	(2.64)	(2.76)	(4.31)	(4.31)	(0.12)	(0.10)
R_modalweak		-0.001		-0.012		-0.008
		(-0.10)		(-0.82)		(-0.68)
Download	-0.035**	-0.034**	-0.041***	-0.041***	-0.013	-0.013
	(-2.29)	(-2.24)	(-2.76)	(-2.75)	(-1.37)	(-1.41)
Short×R_uncertainty	-0.027		-0.019		-0.024**	
	(-1.41)		(-1.59)		(-2.53)	
Short×R_negative	0.004	0.006	0.000	0.002	0.004	0.005
	(0.25)	(0.35)	(0.01)	(0.10)	(0.27)	(0.33)
Short×R_modalweak		-0.029*		-0.022**		-0.001
		(-1.71)		(-2.10)		(-0.23)
Short×Download	-0.023*	-0.022	0.006	0.006	-0.009*	-0.009
	(-1.69)	(-1.64)	(0.57)	(0.59)	(-1.69)	(-1.63)
Ln(n_words)	-0.006	-0.028	0.017	0.005	-0.010	0.003
	(-0.23)	(-1.39)	(0.61)	(0.25)	(-0.57)	(0.23)
Ln(n_filing)	-0.033*	-0.033*	-0.045	-0.044	-0.008	-0.008
	(-1.84)	(-1.82)	(-1.54)	(-1.53)	(-0.43)	(-0.43)
Ln(SZ)	-0.329***	-0.326***	-0.174**	-0.170**	-0.030	-0.033
	(-2.72)	(-2.68)	(-2.35)	(-2.29)	(-0.32)	(-0.34)
Ln(B/M)	0.059	0.060	0.328***	0.329***	0.040	0.038
	(1.26)	(1.27)	(10.87)	(10.88)	(1.50)	(1.45)
Pr1y	0.051***	0.050***	-0.007	-0.008	-0.002	-0.000
	(2.86)	(2.76)	(-0.47)	(-0.52)	(-0.14)	(-0.01)
SUE	-0.001	-0.000	-0.014	-0.013	-0.007	-0.007
	(-0.03)	(-0.02)	(-0.99)	(-0.98)	(-0.64)	(-0.66)
Intercept	0.031***	0.030***	0.038***	0.038***	0.013	0.013
	(3.44)	(3.21)	(6.15)	(6.30)	(1.51)	(1.55)
Observations	14,979	14,979	14,760	14,760	14,821	14,821
Adjusted R^2	-0.02	-0.03	0.05	0.05	0.23	0.23

Table 10: Predictability of crash risk

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

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

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, earnings from IBES, 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
Short	0.003	0.003	0.000	-0.000	-0.008	-0.008
	(0.25)	(0.21)	(0.00)	(-0.03)	(-0.67)	(-0.71)
R_uncertainty	-0.054**		-0.051*		-0.047*	
	(-2.15)		(-1.92)		(-1.83)	
R_negative	0.016	0.013	0.018	0.015	-0.001	-0.003
	(0.68)	(0.58)	(0.86)	(0.78)	(-0.05)	(-0.12)
R_modalweak		-0.005		-0.007		-0.012
		(-0.26)		(-0.34)		(-0.65)
Download	-0.014	-0.015	-0.016	-0.017	-0.017	-0.018
	(-0.97)	(-1.03)	(-1.12)	(-1.17)	(-1.37)	(-1.44)
Short×R_uncertainty	0.020*		0.018		0.015	
	(1.84)		(1.65)		(1.32)	
Short×R_negative	0.002	0.001	0.003	0.003	-0.004	-0.005
	(0.24)	(0.16)	(0.43)	(0.34)	(-0.42)	(-0.47)
Short×R_modalweak		0.015*		0.013		0.013
		(1.68)		(1.49)		(1.38)
Short×Download	0.004	0.003	0.003	0.003	0.000	0.000
	(0.33)	(0.28)	(0.34)	(0.29)	(0.05)	(0.00)
Ln(n_words)	-0.034	0.011	-0.030	0.011	-0.020	0.015
	(-1.03)	(0.44)	(-0.93)	(0.44)	(-0.67)	(0.68)
Ln(n_filing)	-0.010	-0.011	-0.011	-0.011	-0.016	-0.016
	(-0.49)	(-0.50)	(-0.48)	(-0.49)	(-0.72)	(-0.72)
Ln(SZ)	0.960***	0.955***	0.855***	0.850***	0.578***	0.574***
	(9.29)	(9.21)	(8.77)	(8.70)	(7.45)	(7.37)
Ln(B/M)	0.004	0.001	0.010	0.008	0.004	0.002
	(0.13)	(0.05)	(0.38)	(0.30)	(0.16)	(0.07)
Pr1y	-0.009	-0.008	-0.013	-0.012	-0.002	-0.001
	(-0.89)	(-0.83)	(-1.32)	(-1.28)	(-0.18)	(-0.13)
SUE	0.017*	0.017*	0.016**	0.016*	0.007	0.006
	(1.99)	(1.96)	(2.00)	(1.98)	(0.80)	(0.77)
Intercept	-0.082***	-0.081***	-0.073***	-0.071***	-0.049***	-0.048***
	(-9.39)	(-9.27)	(-8.78)	(-8.75)	(-8.03)	(-7.77)
Observations	14,810	14,810	14,810	14,810	14,810	14,810
Adjusted R^2	0.07	0.07	0.06	0.06	0.03	0.03

Appendix: Variable definition

Short sale variable	
Short	Shorting volume ratio, daily shorting volume/shares outstanding, on the filing date
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 BHAR3d, the event window is [1, 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.
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.
SUE	SUE is a firm's standardized unexplained earnings, defined as the realized earnings per share (EPS) minus EPS from quarter t-4, scaled by the stock price at quarter t-4.

Appendix Table 1: Return predictability of shorting volume and textual information around filing dates

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

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

where $BHAR4d_{i,t}$ is the 4-day cumulative abnormal return (shorting volume) during the event window [0, 3]. $Short_{i,t}$ is the shorting volume ratio on the filing dates. *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, earnings from IBES, 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
Short	0.081**		0.079**	0.076***	0.085***	0.083***
	(2.46)		(2.50)	(2.69)	(2.67)	(3.02)
R_uncertainty		-0.042*	-0.043**		-0.030	
		(-1.92)	(-2.03)		(-1.30)	
R_negative		0.027	0.028	0.030	0.037*	0.039**
		(1.49)	(1.46)	(1.54)	(1.92)	(2.00)
R_modalweak				-0.047***		-0.045***
				(-3.70)		(-3.20)
Download					-0.038***	-0.038***
					(-2.78)	(-2.79)
Short×R_uncertainty			-0.060**		-0.058**	
			(-2.22)		(-2.02)	
Short×R_negative			0.047**	0.054**	0.047**	0.052**
			(2.00)	(2.20)	(2.09)	(2.25)
Short×R_modalweak				-0.080**		-0.081**
				(-2.32)		(-2.17)
Short×Download					-0.016	-0.014
					(-0.75)	(-0.70)
Ln(n_words)	-0.018	-0.059	-0.070**	-0.051**	-0.058	-0.050**
	(-0.77)	(-1.62)	(-2.00)	(-2.21)	(-1.61)	(-2.11)
Ln(n_filing)	-0.071**	-0.074***	-0.071***	-0.071***	-0.083***	-0.083***
	(-2.64)	(-2.72)	(-2.66)	(-2.67)	(-2.85)	(-2.89)
Ln(SZ)	0.434***	0.447***	0.462***	0.466***	0.442***	0.448***
	(4.85)	(4.90)	(5.32)	(5.54)	(5.09)	(5.34)
Ln(B/M)	0.172***	0.168***	0.173***	0.171***	0.173***	0.172***
	(6.25)	(6.26)	(6.38)	(6.18)	(5.79)	(5.64)
Pr1y	-0.073***	-0.068***	-0.078***	-0.079***	-0.077***	-0.079***
	(-3.44)	(-2.96)	(-3.55)	(-3.70)	(-3.46)	(-3.61)
SUE	0.077***	0.077***	0.078***	0.078***	0.079***	0.079***
	(4.58)	(4.53)	(4.66)	(4.68)	(4.64)	(4.62)
Intercept	-0.006**	-0.010***	-0.006**	-0.004	-0.016***	-0.014**
	(-2.15)	(-3.64)	(-2.12)	(-1.36)	(-2.73)	(-2.50)
Observations	16,542	16,542	16,542	16,542	15,303	15,303
Adjusted R^2	0.09	0.08	0.09	0.09	0.09	0.09