



Nothing but good intentions: the search for equity and stock price crash risk

Doron Reichmann¹ · Rouven Möller¹ · Tobias Hertel¹

Accepted: 3 March 2022 / Published online: 30 March 2022
© The Author(s) 2022

Abstract

In this study, we examine whether and how managers' intentions to raise equity are associated with future stock price crash risk. Therefore, we apply modern information retrieval techniques to corporate textual disclosures by constructing document embeddings that preserve contextual relationships in managers' discussions on liquidity and capital resources. Using these document embeddings, we construct a continuous measure of managers' intentions to raise equity. We document that the search for equity is associated with higher future stock price crash risk. Further analyses suggest that managers with stronger intentions to raise equity are more likely to block negative news flow and that these intentions reinforce the effects of earnings manipulation and textual obfuscation on stock price crash risk. In summary, our results suggest that managers' search for equity incentivises managerial bad news hoarding.

Keywords Stock price crash risk · Information opacity · Agency theory · Natural language processing · Textual disclosures · Equity finance · Information retrieval

JEL Classification G12 · G14 · G32 · M41

1 Introduction

Since the seminal work of Jin and Myers (2006) firm-specific stock price crashes are linked to bad news hoarding by firm managers. Due to incentives such as compensation packages, empire building, and reputation concerns (Ball 2009; Graham et al.

✉ Doron Reichmann
doron.reichmann@rub.de

Rouven Möller
rouven.moeller@rub.de

Tobias Hertel
tobias.hertel@rub.de

¹ Department of Banking and Finance, Ruhr University Bochum, Bochum, Germany

2005), managers have the tendency to hide adverse information from outside investors which leads to bad news being stockpiled within the firm. Once the accumulated bad news reaches a tipping point where it can no longer be contained, stock prices are prone to crash. Given the devastating effects of crash risk on investor welfare, literature in the field of accounting and finance has increasingly turned to understanding the determinants of bad news hoarding behaviour. In this study we examine managers' intentions to offer equity (hereafter: equity intent) as a potential incentive to hoard bad news. In the case of equity offerings, managers are known to time the market with regard to the current market valuation (Graham and Harvey 2001). More specifically, firm managers seeking equity are incentivised to provide more optimistic information prior to the equity offering to raise more capital, even if this contrasts the firm's true performance (Lang and Lundholm 2010). Given this evidence, we believe that managers who try to persuade investors to invest in their company are more likely to hoard bad news. However, whether managers' equity intent serves as a driver for bad news hoarding behaviour as reflected in future stock price crash risk remains an open empirical question. To address this issue, we apply modern techniques of information retrieval to firm disclosures to identify cues revealing managers' equity intent and quantify its effects on future stock price crash risk.

A prominent example of managerial bad news hoarding with severe effects on investor welfare is the case of Marrone Bio Innovations. Former COO Hector M. Absi Jr. allegedly profited from hiding various sales concessions from independent auditors and outside investors.¹ When the SEC uncovered the alleged fraud in August 2014, the company's share price plunged by more than 40%. Given the severe effects of opportunistic management behaviour, section 954 of the 2010 Dodd-Frank Act mandates that publicly listed companies adopt clawback provisions that require managers to return part of all benefits they received as the result of misstated accounting numbers. However, Bao et al. (2018) provide empirical evidence that clawback adoptions do not help mitigating crash risk as managers are induced to hide bad news through different channels than manipulated accounting numbers. Since stock price crash risk can neither be mitigated by portfolio diversification (Sunder 2010), literature in the field of accounting and finance has focused on identifying determinants of managerial bad news hoarding to help investors detecting crash prone companies via screening.² Interestingly, we find that, prior to the stock price crash of Marrone Bio Innovations in 2014, the company openly stated in its

¹ Note that accounting fraud constitutes an extreme and illegal form of bad news hoarding. However, it is documented that managers have significant discretion over both accounting numbers and textual disclosures (Dechow et al. 1995; Hung and Qiao 2017; Lo et al. 2017; Roychowdhury 2006). Hence, bad news hoarding is likely to come in shades. While some managers may engage in extreme forms of bad news hoarding that result in accounting scandals, other managers will rather aim to obfuscate financial reporting by remaining comfortably within the scope of discretion permitted by regulators.

² Sunder (2010) differentiates between two perspectives on risk mitigation. The first risk perspective describes the uncertainty of outcomes, which is symmetric in losses and gains and can be reduced through diversification. The second risk perspective focuses on the magnitude and probability of losses and can only be reduced via screening mechanisms. While the overall uncertainty of outcome of a specific firm (stock volatility) can be reduced through diversification, the most appropriate response towards extreme tail-risks (i.e., stock price crashes) are screening mechanisms.

10-K filing for the fiscal year 2013 that it was searching for additional external funding (e.g. equity intent) in order to expand its operational business. Considering our opening arguments, the intention to raise equity could have served as an incentive for the management to provide an overly optimistic outlook and hide bad news from investors. In this case, screening companies for managers' equity intent could have helped investors to better assess firm-specific crash risk and preserve shareholder welfare.

Although firm fundamentals (e.g., leverage ratio, market-book-ratio) help to explain firm financing choices on average, balance sheet items cannot reveal corporate financing *intentions*. To grasp managers' equity intent, we build on a concept proposed by Hoberg and Maksimovic (2014) and search for commonly used phrases in the Liquidity and Capital Resources (LIQ+CAP) section, a mandated disclosure of the MD&A section in 10-K files, that managers use to express their equity intent. However, managers' equity intent is likely to come in shades. The authors suggest that, even if managers do not actively offer equity eventually, the mere intention to do so should be reflected in their choice of words when discussing the capital resources of their firm. Consequently, we follow Hoberg and Maksimovic (2014) to compile a continuous measure of equity intent that allows us to consider early signals that may mark the start of managers' tendencies to hoard bad news. Specifically, we estimate cosine similarities between vector representations of LIQ+CAP sections in which managers explicitly state that they are issuing equity and all other LIQ+CAP sections in our sample. Instead of using a standard Bag-of-Words (BOW) model that is frequently used in accounting and finance research to estimate document similarities, we construct neural network embeddings of each LIQ+CAP section using the *Doc2Vec* model developed by Le and Mikolov (2014) that is proven to outperform standard BOW models as it accounts for semantic relationships.

After the identification of equity intent, we quantify its effect on firm specific future stock price crash risk. We follow the extant literature and analyse the effect of equity intent in conjunction with earnings manipulation—another popular determinant of stock price crash risk related to bad news hoarding (Hutton et al. 2009; Kim et al. 2019). Our results suggest that firms with stronger equity intent are more likely to suffer from stock price crashes in the future. These findings hold for different popular stock price crash risk measures and remain robust after controlling for firms' earnings manipulation, which is a popular vehicle to hide adverse information as well as an extensive set of control variables. Moreover, the effects have a comparable magnitude like the effect earnings manipulation and textual obfuscation exert on stock price crash risk. To strengthen the case that equity intent is in fact a motive for managers to hoard bad news, we conduct a causal mediation analysis where we use the fraction of negative words in each firm's MD&A section as a mediator variable. In line with the theory of bad news hoarding, our results show that stronger equity intent incentivises managers to withhold negative information by using less negative language in their MD&A, supporting the view that bad news hoarding serves as an important economic mechanism predicting the likelihood of future price crashes following a sudden release of accumulated negative information.

In additional analyses we examine how managers' equity intent relates to the usage of earnings manipulation and textual obfuscation. The results show that equity

intent reinforces the effects of these two popular vehicles of bad news hoarding. While earnings manipulation and textual obfuscation offer only marginal predictive power for future stock price crash risk for firms with low equity intent, they are strongly significant determinants for future stock price crash risks for high equity intent firms. Furthermore, we conduct a series of analyses to assess the robustness of our main results. Our robustness tests show that equity intent is an early warning indicator for stock price crash risk up to a prediction window of three years while predictive power decreases for wider prediction windows. In the next analysis we evaluate whether equity intent also predicts upward stock price jumps. If equity intent is a motive for bad news hoarding, we should not be able to predict upward stock price jumps in a similar way as stock price crashes. In fact, the results are generally consistent with this notion. We further check whether our findings are driven by model specifications of the applied neural network to estimate the LIQ+CAP embeddings. By exploring an alternative approach to estimate LIQ+CAP embeddings, we find that our main results remain robust for all stock price crash risk measures. Finally, we test the sensitivity of our results by controlling for several additional business risk and textual proxies that are associated with stock price crash risk. Again, our results remain virtually unchanged.

This paper contributes to the literature in the following way: In line with the SEC's plan to use textual analyses to detect managerial opportunistic behaviour (Eaglesham 2013), we contribute to the literature that uses textual information to complement the prediction of stock price crashes. Consistent with El-Haj et al. (2019), who state that "mainstream accounting and finance research appears to be behind the curve in terms of computational linguistic sophistication" (p. 266), prior works in the field of stock price crash risk mainly rely on basic content analysis methods, providing evidence that more complex and ambiguous company disclosures obfuscates financial information and thus, facilitates bad news hoarding (Ertugrul et al. 2017; Kim et al. 2019). Instead of identifying a new *vehicle* that managers use to hide bad news from outside investors, we draw on the bad news hoarding theory of Jin and Myers (2006) which requires the existence of certain *incentives* that drive managers decision to hoard bad news in the first place. Hence, by raising the question of whether company disclosures contain information about managerial incentives to hide bad news, we add a new perspective on how to use company filings to identify determinants of future crash risk. Furthermore, we argue that, as compared to prior works relying on broad textual characteristics of company files, our approach is superior in detecting managers' bad news hoarding as it identifies a theoretically grounded incentive to use asymmetric information in efforts to time markets within managers' capital structure decisions.

The remainder of this paper is structured as follows: Sect. 2 reviews literature in the field of stock price crash risk and provides our research hypothesis. Section 3 describes our data, the crash risk measures, the neural network embeddings we use to identify equity intent and the models to predict future crash risk. In Sect. 4 we present our prediction results before Sect. 5 provides additional analyses and Sect. 6 concludes the paper.

2 Literature review and hypothesis

Since Jensen and Meckling (1976) first introduced agency costs, a large stream of literature analyses how linking the performance measurement of managers to certain budgetary or financial targets exacerbates existing conflicts of interests between them and outside investors that incentivise managers to engage in earnings management (Franco-Santos et al. 2012; Jensen 2005). Likewise, since the current market valuation of a firm can influence its managers' strategic capital structure decisions, in theory, managers are incentivised to withhold information that could negatively affect the firms' market valuation (Alti 2003; Baker and Wurgler 2002; DeAngelo et al. 2010; Graham and Harvey 2001). Unsurprisingly managers that engage in earnings management in this context, also try to maximize their seasoned equity offerings (SEO) proceeds by increasing SEO pricing (Kim and Park 2005) but at the same time experience a declining operating performance as well as a distinctly post-SEO underperformance as the real consequences of their earnings management decisions reach the market (Cohen and Zarowin 2010; DuCharme et al. 2004; Walker and Yost 2008). However, thus far the literature has neglected that managerial opportunistic behaviour may also have much larger implications for investors welfare due to its impact on stock price crash risk. Hence, we contribute to this literature by analysing how a firms equity intent serves as a motive for opportunistic behaviour in the form of bad news hoarding, which leads to stock price crashes.

In this regard Jin and Myers (2006) set up a model in which the degree of opacity affects the division of risk bearing between firm insiders (managers) and outside investors. Managers' ambitions to capture parts of the firm's cash flow is limited by outside investors' perception of the firm's cash flow and value. If cash flows exceed investors' expectations, managers are able to capture a larger fraction of it. Conversely, if managers underperform relative to investors' expectations, they suffer reduced compensation and might even run the danger to lose their job. Thus, managers are incentivised to overstate financial performance and increase opacity by strategically withholding negative information from outside investors. In this sense, managers' disclosure preferences are not aligned with those of outside investors (Kothari et al. 2009). However, there is a natural limit to the accumulation of bad news, because the probability that outsider investors grasp the firm's true value continuously increases the more negative news are accumulated. When further obfuscations become too costly or difficult to maintain, all the negative information are released at once, resulting in the firm's stock price to crash. Given the subsequent destruction of shareholder welfare, managers' incentives to engage in bad news hoarding and the underlying mechanisms of this practice have received increasing attention from financial research. Previous literature has already established the positive relation between information asymmetries and stock price crash risk. For instance, incentive-based management compensation, CEO overconfidence (Kim et al. 2016), and overall career concerns (Kothari et al. 2009) provide motives for managers to uphold and expand information asymmetries against outside investors, thus, increasing the risk of a stock price crash (Kim et al. 2011).

In search for other motives for managers to obfuscate corporate disclosure, we turn our focus on corporate financing decisions, more precisely managers' equity intent. Besides the obvious importance for the continuation and success of a firm's business activities, financial management decisions are subject to agency conflicts between managers and outside investors, especially in the case of equity financing. While we do not plan to extend the comprehensive literature on choosing the optimal financial strategy, we are interested in the possibilities of opportunistic management behaviour (and its financial consequences) that are associated with managers' equity intent.

In general, managers should be interested in a successful SEO for at least two reasons. Firstly, their future courses of action inside the firm benefits from higher financial resources. Secondly, they may fear the negative resonance following a disappointing SEO from a career perspective. Conversely, a positive stock performance and a successful equity issuance should enhance their job security and increase their value on the labour market (Chi and Gupta 2009). Consistently, the influential survey of corporate financing decisions by Graham and Harvey (2001) showed that managers proactively chose their equity issuances with regard to the current market valuation of the firm. Since firm outsiders are unable to assess the true value of a firm (given semi-informationally efficient markets), managers are able to exploit these information asymmetries to maximize the corresponding capital inflow. Lang and Lundholm (2010) observe changes in corporate disclosure policies around the period of an SEO. They show that firms seeking for equity tend to provide more detailed and more optimistic disclosure prior to the equity offering, leading to an increase of their stock price prior to the offering announcement. While the increase of the stock price may also be the result of improved economic conditions, Lang and Lundholm (2010) provide evidence, that this effect also holds true for firms without an economic improvement which "hype up" their stock. Interestingly, firms dialling up their disclosure suffer from comparably higher price declines after the actual announcement of the equity offering, indicating a market correction for "hyped up" stocks. Bergmann and Roychowdhury (2008) show that managers strategically adapt the level of disclosure (long-horizon earnings forecasts) to mould investor expectations and foster the firm's valuation. In the case of high investor sentiment, managers are expected to decrease the level of voluntary disclosure to manage future expectations and maintain optimistic firm valuations. Thus, managers are incentivised to uphold high investor sentiment by concealing negative information in the interest of optimizing their SEO. Due to the limits of bad news accumulation, an opportunistic manager should offer equity when the (positive) difference between outside investors' assessment and firm's true value is the highest. In this sense, the search for equity could be interpreted as signal that managers do not expect the market's perception of the firm's value to further increase in the near future (Krasker 1986). Following the notion of Bergmann and Roychowdhury (2008), the search for equity would mark the beginning of a decrease in a firm's disclosure level and the starting point of negative news concealment. Given this evidence, managers who intend to issue equity in the foreseeable future might be more likely to withhold negative news from outside investors to secure a successful SEO and maximize their capital inflow. We thus posit the following hypothesis:

Hypothesis Managers' equity intent incentivises bad news hoarding behaviour as reflected by increasing future stock price crash risk.

The analysis above suggests that managers' equity intent may help investors to better grasp potential risks arising from bad news hoarding behaviour. However, managers' *intentions* cannot be extracted from simple balance sheet data and thus, require a rigorous identification strategy. Therefore, we construct neural network embeddings from textual disclosures to estimate managers' equity intent based on linguistic characteristics. Moreover, since firms seeking equity might generally be associated with a riskier business profile, we employ a battery of variables capturing fundamental risk to alleviate the concern that our results are simply driven by business risk.

3 Sample selection, data, and research design

3.1 Sample selection

Table 1 documents the sample selection process. Following Schmidt et al. (2019) in screening data from *Refinitiv Datastream* and *Worldscope*, our sample starts with available US firm-years from 1994 to 2019. After removing financial firms (SIC codes 6000–6999) and utilities (SIC codes 4900–4999), we collect weekly returns for each firm-year. To avoid bias arising from thin trading, we remove firms with a fiscal year-end price lower than \$1.³ In addition, we remove observations with less than 26 weeks of stock return data and observations with nonpositive total assets at the beginning of a fiscal year.⁴ In order to estimate managers' equity intent, we next collect all available 10-K files from the SEC's EDGAR system for our sample of firms and extract the MD&A section using the parsing procedure proposed by Reichmann and Reichmann (2022).⁵ Having extracted the MD&A section, we compile Regular Expressions to extract the LIQ+CAP section. Since the following methodology relies on a clean identification of the LIQ+CAP section, we describe

³ Note that penny stocks are often traded in thin markets that have a high price volatility and low liquidity. Since we seek to examine the relationship between managers' equity intent and their tendencies to hoard bad news, stock price crashes arising from difficult market conditions could bias our empirical identification. Therefore, we follow the extant literature and remove low-priced firms from our sample (e.g. Kim et al. 2011; Kim et al. 2016; Kim and Zhang 2016; Kim et al. 2019.).

⁴ Note that, while some studies in the field of crash risk also exclude firms with negative book values (e.g. Al Mamun et al. 2021; Wu and Lai 2020), these firms have become much more common in recent years, are often financially healthy, and report strong earnings (Luo et al., 2021). Since Luo et al. (2021) show that the extreme leverage ratios of such firms are mostly driven by high investment demand, they also appear susceptible to the usage of equity financing and thus, are clearly of interest for our research question. However, unreported results suggest that our inferences remain unchanged when excluding negative book value firms.

⁵ For the textual analysis, we use all available 10-K filings *before* dropping observations with insufficient data to measure crash risk. This significantly increases our sample size for the textual analysis and consequently, improves the generalizability of our textual proxies.

the parsing procedure in greater detail in Appendix 1. After dropping observations that have no machine-readable LIQ+CAP section and insufficient financial data to construct control variables, 28,382 observations from 2825 firms remain in the final sample.

3.2 Measuring managers' equity intent

While common balance sheet data does not provide any cues on managers' equity intent, Hoberg and Maksimovic (2014) show that managers address financing needs within the LIQ+CAP section, a mandated disclosure of the MD&A section in 10-K files. Thus, the authors compile a set of phrases managers commonly use in the LIQ+CAP section to express their equity intent. These include the following:

issuing equity securities OR expects equity securities OR through equity financing OR sources equity financing OR seek equity investments OR seek equity financings OR access equity markets OR raised equity arrangements OR undertake equity offerings OR sell common stock OR issuing common stock OR selling common stock OR use equity offerings OR offering equity securities OR planned equity offering OR seek equity offering OR raise equity offering OR equity offering would add OR additional equity offering OR considering equity offering OR seek equity financing OR pursue equity offering OR consummates equity offering OR raises equity capital OR raise equity offering OR sources equity offering

Using the above list of words and phrases, we identify a sample of firms that explicitly express equity intent.⁶ However, manager's equity intent is likely to come in shades. While some managers will explicitly state to seek equity, others may still consider it even though they do not explicitly state to do so. Based on the idea that managers with equity intent are likely to discuss similar content (i.e. use similar words) within the LIQ+CAP section, Hoberg and Maksimovic (2014) suggest to construct continuous measures of managers' equity intent. Specifically, the authors estimate cosine similarities to score how proximate the overall LIQ+CAP vocabulary is to that of firms that explicitly express their equity intent. Therefore, the authors employ a BOW approach to convert each LIQ+CAP section to a vector representation of word counts. Formally, if the full sample of LIQ+CAP sections is made of $|V|$ unique words (the total vocabulary), the BOW approach represents a given LIQ+CAP section j as a vector with a length of $|V|$, where each entry is assigned to a word in the vocabulary and its corresponding word count, $count_{i,j}$. However, research in the field of computational linguistic points out two

⁶ Note that we test whether either one of the phrases appears in a sentence of the LIQ+CAP section. Therefore, these phrases can slightly vary. For instance, we also identify phrases such as "we seek to undertake a new equity offering" even though the phrase is "undertake equity offering". In total we identify 6,499 LIQ+CAP sections in which managers explicitly state their equity intent which makes up 12.6% of our total sample. This is generally consistent with Hoberg and Maksimovic (2014) who identify 12.8% of firms within their sample as equity-focused.

Table 1 Sample selection

Data filters	Dropped observations	Sample size	Number of firms
Refinitiv US availability (1994–2019)		442,260	17,010
Exclude financial and utility firms	163,696	278,564	10,712
Return data available	184,667	93,897	10,433
Fiscal year-end price ≥ 1 \$	13,651	80,246	8305
Weekly returns in a fiscal year ≥ 26	2233	78,013	7915
Nonpositive total assets	3521	74,492	7589
Match with 10-K filings	30,646	43,846	3833
MD&A parsing	2540	41,306	3649
LIQ+CAP parsing	8026	33,280	3437
Missing data for control variables	4898	28,382	2825

This table reports the impact of various data filters on the initial sample retrieved from *Refinitiv Datastream* and *Worldscope*

major drawbacks of this method (Le and Mikolov 2014): firstly, the BOW approach does not account for word order. As long as the same words are used, different documents can have exactly the same vector representations. Secondly, BOW vectors do not account for the semantics of words. For instance, when using the BOW approach, the words “profit”, “revenue”, and “water” are considered equally distant, even though “profit” and “revenue” are semantically much closer than “profit” and “water”. To address these issues, Le and Mikolov (2014) developed *Doc2Vec*, an unsupervised machine learning algorithm that learns dense vector representations from a sample of documents.

The authors build on the famous *Word2Vec* model developed by Mikolov et al. (2013a, b) which represents a neural network that attempts to estimate so-called word embeddings of each word found in a sample of documents. Word embeddings can be understood as learned vector representations of words that capture their semantic information, or put differently, their meaning. Therefore, *Word2Vec* relies on an old concept in linguistics suggesting that words that have similar meaning are more likely to co-occur with similar word-neighbours (Harris 1954). Hence, the neural network “reads” through a sample of documents and thereby learns to predict the neighbouring words of each word it encounters. To improve its predictions, the model continually updates the weights of hidden layers that are used to predict the neighbouring words. When learning is completed after a certain number of iterations, these layer weights are used as effective vector representations of words (i.e. word embeddings). The word embeddings all have a fixed, pre-defined dimension d that captures the semantic information learned from the co-occurrence relationship of each word and its neighbours. As a result, words that co-occur with similar neighbouring words are assigned with more similar word embeddings, that, when plotted into vector space, occupy closer locations to one another.

While *Word2Vec* was developed to estimate vector representations of words, *Doc2Vec* has become a widely used technique to create vector representations of

whole documents such as our LIQ+CAP sections. Le and Mikolov (2014) propose two different architectures to estimate vector representations of documents. The Distributed Memory (DM) model is based on *Word2Vec* but also adds a $d \times N$ document matrix to the modelling process, where d denotes a given number of dimensions and N denotes the total number of documents in the sample. Therefore, besides training word embeddings, the model also trains a d -dimensional feature vector that is unique to each of the N documents in the sample. In comparison, the Distributed Bag Of Words (DBOW) model is less complex as it ignores word order and only uses the document vector to predict a randomly sampled set of words from the respective document. In both cases, the models run against a loss function and continuously update the document matrix to achieve better predictions. After finalizing a given number of iterations, the document matrix contains d -dimensional document vectors for each of the N documents. In their influential paper, Le and Mikolov (2014) test how these so-called document embeddings compare to BOW vectors by estimating cosine similarities between the content of web pages and respective search queries. Their findings show that, compared to a standard BOW model, the usage of *Doc2Vec* gives a 53% relative improvement in terms of error rate when performing information retrieval tasks.⁷ Consequently, we construct document embeddings in the spirit of Le and Mikolov (2014).

We begin by employing extensive text-preprocessing steps to reduce textual noise, improve the generalizability of our *Doc2Vec* model and form phrases that are specific to the LIQ+CAP sections. Specifically, we closely follow the recommendations of Reichmann and Reichmann (2022) by (i) splitting the document into sentences for further processing, (ii) replacing named entities such as firm names, dates, or money values to predefined tags, (iii) employing lemmatization to reduce feature dimensionality,⁸ (iv) forming phrases in the spirit of Mikolov et al. (2013b) and (v) removing stopwords that are defined as words that do not add meaning to a document.⁹ In the following, we illustrate the impact on the text preprocessing steps on an exemplary sentence taken from our sample of LIQ+CAP sections¹⁰:

“Investments in property, plant and equipment were \$128 million in 2009 and \$154 million in 2008”.

After employing preprocessing steps (ii) to (v), the sentence reads as follows:

“investment property_plant_and_equipment -money- -date- -money- -date-”.

⁷ Note, that the authors only report the results of a combined measure that uses both the DM and DBOW *Doc2Vec* architectures. Furthermore, as a baseline, the authors apply a BOW approach that uses term frequency-inverse document frequencies (TF-IDF) weights as the raw count approach used in Hoberg and Maksimovic (2014) performs even worse. Even when using a more sophisticated method such as a weighted Bag-Of-Bigrams with TF-IDF weights, the authors still report a relative improvement of 32% in terms of error rate.

⁸ Note that we use *spaCy* for preprocessing step (ii) and (iii).

⁹ We use the generic stopword list proposed by Loughran and McDonald (<https://sraf.nd.edu/textual-analysis/resources/#StopWords>).

¹⁰ Specifically, the sentence is taken from the Agilent Technologies 10-K filing in 2009 (0001047469-09-010861).

(ii) Named entities such as the money value “\$128 million” and dates like “2009” are replaced with the tags “-money-” and “-date-”¹¹; (iii) lemmatization reduces the word “Investments” (plural) to its base-form “investment” (singular); (iv) the phrase “property, plant and equipment” is found to significantly co-occur within our sample of LIQ+CAP sections, so that it is concatenated to a single token using underscores (“property_plant_and_equipment”);¹² and (v) stopwords outside of phrases like “and”, “in”, and “were” are removed.¹³ As the result shows, the sentence has become shorter and is less likely to reflect linguistic cues that are specific to a certain point in time which helps to reduce model dimensionality and improves the generalizability of the document embeddings.

Next, we use the *Doc2Vec* module of the *gensim* library in Python to estimate vector representations of each LIQ+CAP section. While the DBOW and the DM model only differ slightly, the simpler DBOW model ignores word order which could reduce its ability to capture semantic information. Therefore, for our baseline analysis, we employ the DM model and later apply the DBOW model to test the robustness of our findings. For the implementation of the *Doc2Vec* model, we first have to define the dimensionality d of the document embeddings which the model learns during the training process. While larger vector representations may yield better results, they also require more costly computations. However, based on the findings of Pennington et al. (2014) who show that embedding dimensions greater than 300 offer little improvement in quality, we follow the vast majority of the literature that employs textual embedding models and set number of embedding dimensions equal to 300 (e.g. Lau and Baldwin 2016; Li et al. 2021; Matin et al. 2019). For the remaining parameters, we closely follow the recommendations of Lau and Baldwin (2016) who provide optimal *Doc2Vec* hyper-parameter settings for semantic textual similarity tasks.¹⁴ After 1000 iterations of training over the whole sample of LIQ+CAP sections, we assign each section a 300-dimensional document vector. We then construct an equity focus vector by averaging the vectors of all LIQ+CAP sections in which managers explicitly express their equity intent. Finally, we construct $EQUITY_INTENT_{it}$ as cosine similarities from each LIQ+CAP sections to the equity focus vector. While $EQUITY_INTENT_{it}$ naturally ranges between 0 and 1, higher values indicate a stronger equity intent.

In order to better understand how our main variable of interest is measured, Fig. 1 graphically displays a random sample of 500 LIQ+CAP embeddings and their

¹¹ Note that *spaCy* also detects named entities that consist of multiple words such as “Apple Inc” or “the following year”.

¹² Note that “Property, plant and equipment” refers to a common balance sheet item. Hence, its meaning is not a simple composition of the meanings of its individual words. Since these phrases have a different meaning than their individual words, they should rather be treated as unique tokens (Mikolov et al. 2013b).

¹³ Note that, since stopwords can help understanding commonly used phrases, we remove stopwords after forming phrases. Therefore, the phrase “property_plant_and_equipment” still includes the stopword “and”.

¹⁴ Following Lau and Baldwin (2016), we use a vector size of $d = 300$, a window size of 5, use a down-sampling threshold of $1e^{-6}$, draw 5 “noise words” through negative sampling. Given the size of our dataset we also ignore words accruing less than 5 times.

distant to the equity intent vector by employing t-Distributed Stochastic Neighbour Embeddings (t-SNE) techniques to visualize the 300-dimensional vectors in a 3D scatter plot.¹⁵ Each dot represents the document embedding of a randomly drawn LIQ+CAP section. The redder the dots, the higher $EQUITY_INTENT_{it}$ (i.e. the closer they are to the equity intent vector). Consequently, a red dot that occupies the exact same space as the equity intent vector would have a value of $EQUITY_INTENT_{it} = 1$. Put simply, we would expect that redder firms have a higher propensity to experience a one-year-ahead stock price crash, as they exhibit a stronger intent to issue equity.

In order to validate our measure of equity intent, we examine the correlations between a firm's equity intent and various firm characteristics such as firm size ($LOGMV_{it}$), the market-to-book ratio (MTB_{it}), leverage (LEV_{it}), operating performance (ROA_{it}) as well as risk measures including firm-specific return volatility ($SIGMA_{it}$) and cash flow volatility ($CFVOL_{it}$). In addition, we include proxies for firms' inherent information asymmetry using an intangibility measure ($ADJROTA_{it}$) and R&D expense ($PROP_COST_{it}$). Finally, we also test correlations with firm age (AGE_{it}). The results are reported in the Table 2A. We find that smaller, younger, badly performing, and high-growth firms have a higher equity intent. Furthermore, firms that are less capable of using debt financing, firms with higher cash flow volatility, and firms with higher information asymmetry are more likely to search for equity. In summary, we conclude that our measure of equity intent correlates with common firm characteristics that are likely to constrain firms' financing options to equity financing.

3.3 Measuring crash risk

To measure firm-specific stock price crash risk, we calculate firm-specific weekly returns (W) using the following market-industry index model estimated for each firm and fiscal year:

$$r_{i\tau} = \alpha_i + \beta_{1i}r_{m\tau-1} + \beta_{2i}r_{j\tau-1} + \beta_{3i}r_{m\tau} + \beta_{4i}r_{j\tau} + \beta_{5i}r_{m\tau+1} + \beta_{6i}r_{j\tau+1} + \varepsilon_{i\tau}$$

where r_i , r_j , and r_m are the returns in week τ for stock i , the value-weighted Fama–French index for industry j , and the CRSP value-weighted market index m , respectively. Since 10-K files are usually filed within a three month period after a firm's official fiscal year-end, we define a fiscal year as the 12-month period ending three month after a firm's official fiscal year-end to avoid look-ahead bias (Kim et al. 2019).¹⁶ To account for non-synchronous trading in our estimation (Dimson

¹⁵ Note that the embedding dimensions are not interpretable by humans. Each dimension attempts to capture semantic information of a given LIQ+CAP section which the *Doc2Vec* model learned during the training process.

¹⁶ In line with previous research, we rely on weekly returns as it increases variance R^2 s compared to the utilization of monthly data and thus, should result in more accurate crash risk measures. However, our results remain robust when following Callen and Fang (2013, 2015a, b) in measuring crash risk who use daily instead of weekly returns.

1979), we augment the index model with lead and lag terms for market and industry returns.¹⁷ The firm-specific weekly return for firm i and week τ , $W_{i\tau}$, is the natural logarithm of one plus the residual of the equation above. For the selection of crash risk measures, we follow the existing literature and calculate the following four measures (Al Mamun et al. 2021; Chen et al. 2001; Hong et al. 2017; Jin and Myers 2006; Kim et al. 2011, 2019; Kim and Zhang 2016). First, the negative skewness of weekly stock returns:

$$NCSKEW_{it} = - \left[n(n-1)^{\frac{3}{2}} \sum_{\tau=1}^n W_{i\tau}^3 \right] / \left[(n-1)(n-2) \left(\sum_{\tau=1}^n W_{i\tau}^2 \right)^{\frac{3}{2}} \right]$$

$NCSKEW_{it}$ is calculated by dividing the negative of the third moment of firm-specific weekly returns by the standard deviation of firm-specific weekly returns raised to the third power over all n weeks of the fiscal year. Through scaling the raw third moment by the standard deviation cubed we are able to compare firms with different stock price variances (Chen et al. 2001). By putting a minus sign in front of the nominator, high values of $NCSKEW_{it}$ correspond with a more left-skewed distribution and an increased stock price crash risk for firm i in year t . Hence, we adopt the common convention that more left-skewed return distributions are associated with frequent small gains but bear the risk of extreme negative outliers (i.e. stock price crashes).

As measures based on third moments are potentially overly affected by extreme observations, we apply the “down-to-up” volatility $DUVOL_{it}$ as our second crash risk variable. $DUVOL_{it}$ equals the natural logarithm of the ratio of the standard deviation of firm-specific weekly returns in *down* weeks to the standard deviation of firm-specific weekly returns in *up* weeks:

$$DUVOL_{it} = \log \left[\frac{(n_u - 1) \sum_{DOWN} W_{i\tau}^2}{(n_d - 1) \sum_{UP} W_{i\tau}^2} \right]$$

where n_u and n_d represent the number of *up* and *down* weeks defined as weeks in which the weekly return $W_{i\tau}$ exceeds or falls below the average return of a fiscal year, respectively. Based on the pattern of incremental gains and larger losses, “crash prone” firms should be characterised by higher standard deviation of firm-specific weekly returns in *down* weeks compared to *up* weeks. Thus, high values of $DUVOL_{it}$ should correspond with an increased stock price crash risk for firm i in year t .

Thirdly, we use $COUNT_{it}$ defined as the difference in frequencies between negative stock price crashes and positive upward stock price jumps in firm-specific

¹⁷ Note that the lead and lag terms barely correlate across time (i.e. <10%). However, the market ($r_{m\tau}$) and industry ($r_{j\tau}$) returns show high correlation coefficients of approximately 79%. Even though these correlations are unlikely to bias the residuals of the market-industry index model, we also estimate firm-specific weekly returns using an extended market model as in Kim et al. (2021) that excludes industry returns and instead controls for additional lead and lag terms of market returns. Our results remain robust.

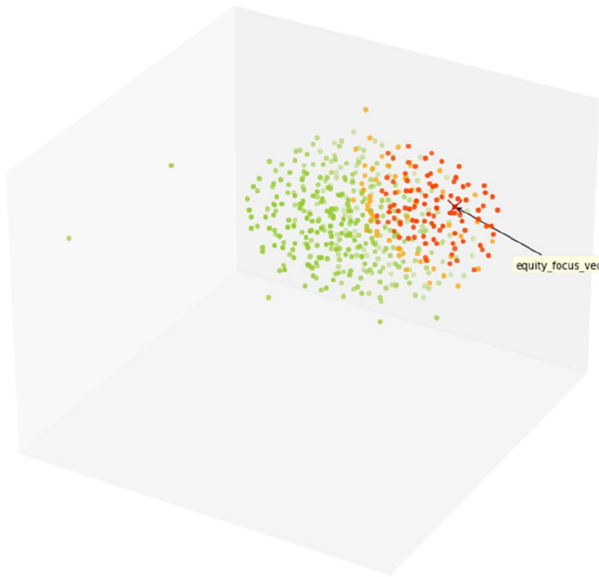


Fig. 1 Measuring managers' equity intent

returns. Stock price crashes (upward jumps) are defined as a firm-specific weekly return that falls (rises) 3.09 standard deviations below (above) the annual mean.¹⁸ Similar to prior works, we calculate $COUNT_{it}$ as the difference between stock price crashes and upward jumps (Jin and Myers 2006). Based on the same argumentation as our previous measures, skewed-return distributions, or put differently, “crash prone” firms should be characterised by more stock price crashes than upward jumps. In conclusion, higher values of $COUNT_{it}$ indicate increased stock price crash risk for firm i in year t . Finally, we include the indicator variable $CRASH_{it}$ that equals one if a firm experiences a stock price crash within a fiscal year and zero otherwise. Therefore, $CRASH_{it}$ captures the realization of a firm-specific stock price crash.

To evaluate our hypothesis that managers' equity intent increases future stock price crash risk through bad news hoarding behaviour, we regress our crash risk measures $NCSKEW_{it+1}$, $DUVOL_{it+1}$, $COUNT_{it+1}$, and $CRASH_{it+1}$ on our text-based equity intent variable $EQUITY_INTENT_{it}$ as well as a set of control variables:

$$CRASHRISK_{it+1} = \beta_0 + \beta_1 EQUITY_INTENT_{it} + \sum \beta_k CONTROL S_{it}^k + \varepsilon_{it}$$

The set of control variables includes the non-linear relationship between financial opacity, measured as the three years moving sum of discretionary accruals ($OPAQUE_{it}$) and its squared term, since Hutton et al. (2009) document a concave relation between financial opacity and crash risk. Firm size, measured as the natural

¹⁸ The threshold of 3.09 standard deviations reflects the critical value for 0.1% of the distribution of weekly returns.

Table 2 Descriptive statistics

Panel A: descriptive statistics

Variable	N	Min	Q1	Mean	Median	Q3	Max	SD
<i>Crash risk measures</i>								
$NCSKEW_{it+1}$	28,382	- 2.194	- 0.425	0.128	0.030	0.564	3.339	0.943
$DUVOL_{it+1}$	28,382	- 0.954	- 0.287	0.022	- 0.035	0.243	1.870	0.484
$COUNT_{it+1}$	28,382	- 1.000	0.000	- 0.001	0.000	0.000	2.000	0.698
$CRASH_{it+1}$	28,382	0.000	0.000	0.234	0.000	0.000	1.000	0.423
<i>Independent Variables</i>								
$EQUITY_INTENT_{it}$	28,382	0.147	0.223	0.281	0.264	0.322	0.580	0.083
$OPAQUE_{it}$	28,382	0.019	0.123	0.408	0.234	0.436	2.014	0.596
$LOGMV_{it}$	28,382	6.919	11.330	12.908	12.981	14.457	18.475	2.231
MTB_{it}	28,382	- 54.895	1.225	3.344	2.250	4.127	64.783	9.810
LEV_{it}	28,382	0.019	0.270	0.599	0.462	0.645	11.945	1.059
ROA_{it}	28,382	- 1.419	- 0.029	- 0.004	0.070	0.138	0.548	0.303
$DTURN_{it}$	28,382	- 0.515	- 0.018	0.001	0.003	0.003	0.494	0.112
$NCSKEW_{it}$	28,382	- 2.217	- 0.431	0.113	0.019	0.540	3.237	0.930
$SIGMA_{it}$	28,382	0.002	0.031	0.059	0.049	0.076	0.251	0.044
RET_{it}	28,382	- 3.165	- 0.281	- 0.275	- 0.116	- 0.048	- 0.000	0.466
$MODFOG_{it}$	28,382	9.825	12.358	13.280	13.248	14.147	17.534	1.379
$NEGW_{it}$	28,382	0.002	0.009	0.013	0.012	0.015	0.028	0.005

Panel B: Univariate comparisons

<i>Crash Risk Measures</i>	$EQUITY_INTENT_{it}$ tercile group			p- value:	(3)-(1)
	(1) Low	(2) Mid	(3) High		
$NCSKEW_{it+1}$	0.073	0.097	0.215		0.000
$DUVOL_{it+1}$	- 0.009	0.001	0.072		0.000
$COUNT_{it+1}$	- 0.004	- 0.019	0.021		0.016
$CRASH_{it+1}$	0.226	0.222	0.252		0.000

Panel A presents the descriptive statistics on crash risk, equity intent, and controls. The crash variables cover the period 1995–2019, while our equity intent measure and the control variables cover the period 1994–2018. All variables are winsorized at the top and bottom 1%. In Panel B, we sort our sample into three groups by the equity intent measure ($EQUITY_INTENT_{it}$), report the average one-year-ahead crash risk for each group, and test the difference in crash risk between the high- and low equity intent groups

logarithm of a firm’s market capitalization ($LOGMV_{it}$), controls for various aspects of a firm’s operational and business environment. Prior works show that growth firms, captured by the market-to-book ratio (MTB_{it}), are more crash-prone (Hutton et al. 2009; Kim et al. 2019), whereas firms that are less likely to experience stock price crashes should be more capable of obtaining debt (LEV_{it}). Since Graham et al. (2005) suggests that poorly performing firms are more likely to withhold bad news, we control for operating performance (ROA_{it}) and the mean of firm-specific weekly returns during a fiscal year (RET_{it}). Chen et al. (2001) find that investor heterogeneity, measured as the detrended level of turnover ($DTURN_{it}$), is positively

associated with crash risk. Finally, firm-specific skewness ($NCSKEW_{it}$) and volatility ($SIGMA_{it}$) is documented to correlate with crash risk (Kim et al. 2019; Wu and Lai 2020). All of our regressions also include year and industry fixed effects (based 2-digit SIC industry classification) to account for year- and industry-wide variation in crash risk patterns. The reported standard errors are robust to heteroscedasticity (MacKinnon and White 1985).

4 Results

4.1 Univariate statistics

Panel A of Table 2 contains descriptive statistics for our main variables $NCSKEW_{it+1}$, $DUVOL_{it+1}$, $COUNT_{it+1}$, $CRASH_{it+1}$, and $EQUITY_INTENT_{it}$, as well as our additional control variables. The average cosine similarity between an individual firms' equity intent and the equity intent vector equals 0.281. The mean values (medians) of the variables $NCSKEW_{it+1}$ and $DUVOL_{it+1}$ of 0.128 (0.030) and 0.022 (− 0.035), respectively, indicate that the distribution of weekly returns of the firms in our sample exhibits a pronounced negative skewness. Consistently, we find that, on average, 23.4% of the observations experience a stock price crash as indicated by the mean value of $CRASH_{it+1}$ which is generally consistent with the descriptive statistics reported by Kim et al. (2019). Interestingly for $COUNT_{it+1}$ we find a marginally negative mean of − 0.001 which indicates that we observe slightly more stock price jumps than crashes. But the direct comparison with the other two measures, demonstrates that if a crash happens it tends to be larger than a corresponding jump and thus detrimental for investor welfare.¹⁹

As a first step of our analysis, we use univariate comparisons between equity intent and our crash risk measures. Therefore, we sort our sample into tercile groups by the equity intent variable $EQUITY_INTENT_{it}$ and present the mean values of the one-year-ahead crash risk measures for each group in Panel B of Table 2. The results demonstrate that firms with low equity intent are more likely to experience upward stock price jumps rather than crashes as indicated by the negative means for two of our dependent variables. More importantly, we find that the stock price crash risk increases monotonically from the low-equity intent group to the high-equity intent group when we measure crash risk by $NCSKEW_{it+1}$ and $DUVOL_{it+1}$. Furthermore, the differences between the high- and low-equity intent group are statistically significant at the 1%-level. For $COUNT_{it+1}$, we find that crash risk decreases from low- to medium-equity intent only to peak for the high-equity intent group. The differences

¹⁹ We recognise that $COUNT_{it+1}$ exhibits only small variations within our sample since it only represents extreme values of the return distribution. In theory, this may lead to lower reliability and fit of OLS regression models. To account for this, we recode $COUNT_{it+1}$ to positive values by adding a constant equal to the minimum sample value of $COUNT_{it+1}$ to each observation and re-estimated all our models using Poisson regressions. We find that the results are qualitatively the same. For this reason, and to maintain the comparability with the existing literature, we report the results of the OLS regressions below. All other results are available on request.

in crash risk between the high- and low-equity intent groups are statistically significant at the 5%-level. Similarly, turning to the $CRASH_{it+1}$ measure, we find that 22.6% of the observations in the low-equity intent group experience at least one stock price crash during a fiscal year, whereas for the high-equity intent group, 25.2% of the observations experience at least one stock price crash risk during a fiscal year. The difference in means between the low- and high-equity intent group is statistically significant on the 1%-level. However, to get a deeper understanding how equity intent affects the level of crash risk, we need to consider other factors influencing crash risk within a multivariate analysis.

4.2 Main results

Table 3 shows the results of our regressions to predict one-year-ahead crash risk. The regression models differ only with respect to the dependent variables, so that column 1 shows the results for $NCSKEW_{it+1}$, column 2 for $DUVOL_{it+1}$ and column 3 for the results for $COUNT_{it+1}$. We find that the coefficient of $EQUITY_INTENT_{it}$ is positive and statistically significant on the 1%-level for all four crash risk measures, $NCSKEW_{it+1}$, $DUVOL_{it+1}$, $COUNT_{it+1}$, and $CRASH_{it+1}$. These results are in line with our hypotheses that a stronger equity intent helps to predict future stock price crash risk as it may represent a motive for bad news hoarding. Consistent with the extant literature on the relationship between earnings management and crash risk (Hutton et al. 2009) we find the coefficients for $OPAQUE_{it}$ ($OPAQUE_{it}^2$) to be positive (negative) and statistically significant at 1%-level or 5%-level in all model specifications. Even more importantly these results demonstrate that a firm's equity intent predicts crash risk in $t + 1$ after controlling for the level of earnings management and insofar is a potent predictor for bad news hoarding that transcends the effects of earnings manipulation. In line with expectations, the R-squared of the regression models for $COUNT_{it+1}$ and $CRASH_{it+1}$ indicate a weaker model fit due to the relatively small variations in these variables over the whole sample period since both only contain information on tail events of the return distribution. Moreover, this might also contribute to less statistical significance for some of the control variables because of relatively higher standard errors. However, the overall model fit of our models is consistent with results reported in prior literature on stock price crash risk which is reflective of the complex nature of predictions models for rare events such as stock price crashes. We also estimate the economic significance of $EQUITY_INTENT_{it}$ on the realization on a one-year-ahead crash event as captured by $CRASH_{it+1}$. Therefore, we use a margins model to estimate partial derivatives of the regression equation with respect to each variable in the model for each unit in the data. We find that a one standard deviation increase in $EQUITY_INTENT_{it}$ is associated with an 0.80% increased probability of a one-year-ahead crash.²⁰ This effect is comparable to other determinants of crash risk identified by prior research. For instance, Kim et al.

²⁰ We also calculate standardized coefficients for all OLS estimates of $EQUITY_INTENT_{it}$. We find that a one standard deviation increase in equity intent exerts a fairly consistent positive influence on future crash risk for all OLS models ($NCSKEW_{it+1}$: 0.036; $DUVOL_{it+1}$: 0.043; $COUNT_{it+1}$: 0.025).

(2019) show that a one standard deviation in their modified readability measure is associated with a 0.52% increased crash probability, whereas Hutton et al. (2009) show that a one standard deviation increase in $OPAQUE_{it}$ is associated with a 1.73% increase in the probability of a future crash event. In summary, we consider the effect of $EQUITY_INTENT_{it}$ to be economically significant. Therefore, our findings may help investors secure their investment performance against stock price crashes.

Up to this point, our results present a positive association between a firm's equity intent and its crash risk. To further examine the role of equity intent as a motive for bad news hoarding, we conduct a mediation analysis. In this analysis we assess the extent to which the effect of equity intent could be explained by another mediator variable that is likely to capture bad news hoarding by managers. In this regard, we assume that managers who try to block the flow of negative information to the public, intentionally or unintentionally use less negative language to explain their assessment of their firm. Therefore, we construct the mediation variable $NEGW_{it}$ that represents the fraction of negative words used by firm managers within the MD&A section. We estimate the fraction of non-negated negative words within the MD&A section by using the Fin-Neg wordlist proposed by Loughran and McDonald (2011).²¹ Following the recommendations of Baron and Kenny (1986), the subsequent mediation analysis proceeds in three steps. Firstly, as documented under "Step 1" in Table 4, we regress the mediator $NEGW_{it}$ on our main independent variable $EQUITY_INTENT_{it}$. The results show that more equity intent of a firm is associated with significantly (at the 1%-level) less negative words within its MD&A section, indicating that managers who are seeking equity allow less negative news flow. Secondly, we regress the crash risk measures on $EQUITY_INTENT_{it}$ as shown in "Step 2" of Table 4. Hence, the second step simply repeats our baseline analysis and shows that higher equity intent corresponds to higher crash risk. Thirdly, we regress the crash risk measures on both our main independent variable $EQUITY_INTENT_{it}$ and on the mediator $NEGW_{it}$. We find that while the coefficient of $EQUITY_INTENT_{it}$ remains positive and statistically significant, negative news flow helps mitigating future crash risk as indicated by the negative and statistically significant coefficient of $NEGW_{it}$.²² Finally, we estimate the average causal mediation effect (ACME). The ACME stands for the indirect effect of equity intent on crash risk that goes through negative language of managers assessed negative wording in the MD&A section of a firm's 10-K report. For the calculation of the ACME we use 1000 simulations and non-parametric bootstrapped confidence intervals. The results are presented in columns (2)–(4). The ACME is statistically significant at the 1%-level. This finding

²¹ We rely on the level of word counts since Loughran and McDonald (2011) note that differencing methods rely on the assumption that investors can remember the frequency of negative words used in 10-Ks that were filed one year ago. Moreover, their results suggest that much of the variation in differences is likely to be driven by random variation in the frequency of commonly used words.

²² Note that consistent with the prediction of Baron and Kenny (1986), we find that in "Step 3" of Table 1, the coefficients of $EQUITY_INTENT_{it}$ are slightly smaller as compared to the regression in "Step 2". Baron and Kenny (1986) argue that it is critical to examine not only the significance of the coefficients but also their absolute magnitude. Since the independent variable and the mediator should be correlated one would expect reduced power in the test of the coefficients in "Step 3" as shown by our results.

Table 3 Impact of equity intent on stock price crash risk

	Dependent variable			
	<i>NCSKEW</i> _{<i>it</i>+1}	<i>DUVOL</i> _{<i>it</i>+1}	<i>COUNT</i> _{<i>it</i>+1}	<i>CRASH</i> _{<i>it</i>+1}
<i>EQUITY_INTENT</i> _{<i>it</i>}	0.412*** (5.77)	0.254*** (7.28)	0.211*** (3.73)	0.545*** (2.87)
<i>OPAQUE</i> _{<i>it</i>}	0.269*** (9.20)	0.176*** (12.13)	0.104*** (4.53)	0.161** (2.04)
<i>OPAQUE</i> _{<i>it</i>} ²	- 0.056*** (- 6.57)	- 0.038*** (- 8.82)	- 0.019*** (- 2.86)	- 0.024 (- 1.01)
<i>LOGMV</i> _{<i>it</i>}	0.000 (- 0.00)	- 0.008*** (- 4.61)	0.019*** (7.24)	0.046*** (5.14)
<i>MTB</i> _{<i>it</i>}	0.002*** (2.97)	0.001*** (4.12)	0.001 (1.47)	0.001 (0.63)
<i>LEV</i> _{<i>it</i>}	- 0.015 (- 1.55)	- 0.015*** (- 3.31)	- 0.002 (- 0.37)	- 0.004 (- 0.13)
<i>ROA</i> _{<i>it</i>}	- 0.087*** (- 3.81)	- 0.068 *** (- 6.00)	0.027 (1.49)	0.313*** (5.04)
<i>DTURN</i> _{<i>it</i>}	0.108** (2.26)	0.093*** (3.97)	0.049 (1.35)	- 0.050 (- 0.38)
<i>NCSKEW</i> _{<i>it</i>}	0.157*** (23.57)	0.099*** (30.54)	0.061*** (12.32)	0.124*** (7.69)
<i>SIGMA</i> _{<i>it</i>}	- 11.230*** (- 27.53)	- 9.244*** (- 38.57)	- 3.292*** (- 9.75)	0.150 (0.13)
<i>RET</i> _{<i>it</i>}	- 0.870*** (- 22.82)	- 0.623*** (- 30.79)	- 0.194*** (- 6.60)	- 0.083 (0.74)
Year FE	YES	YES	YES	YES
Industry FE	YES	YES	YES	YES
Observations	28,382	28,382	28,382	28,382
Adjusted/Pseudo R ²	0.129	0.222	0.040	0.024

This table presents the results for the OLS regressions of *NCSKEW*_{*it*+1}, *DUVOL*_{*it*+1} and *COUNT*_{*it*+1}, and the logistic regression of *CRASH*_{*it*+1} on our equity intent measure *EQUITY_INTENT*_{*it*} for the time period from 1994 to 2018. The variable *NCSKEW*_{*it*+1} is the negative skewness of firm-specific weekly returns over fiscal year *t* + 1; *DUVOL*_{*it*+1} is the natural logarithm of the ratio of the standard deviation of firm-specific weekly returns on down weeks to the standard deviation of firm-specific weekly returns on up weeks, where the down and up weeks are those with firm-specific weekly returns below and above the mean over fiscal year *t* + 1, respectively; *COUNT*_{*it*+1} is the frequency differences between stock price crashes and upward stock price jumps in fiscal year *t* + 1; *CRASH*_{*it*+1} is an indicator variable that equals 1 if a stock price crash occurred in in fiscal year *t* + 1 and 0 otherwise. We define a fiscal year as the 12-month period ending three months after the official fiscal year-end to avoid look-ahead bias. The variable *EQUITY_INTENT*_{*it*} is the cosine similarity between a given firms' LIQ+CAP embedding to the equity focus vector that is defined as the average of all LIQ+CAP embeddings in which managers explicitly express their equity intent. All models also include an unreported intercept. The t- and z-statistics reported in parentheses are based on robust standard errors. *, **, and *** represent significance at the 10%, 5%, and 1% levels, respectively

indicates that, a significant part of the effect of equity intent on crash risk can be attributed to bad news hoarding through less negative language used by firm managers in their MD&A section. In line with the predictions of Jin and Myers (2006) as well as the vast majority of empirical research on crash risk, these findings indicate

Table 4 Bad news hoarding—mediating effect of negative wording in the MD&A section

	Step 1		Step 2				Step 3			
	<i>NEGW_{it}</i>		<i>NCSKEW_{it+1}</i>	<i>DUVOL_{it+1}</i>	<i>COUNT_{it+1}</i>	<i>CRASH_{it+1}</i>	<i>NCSKEW_{it+1}</i>	<i>DUVOL_{it+1}</i>	<i>COUNT_{it+1}</i>	<i>CRASH_{it+1}</i>
<i>EQUITY_INTENT_{it}</i>	-0.008*** (-21.48)		0.412*** (5.77)	0.254*** (7.28)	0.211*** (3.73)	0.545*** (2.87)	0.371*** (5.17)	0.231*** (6.61)	0.187*** (3.27)	0.447** (2.35)
<i>NEGW_{it}</i>	-	-	-	-	-	-	-5.259*** (-4.63)	-2.857*** (-5.23)	-3.140*** (-3.49)	-6.532*** (-2.00)
ACME	-	-	-	-	-	-	0.041***	0.022***	0.025***	0.009***
Controls	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Year FE	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Industry FE	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Observations	28,382	28,382	28,382	28,382	28,382	28,382	28,382	28,382	28,382	28,382
Adjusted/Pseudo R ²	0.124	0.129	0.222	0.222	0.040	0.024	0.127	0.223	0.039	0.024

This table presents the results for the mediation analysis as proposed by Baron and Kelly (1986). In “Step 1”, we regress the mediator $NEGW_{it}$, which is defined as the fraction of negative words in a given MD&A section, on our main independent variable $EQUITY_INTENT_{it}$. The variable $EQUITY_INTENT_{it}$ is the cosine similarity between a given firms’ LIQ+CAP embedding to the equity focus vector that is defined as the average of all LIQ+CAP embeddings in which managers explicitly express their equity intent. In “Step 2”, we run OLS regressions of $NCSKEW_{it+1}$, $DUVOL_{it+1}$ and $COUNT_{it+1}$, and the logistic regression of $CRASH_{it+1}$ on our equity intent measure $EQUITY_INTENT_{it}$. The variable $NCSKEW_{it+1}$ is the negative skewness of firm-specific weekly returns over fiscal year $t + 1$; $DUVOL_{it+1}$ is the natural logarithm of the ratio of the standard deviation of firm-specific weekly returns on down weeks to the standard deviation of firm-specific weekly returns on up weeks, where the down and up weeks are those with firm-specific weekly returns below and above the mean over fiscal year $t + 1$, respectively; $COUNT_{it+1}$ is the frequency differences between stock price crashes and upward stock price jumps in fiscal year $t + 1$; $CRASH_{it+1}$ is an indicator variable that equals 1 if a stock price crash occurred in in fiscal year $t + 1$ and 0 otherwise. We define a fiscal year as the 12-month period ending three months after the official fiscal year-end to avoid look-ahead bias. In “Step 3”, we also add the mediator $NEGW_{it}$ to the regression model. ACME denotes the average causal mediation effect of $EQUITY_INTENT_{it}$ that goes through $NEGW_{it}$. The causal mediation analysis is performed using 1,000 simulations and non-parametric bootstrapped confidence intervals. All models also include an unreported intercept. The t- and z-statistics reported in parentheses are based on robust standard errors. *, **, and *** represent significance at the 10%, 5%, and 1% levels, respectively

that managers' bad news hoarding behavior serves an important economic mechanism through which managers' equity intent affect future crash risk.

Our paper illustrates that a firms equity intent acts a motive for bad news hoarding. Given that managers actively engage in earnings manipulation (Hutton et al. 2009) as well as textual obfuscation to hide negative news (Kim et al. 2019) one could naturally ask whether these instruments of bad news hoarding are more likely to increase crash risk when managers are incentivised to hoard bad news due to their equity intent. Consistent with the literature, we use the FOG Index to assess the textual obfuscation based on the readability of the 10-K report. We calculate the FOG Index as $(\text{words per sentence} + \text{percentage of complex words}) \times 0.4$ with a higher FOG Index indicating that the report is more difficult to read, where complex words are defined as words exceeding three syllables. The FOG Index indicates the number of years of formal education a reader of average intelligence needs to understand a given text on the first reading. We follow the recommendations of Kim et al. (2019) and adjust the FOG Index for more than 2000 multisyllabic words that occur frequently in financial reports but are unlikely to be considered complex from an investor's perspective. After modifying the FOG Index, we derive $MODFOG_{it}$ as proposed by Kim et al. (2019). If a firms equity intent reinforces the effects of earnings manipulation and textual obfuscation on crash risk, we would expect that $OPAQUE_{it}$ and $MODFOG_{it}$ have a stronger predictive power when managers have equity intent. Hence, we examine the effects of earnings manipulation and textual obfuscation for two different groups based on their equity intent. Similar to the univariate comparison in Table 2, we conduct a split sample analysis by comparing firms within the lower tercile (low equity intent) and the upper tercile (high equity intent) group of $EQUITY_INTENT_{it}$. In Panel A of Table 5 the results for the low-equity intent group indicate that textual obfuscation is solely significant at the 5%-level for our second crash risk measure $DUVOL_{it+1}$. Moreover, $OPAQUE_{it}$ significantly predicts crash risk in $t + 1$ for $NCSKEW_{it+1}$ (on the 5%-level) and $DUVOL_{it+1}$ (on the 1%-level) and $CRASH_{it+1}$ (on the 5%-level only for the squared term of $OPAQUE_{it}$). In line with our expectations, the results for the high-equity intent group presented in Panel B of Table 5 demonstrate that the effects of both instruments of bad news hoarding (earnings manipulation and textual obfuscation) increase in effect size and statistical significance. Thus, textual obfuscation increases crash risk in $t + 1$ at least at the 5%-level for all crash risk measures. Similarly, earnings manipulation also increases crash risk in $t + 1$ at least on the 5%-level for all measures of crash risk.²³ Overall, these findings indicate that managers' equity intent increases the adverse consequences of common bad news hoarding tool which further strengthens our view that managers equity intent encourages managers to hide adverse information.

²³ In addition, we find larger and more significant effects of most control variables for firms with high equity intent. Thus, in line with expectations, growth firms are even more crash prone if their managers have a strong incentive to hide adverse information, e.g., have a high equity intent. For brevity, we do not report the results for the control variables.

Table 5 Impact of equity intent on the association between earnings manipulation, textual obfuscation and stock price crash risk

	Dependent variable			
	$NCSKEW_{it+1}$	$DUVOL_{it+1}$	$COUNT_{it+1}$	$CRASH_{it+1}$
<i>Panel A: Low equity intent</i>				
$MODFOG_{it}$	0.011 (1.50)	0.008** (2.40)	0.002 (0.36)	0.022 (1.01)
$OPAQUE_{it}$	0.133** (2.31)	0.080*** (2.94)	0.034 (0.72)	- 0.182 (- 1.07)
$OPAQUE_{it}^2$	- 0.001 (- 0.49)	- 0.011 (- 1.23)	0.014 (0.81)	0.119** (2.01)
Controls	YES	YES	YES	YES
Year FE	YES	YES	YES	YES
Industry FE	YES	YES	YES	YES
Observations	9461	9461	9461	9,461
Adjusted/Pseudo R ²	0.045	0.081	0.019	0.029
<i>Panel B: High equity intent</i>				
$MODFOG_{it}$	0.042*** (6.08)	0.036*** (8.96)	0.017*** (2.64)	0.054** (2.51)
$OPAQUE_{it}$	0.276*** (6.26)	0.187*** (8.62)	0.130*** (3.69)	0.255** (2.40)
$OPAQUE_{it}^2$	- 0.061*** (- 4.97)	- 0.042*** (- 6.76)	- 0.028*** (- 2.88)	- 0.072** (- 2.52)
Controls	YES	YES	YES	YES
Year FE	YES	YES	YES	YES
Industry FE	YES	YES	YES	YES
Observations	9,461	9,461	9,461	9,461
Adjusted/Pseudo R ²	0.240	0.387	0.073	0.033

This table presents the results for the OLS regressions of $NCSKEW_{it+1}$, $DUVOL_{it+1}$ and $COUNT_{it+1}$, and the logistic regression of $CRASH_{it+1}$ on the readability measure $MODFOG_{it}$ and the earnings manipulation measures $OPAQUE_{it}$ and $OPAQUE_{it}^2$ in the time period from 1994 to 2018. The variable $NCSKEW_{it+1}$ is the negative skewness of firm-specific weekly returns over fiscal year $t + 1$; $DUVOL_{it+1}$ is the natural logarithm of the ratio of the standard deviation of firm-specific weekly returns on down weeks to the standard deviation of firm-specific weekly returns on up weeks, where the down and up weeks are those with firm-specific weekly returns below and above the mean over fiscal year $t + 1$, respectively; $COUNT_{it+1}$ is the frequency differences between stock price crashes and upward stock price jumps in fiscal year $t + 1$; $CRASH_{it+1}$ is an indicator variable that equals 1 if a stock price crash occurred in in fiscal year $t + 1$ and 0 otherwise. We define a fiscal year as the 12-month period ending three months after the official fiscal year-end to avoid look-ahead bias. The variable $MODFOG_{it}$ is the modified FOG Index proposed by Kim et al. (2019). The variable $OPAQUE_{it}$ is calculated as the moving sum of the absolute value of abnormal accruals in the prior three years, where abnormal accruals are estimated using the modified Jones model. The variable $OPAQUE_{it}^2$ represents the squared term of $OPAQUE_{it}$ which captures potential non-linearities of firms' earnings management on stock price crash risk. We split the sample by our equity intent measure $EQUITY_INTENT_{it}$, where each firm is included into the low (high) equity intent subsample if its equity intent falls into the first (last) tercile of the equity intent distribution of the whole sample. Panel A presents the regression results for the low equity intent group and panel B presents the regression results for the high equity intent group. All models also include an unreported intercept. The t- and z-statistics reported in parentheses are based on robust standard errors. *, **, and *** represent significance at the 10%, 5%, and 1% levels, respectively

5 Additional analyses

In order to test the sensitivity of our results, we conduct several robustness tests. Firstly, because investors that try to hedge against future crash risk might be interested in the persistence with which equity intent signals bad news hoarding, we expand the prediction window of all four crash risk measures. While our baseline analysis employs a 12-month ahead prediction window, managers may accumulate bad news over several years. Thus, we test the relationship between managers' equity intent and the 24-, 36-, and 48-month ahead stock price crash risk. The results are displayed in Table 6. We find that the coefficients of $EQUITY_INTENT_{it}$ remain statistically significant at least on the 5%-level for all four crash variables when employing a 24-month window, and even remain significant on at least 5% for $NCSKEW_{it+3}$ and $DUVOL_{it+3}$ when employing a 36-month window. Even for a 48-month prediction window, the coefficients of $EQUITY_INTENT_{it}$ remain statistically significant on the 10%-level when predicting $DUVOL_{it+4}$. Consistent with the notion that bad news hoarding becomes more difficult to maintain over the course of time the coefficients and their respective t-values of $EQUITY_INTENT_{it}$ steadily decreases with an increasing prediction window, similar the models' R-squared. Given the decreasing importance of $EQUITY_INTENT_{it}$ over time, our results suggest that investors should screen 10-K filings on a yearly basis to obtain better predictions of future crash risk.

Secondly, we examine the association between managers' equity intent and upward stock price jumps. Our results generally suggest that managers' equity intent is associated with increasing bad news hoarding behaviour. While a gradual increase of the stock price as a result of managers "hyping up" the firm's stock price prior to an SEO seems plausible, sudden upward price jumps seem comparably unlikely, as they would indicate positive market reaction following the revelation of unexpected (good) news. In this case, one would expect that $EQUITY_INTENT_{it}$ is associated with future stock price crashes as a consequence of the sudden release of the accumulated negative information, but should either have no, or a negative correlation with upward stock price jumps. Therefore, we decompose the crash risk measure $COUNT_{it+1}$ into stock price crashes and upward jumps and estimate both proxies separately. Unreported results show that $EQUITY_INTENT_{it}$ is only positively associated with stock price crashes and shows a negative correlation with upward jumps, that is statistically significant only at the 10%-level. Similar to the conception of a placebo test, the theoretical consistency of these results suggest that our inferences are unlikely to be driven by unobservable forces.

Thirdly, we construct an alternative measure for our main variable of interest $EQUITY_INTENT_{it}$. Given the usage of neural networks, we acknowledge that the variable is subject to a large parameter space and uncountable interactions between neurons. While our baseline analysis employs the DM model, we next estimate word embeddings using the DBOV model proposed by Le and Mikolov (2014). The DBOV model is generally considered to be less complex compared to the DM model as it only uses the document matrix to predict a randomly sampled set of consecutive words from a document. For the implementation of the DBOV model, we follow the recommendations of Lau and Baldwin (2016) and estimate 300-dimensional document embeddings in 400 iterations over the whole sample of LIQ+CAP sections.²⁴ Similar to our baseline analysis, we use these document embeddings to estimate an equity intent vector and calculate cosine similarities from each LIQ+CAP embeddings to the equity intent vector to derive $EQUITY_INTENT_DBOV_{it}$. Table 7 shows the results of estimating our crash risk measures using $EQUITY_INTENT_DBOV_{it}$. The results remain virtually unchanged, indicating that our inferences are not merely driven by model and parameter choices.

Fourthly, we include additional control variables to address concerns of omitted variable bias.²⁵ Specifically, we address concerns that managers' equity intent is merely a reflection of a firm's risk profile by controlling for additional variables capturing business risks including a firm's cash flow volatility $CFVOL_{it}$, sales volatility $SALESVOL_{it}$, earnings volatility $EARNVOL_{it}$ as well as the Herfindahl–Hirschman Index based on 3-digit SIC codes HHI_{it} (Giroud and Mueller 2010; Kim et al. 2019). Since our main variable of interest is based on textual data, we also control for textual proxies that prior works found to be significantly associated with crash risk. These include the modified FOG Index $MODFOG_{it}$ proposed by Kim et al. (2019), managers' usage of ambiguous language measured by the fraction of weak modal words within the MD&A $WMOD_{it}$ and the natural logarithm of a 10-K's file size in megabytes $FILESIZE_{it}$ as in Ertugrul et al. (2017).²⁶ Finally, based on the findings of Wu and Lai (2020), we include two additional measures for firm's inherent information asymmetry, namely, intangible intensity $ADJROTA_{it}$ as suggested by Clausen and Hirth (2016) as well as a firm's proprietary costs $PROP_COST_{it}$ measured as a firm's R&D expense scaled by total assets and firm age (AGE_{it}). The results displayed in Table 8 show that the coefficient of $EQUITY_INTENT_{it}$ remains statistically significant for all four crash measures, alleviating concerns of omitted variable bias.²⁷

²⁴ Following Lau and Baldwin (2016), we use a vector size of $d = 300$, a window size of 15, use a down-sampling threshold of $1e^{-5}$, draw 5 "noise words" through negative sampling. Given the size of our dataset we also ignore words accruing less than 5 times.

²⁵ Descriptive statistics for the additional variables are displayed in Table 1A.

²⁶ While Ertugrul et al. (2017) and Kim et al. (2019) estimate the fraction of ambiguous language as well as the FOG Index based on the full 10-K file, we determinately measure the variables based on the MD&A section as this section is more likely to capture characteristics of firm's management. However, constructing measures based on the full 10-K does not alter our results.

²⁷ Note that our results also remain robust when using $EQUITY_INTENT_DBOV_{it}$ instead.

Table 6 Persistence of the impact of equity intent on stock price crash risk

	Dependent variable		
	t+2	t+3	t+4
<i>Panel A: NCSKEW</i>			
<i>EQUITY_INTENT_{it}</i>	0.292*** (3.95)	0.157** (2.05)	0.095 (1.24)
Controls	YES	YES	YES
Year FE	YES	YES	YES
Industry FE	YES	YES	YES
Observations	28,370	28,329	28,295
Adjusted/Pseudo R ²	0.067	0.041	0.025
<i>Panel B: DUVOL</i>			
<i>EQUITY_INTENT_{it}</i>	0.169*** (4.56)	0.106*** (2.75)	0.070* (1.76)
Controls	YES	YES	YES
Year FE	YES	YES	YES
Industry FE	YES	YES	YES
Observations	28,370	28,329	28,295
Adjusted/Pseudo R ²	0.117	0.067	0.040
<i>Panel C: COUNT</i>			
<i>EQUITY_INTENT_{it}</i>	0.132** (2.34)	0.076 (1.35)	0.048 (0.84)
Controls	YES	YES	YES
Year FE	YES	YES	YES
Industry FE	YES	YES	YES
Observations	28,370	28,329	28,295
Adjusted/Pseudo R ²	0.022	0.013	0.008
<i>Panel D: CRASH</i>			
<i>EQUITY_INTENT_{it}</i>	0.567** (2.92)	0.333 (1.61)	0.104 (0.47)
Controls	YES	YES	YES
Year FE	YES	YES	YES
Industry FE	YES	YES	YES
Observations	28,370	28,329	28,295
Adjusted/Pseudo R ²	0.019	0.016	0.015

This table presents the results for the OLS regressions of $NCSKEW_{it+2:4}$ (Panel A), $DUVOL_{it+2:4}$ (Panel B) and $COUNT_{it+2:4}$ (Panel C), and the logistic regression of $CRASH_{it+2:4}$ (Panel D) on our equity intent measure $EQUITY_INTENT_{it}$ for the time period from 1994 to 2018. The variable $NCSKEW_{it+2:4}$ is the negative skewness of firm-specific weekly returns in fiscal years $t+2$ to $t+4$; $DUVOL_{it+2:4}$ is the natural logarithm of the ratio of the standard deviation of firm-specific weekly returns on down weeks to the standard deviation of firm-specific weekly returns on up weeks, where the down and up weeks are those with firm-specific weekly returns below and above the mean in fiscal years $t+2$ to $t+4$, respectively; $COUNT_{it+2:4}$ is the frequency differences between stock price crashes and upward stock price jumps in fiscal years $t+2$ to $t+4$; $CRASH_{it+2:4}$ is an indicator variable that equals 1 if a stock price crash occurred in fiscal years $t+2$ to $t+4$ and 0 otherwise. We define a fiscal year as the 12-month period ending three months after the official fiscal year-end to avoid look-ahead bias. The variable $EQUITY_INTENT_{it}$ is the cosine similarity between a given firms' LIQ+CAP embedding to the equity focus vector that is defined as the average of all LIQ+CAP embeddings in which managers explicitly express their equity intent. All models also include an unreported intercept. The t- and z-statistics reported in parentheses are based on robust standard errors. *, **, and *** represent significance at the 10%, 5%, and 1% levels, respectively

Table 7 Alternative measurement of equity intent

	Dependent variable			
	$NCSKEW_{it+1}$	$DUVOL_{it+1}$	$COUNT_{it+1}$	$CRASH_{it+1}$
$EQUITY_INTENT_DBOW_{it}$	0.509*** (5.49)	0.318*** (7.02)	0.274*** (3.73)	0.626** (2.57)
Controls	YES	YES	YES	YES
Year FE	YES	YES	YES	YES
Industry FE	YES	YES	YES	YES
Observations	28,382	28,382	28,382	28,382
Adjusted/Pseudo R^2	0.126	0.222	0.039	0.024

This table presents the results for the OLS regressions of $NCSKEW_{it+1}$, $DUVOL_{it+1}$ and $COUNT_{it+1}$, and the logistic regression of $CRASH_{it+1}$ on our equity intent measure $EQUITY_INTENT_{it}$ for the time period from 1994 to 2018. The variable $NCSKEW_{it+1}$ is the negative skewness of firm-specific weekly returns over fiscal year $t + 1$; $DUVOL_{it+1}$ is the natural logarithm of the ratio of the standard deviation of firm-specific weekly returns on down weeks to the standard deviation of firm-specific weekly returns on up weeks, where the down and up weeks are those with firm-specific weekly returns below and above the mean over fiscal year $t + 1$, respectively; $COUNT_{it+1}$ is the frequency differences between stock price crashes and upward stock price jumps in fiscal year $t + 1$; $CRASH_{it+1}$ is an indicator variable that equals 1 if a stock price crash occurred in in fiscal year $t + 1$ and 0 otherwise. We define a fiscal year as the 12-month period ending three months after the official fiscal year-end to avoid look-ahead bias. The variable $EQUITY_INTENT_DBOW_{it}$ is the cosine similarity between a given firms' LIQ + CAP embedding to the equity focus vector that is defined as the average of all LIQ + CAP embeddings in which managers explicitly express their equity intent. In this robustness check, all document embeddings are estimated using the Distributed Bag Of Words (DBOW) implementation of *Doc2Vec*. All models also include an unreported intercept. The t- and z-statistics reported in parentheses are based on robust standard errors. *, **, and *** represent significance at the 10%, 5%, and 1% levels, respectively

Finally, we test whether earnings manipulation, a common vehicle managers use to hoard bad news (Hutton et al. 2009), serves as another mediator for the effect of equity intent on stock price crash risks. Specifically, one could argue that managers do not only block negative news flow but also engage in more aggressive earnings management to obfuscate financial disclosures. Therefore, in Table 9, we use $OPAQUE_{it}$ as a mediator variable for $EQUITY_INTENT_{it}$. “Step 1” of the mediation analysis shows that higher equity intent is associated with higher levels of earnings management, suggesting that managers who search for equity are more likely to engage in earnings management. Similarly, the results of “Step 2” and “Step 3” of the mediation analysis are largely consistent with our baseline analysis, indicating that earnings management serves as a significant mediator for effects of managers' equity intent on future crash risk. In summary, the above analyses suggest robustness of our finding that managers equity intent incentivises managers to hoard bad news and hence, increases firm-specific crash risk.

Table 8 Additional controls

	Dependent variable			
	<i>NCSKEW</i> _{<i>it</i>+1}	<i>DUVOL</i> _{<i>it</i>+1}	<i>COUNT</i> _{<i>it</i>+1}	<i>CRASH</i> _{<i>it</i>+1}
<i>EQUITY_INTENT</i> _{<i>it</i>}	0.192** (2.49)	0.121*** (3.30)	0.140** (2.25)	0.454** (2.16)
<i>OPAQUE</i> _{<i>it</i>}	0.239*** (7.33)	0.152*** (9.53)	0.074*** (2.90)	0.124 (1.39)
<i>OPAQUE</i> _{<i>it</i>} ²	- 0.057*** (- 6.07)	- 0.037*** (- 7.89)	- 0.014* (- 1.91)	- 0.041 (- 1.52)
<i>LOGMV</i> _{<i>it</i>}	0.003 (0.81)	- 0.005*** (- 2.65)	0.017*** (5.58)	0.044*** (3.99)
<i>MTB</i> _{<i>it</i>}	0.001* (1.80)	0.001*** (2.78)	0.001 (0.99)	0.001 (0.44)
<i>LEV</i> _{<i>it</i>}	- 0.022** (- 2.46)	- 0.019*** (- 4.29)	- 0.013* (- 1.65)	- 0.020 (- 0.61)
<i>ROA</i> _{<i>it</i>}	- 0.002 (- 0.06)	- 0.009 (- 0.62)	- 0.065*** (2.75)	- 0.407*** (4.91)
<i>DTURN</i> _{<i>it</i>}	0.065 (1.29)	0.067*** (2.73)	0.033 (0.83)	- 0.085 (- 0.60)
<i>NCSKEW</i> _{<i>it</i>}	0.119*** (16.28)	0.076*** (21.58)	0.047*** (8.69)	0.100*** (5.61)
<i>SIGMA</i> _{<i>it</i>}	- 11.420*** (- 21.19)	- 8.637*** (- 29.44)	- 2.910*** (- 7.14)	- 0.889 (- 0.64)
<i>RET</i> _{<i>it</i>}	- 0.826*** (- 17.06)	- 0.601*** (- 23.43)	- 0.155*** (- 4.18)	0.036 (0.25)
<i>CFVOL</i> _{<i>it</i>}	0.037*** (2.58)	0.022*** (2.91)	0.013 (1.08)	0.049 (1.50)
<i>SALESVOL</i> _{<i>it</i>}	0.007 (0.75)	0.004 (0.91)	- 0.002 (- 0.29)	- 0.014 (- 0.55)
<i>EARNVOL</i> _{<i>it</i>}	- 0.014** (- 2.03)	- 0.007** (- 2.04)	- 0.005 (- 0.91)	- 0.011 (- 0.70)
<i>HHI</i> _{<i>it</i>}	0.002 (0.06)	- 0.008 (- 0.49)	0.017 (0.56)	0.162 (1.58)
<i>MODFOG</i> _{<i>it</i>}	0.026*** (5.04)	0.016*** (6.69)	0.009** (2.32)	0.041*** (2.97)
<i>WMOD</i> _{<i>it</i>}	2.371 (0.89)	0.271 (0.22)	1.176 (0.56)	3.365 (0.45)
<i>FILESIZE</i> _{<i>it</i>}	- 0.028*** (- 2.00)	- 0.013** (- 2.03)	- 0.013 (- 1.17)	- 0.087** (- 2.21)
<i>ADJROTA</i> _{<i>it</i>}	0.011* (1.74)	0.002 (0.72)	0.013*** (2.67)	0.040** (2.25)
<i>PROP_COST</i> _{<i>it</i>}	0.150*** (3.32)	0.103*** (4.72)	0.071** (1.99)	0.149 (1.32)
<i>AGE</i> _{<i>it</i>}	- 0.003*** (- 12.37)	- 0.002*** (- 14.58)	- 0.001*** (- 5.31)	- 0.002*** (- 2.76)
Year FE	YES	YES	YES	YES
Industry FE	YES	YES	YES	YES
Observations	23,958	23,958	23,958	23,958
Adjusted/Pseudo R ²	0.092	0.167	0.032	0.025

Table 8 (continued)

This table presents the results of robust checks for the OLS regressions of $NCSKEW_{it+1}$, $DUVOL_{it+1}$ and $COUNT_{it+1}$, and the logistic regression of $CRASH_{it+1}$ on our equity intent measure $EQUITY_INTENT_{it}$ for the time period from 1994 to 2018. In addition to the set of baseline control variables the regression models include $CFVOL_{it}$, the standard deviation of operating cash flows scaled by lagged total assets over the five fiscal years from $t - 4$ to t , $SALESVOL_{it}$ the standard deviation of sales revenue scaled by lagged total assets over the five fiscal years from $t - 4$ to t , $EARNVOL_{it}$ the standard deviation of ROA scaled by lagged total assets over the five fiscal years from $t - 4$ to t , HHI_{it} the Herfindahl–Hirschman Index based on 3-digit SIC codes, $WMOD_{it}$ is the fraction of weak model words within the MD&A, $MODFOG_{it}$ is the modified FOG Index proposed by Kim et al. (2019), $FILESIZE_{it}$ the natural logarithm of the file size (in megabytes) of the MD&A filed in fiscal year t , $ADJROTA_{it}$ is a firm's intangible intensity as proposed by Clausen and Hirth (2016), $PROP_COST_{it}$ denotes a firm's proprietary costs measured as a firm's R&D expense scaled by total assets and AGE_{it} denotes the natural logarithm of a firm's age. All models also include an unreported intercept. The t - and z -statistics reported in parentheses are based on robust standard errors. *, **, and *** represent significance at the 10%, 5%, and 1% levels, respectively

6 Conclusion

Previous research has established the notion that managers are incentivised to strategically withhold negative information from outside investors in order to maximize their own capture of the firm's cash flows (Jin and Myers 2006; Kothari et al. 2009). Since the accumulation of bad news is limited when further obfuscations become too costly or difficult to maintain, all negative information is released at once, resulting in the firm's stock price to crash. Given the subsequent destruction of shareholder welfare, the motives for managers to conduct this behaviour has received increasing attention from financial research literature. While incentive-based management compensation, CEO overconfidence (Kim et al. 2016), and overall career concerns (Kothari et al. 2009) have already been identified as motives to obfuscate corporate disclosure, we turn our focus on corporate financing decisions, thus drawing the connection to arguably one of the most important areas of corporate decision making. Capital structure decisions, especially in the case of equity financing, are known for providing agency conflicts between managers and outside investors, which should incentivise managers to obfuscate corporate disclosure.

Further extending the analysis of drivers of bad news hoarding behaviour, we analyse whether the intention to issue equity incentivises managers to strategically withhold negative information to maximize corresponding capital inflow and strengthen their position of power inside the firm. We examine the role of managers' intentions within the capital structure decisions in shaping firm-specific stock price crash risks, by analysing how managers express the financing needs of their firms in the LIQ+CAP sections of their MD&A. Using neural network embeddings of the LIQ+CAP sections, we identify different degrees of equity intent. Differences in the management's inclination to raise more equity may lead to different levels of bad news hoarding which in turn may impact firm-specific stock price crash risks. We expect that managers expressing equity intent in the firms' disclosures are more prone to bad news hoarding behaviour which helps to predict future stock price crash risk of firms. In fact, we find that a firm's current equity intent has strong predictive power for firm-specific stock price crash risk in $t + 1$. As a central motive for bad news hoarding, equity intent exerts comparable effects on crash risk

Table 9 Bad News Hoarding – Mediating effect of earnings manipulation

	Step 1		Step 2		Step 3		
	<i>OPAQUE_{it}</i>	<i>NCSKEW_{it+1}</i>	<i>DUVOL_{it+1}</i>	<i>COUNT_{it+1}</i>	<i>CRASH_{it+1}</i>	<i>COUNT_{it+1}</i>	<i>CRASH_{it+1}</i>
<i>EQUITY_INTENT_{it}</i>	0.336*** (8.03)	0.440*** (6.27)	0.273*** (7.95)	0.247*** (3.82)	0.585*** (2.99)	0.211*** (3.73)	0.545*** (2.87)
<i>OPAQUE_{it}</i>	–	–	–	–	–	0.104*** (4.53)	0.161** (2.04)
ACME	–	–	–	–	–	0.063***	0.008**
Controls	YES	YES	YES	YES	YES	YES	YES
Year FE	YES	YES	YES	YES	YES	YES	YES
Industry FE	YES	YES	YES	YES	YES	YES	YES
Observations	28,382	28,382	28,382	28,382	28,382	28,382	28,382
Adjusted/Pseudo R ²	0.307	0.125	0.220	0.040	0.024	0.129	0.024

This table presents the results for the mediation analysis as proposed by Baron and Kelly (1986). In “Step 1”, we regress the mediator *OPAQUE_{it}*, which is defined as the moving sum of the absolute value of abnormal accruals in the prior three years, on our main independent variable *EQUITY_INTENT_{it}*. The variable *EQUITY_INTENT_{it}* is the cosine similarity between a given firms’ LIQ + CAP embedding to the equity focus vector that is defined as the average of all LIQ + CAP embeddings in which managers explicitly express their equity intent. In “Step 2”, we run OLS regressions of *NCSKEW_{it+1}*, *DUVOL_{it+1}* and *COUNT_{it+1}* and the logistic regression of *CRASH_{it+1}* on our equity intent measure *EQUITY_INTENT_{it}*. The variable *NCSKEW_{it+1}* is the negative skewness of firm-specific weekly returns over fiscal year *t + 1*; *DUVOL_{it+1}* is the natural logarithm of the ratio of the standard deviation of firm-specific weekly returns on down weeks to the standard deviation of firm-specific weekly returns on up weeks, where the down and up weeks are those with firm-specific weekly returns below and above the mean over fiscal year *t + 1*, respectively; *COUNT_{it+1}* is the frequency differences between stock price crashes and upward stock price jumps in fiscal year *t + 1*; *CRASH_{it+1}* is an indicator variable that equals 1 if a stock price crash occurred in fiscal year *t + 1* and 0 otherwise. We define a fiscal year as the 12-month period ending three months after the official fiscal year-end to avoid look-ahead bias. In “Step 3”, we also add the mediator *OPAQUE_{it}* to the regression model. ACME denotes the average causal mediation effect of *EQUITY_INTENT_{it}* that goes through *OPAQUE_{it}*. The causal mediation analysis is performed using 1,000 simulations and non-parametric bootstrapped confidence intervals. All models also include an unreported intercept. The t- and z-statistics reported in parentheses are based on robust standard errors. *, **, and *** represent significance at the 10%, 5%, and 1% levels, respectively

as other popular obfuscation vehicles. By conducting a mediation analysis, we find that stronger equity intent affects future stock price crash risk through the channel of bad news hoarding. Consistent with this notion, we find that managers' equity intent incentivises managers to obfuscate financial disclosure by manipulating earnings and writing overly complex reports. Moreover, our results remain robust for wider prediction windows up to three years, alternative measures of equity intent as well as for a diverse set of additional factors that potentially influence stock price crash risk.

This paper contributes to the stock price crash risk literature by describing the mechanisms between management intentions in capital structure decisions and future stock price crash risks. As financing decisions present one of the most essential business activities, our results apply relevance for both investors and a broad scope of corporate decision makers. In addition, our approach fits in with current SEC efforts to use data analytics to uncover EPS manipulation as part of their current policy agenda requiring stronger corporate disclosures to better protect investors.²⁸ In contrast to the rather basic textual information complementing prediction models applied in previous research, our approach holds additional prediction power for stock price crashes as it is sensitive towards various textual cues in management statements within firm disclosures indicative for bad news hoarding. More general, it enables to gauge management intentions from written statements and therefore provides an interesting venue for future research analysing the effects of opportunistic management behaviour on financial market outcomes. For example, using our method to identify managers' intentions with regards to capital structure decisions, future research may provide a deeper analysis of how manager time the release of good news prior to raising capital (Lang and Lundholm 2010). While this paper suggests that using information retrieval techniques that go beyond simple summarization techniques such as word counts and readability measures, future works may further explore textual cues of bad news hoarding by applying more sophisticated approaches such as named entity recognition techniques, topic models, or word embeddings.

Appendix 1: Parsing the LIQ + CAP section

Since our estimation of document embedding heavily relies on a clean identification of the LIQ + CAP section we provide a detailed description of our parsing procedure below. We start by extracting the MD&A sections using the parser proposed by Reichmann and Reichmann (2022) which outputs raw text MD&A sections that are stripped of HTML code. Next, the parser replaces HTML tags that are commonly used to indicate certain formatting are replaced with predefined tags such as “###” or “***”. Using these raw text files, we next develop an algorithm to parse the LIQ + CAP section that searches the MD&A section using a maximum of eight iterations. Each iteration is built on the notion that an MD&A follows a consistent layout in terms of formatting. Once we match any of the following specifications, the algorithm outputs the LIQ + CAP section. Otherwise, it continues to the next iteration.

²⁸ See <https://www.wsj.com/articles/sec-digs-deeper-into-companies-eps-manipulation-11633870803>.

1. *We search for formatted section headers written in capital letters:* Specifically, we search for any potential formatting tag that directly starts at the beginning of a new line that is followed by an upper case section header that fits the description of the Liquidity and Capital Resources section.²⁹ If we find a match, we extract the formatting tags and search for a new line starting with the exact same formatting tags followed by either caps-written words or a case-insensitive variation of Item 7a since the LIQ + CAP section is often found at the end of the MD&A.³⁰ We consider this first-best match as a clear indicator of the starting and ending point of the LIQ + CAP section.
2. *We search for formatted section headers with important words taking an initial capital letter:* In titles, all important words often take an initial capital letter. Therefore, in the second iteration, we basically re-run iteration two with the only difference that we require important words to have a capital letter followed by lower case letters (e.g. Liquidity and Capital Resources). Consistently, we also expect the LIQ + CAP section to end with a similar formatted section header.
3. *We search for headers written in capital letters:* We find that not all headers start with a predefined tag denoting HTML-formatting. This is especially true for firm-years from earlier years, where HTML was less common. In this case, we start by looking for headers denoting the start of the LIQ + CAP section written in capital letters. Then, we compile a list of common subsections that usually follow the LIQ + CAP section. Specifically, we compile the following regexes using OR operators:
 - a. $\backslash i \text{CONTRACTUAL} \backslash s + (\text{AND} \backslash s + \text{OTHER} \backslash s +) ? \text{OBLIGATION}$
 - b. $\backslash i (\text{CRITICAL} \backslash \text{RECENT} \backslash \text{SIGNIFICANT}) \backslash s + \text{ACCOUNTING}$
 - c. $\backslash i \text{FORWARD} (\backslash s + | -) \text{LOOKING} \backslash s + \text{STATEMENT}$
 - d. $\backslash i \text{OFF} (\backslash s + | -) \text{BALANCE} \backslash s + \text{SHEET}$
 - e. $\backslash i \text{OPERATING} (. * ?) ? \backslash s + \text{ACTIVITIES}$
 - f. $\backslash i (\text{PLAN} \backslash \text{RESULTS}) \backslash s + \text{OF} \backslash s + \text{OPERATIONS}$
 - g. $\backslash i \text{PAYMENTS} \backslash s + \text{DUE}$
 - h. $\backslash i \text{GOVERNMENT} \backslash s + \text{REGULATION}$
 - i. $\backslash i \text{REPORT} \backslash s + \text{OF} \backslash s + \text{INDEPENDENT}$
 - j. $\backslash i [\text{Tt}] [\text{Ee}] [\text{Mm}] \backslash s + \backslash d \text{A}$
 - k. $\backslash i [\text{Tt}] [\text{Ee}] [\text{Mm}] \backslash s + \backslash d$
4. *We search for headers with important words taking an initial capital letter:* Finally, if iteration 1) to 3) fail, we follow a similar concept as in iteration 2) by rerunning iteration 3) but instead of looking for fully capitalized headers we require important words taking an initial capital letter.

²⁹ Specifically, we compile the following regex as a raw string:

$(\backslash n [^ \text{A-Za-z} \backslash n " \text{ ' } +) (\text{FINANCIAL} \backslash s + \backslash w + , \backslash s + | \text{ANALYSIS} \backslash s + \text{OF} \backslash s +) ? \text{LIQUIDITY} \backslash s + (\text{AND} \backslash \&)$.

³⁰ Sometimes table headers are written in caps, so we do not match $(\backslash \text{bAT} \backslash \text{b} \backslash \text{YEAR} \backslash \text{FISCAL} \backslash \text{DECEMBER})$. The final end regex is compiled by using the exact same formatting text of the starting regex followed by the following raw string:

$(?!(\backslash \text{bAT} \backslash \text{b} \backslash \text{YEAR} \backslash \text{FISCAL} \backslash \text{DECEMBER})) \backslash \text{b} [\text{A-Z}] + \backslash \text{b} \backslash \text{n} [^ \text{A-Za-z} \backslash n " \text{ ' } * ?] [\text{Tt}] [\text{Ee}] [\text{Mm}] \backslash s + \backslash d \text{A}$.

Finally, we find that the algorithm sometimes identifies very small texts as LIQ+CAP sections that are table of contents within the MD&A. Therefore, we require a LIQ+CAP section to exceed 50 characters. Consequently, once the algorithm identifies a starting point in a table of content, it will fail to pass the 50 characters threshold for the whole section as it usually identifies the ending point immediately in the next line of the table. In this case we re-run iteration 1) to 4) after cutting off the MD&A up to the falsely identified starting point, thus forcing the algorithm to find a new starting point outside of the table of contents (Tables 10, 11).

Table 10 Descriptive statistics—additional variables

Variable	N	Min	Q1	Mean	Median	Q3	Max	SD
$EQUITY_INTENT_DBOW_{it}$	28,382	0.236	0.286	0.329	0.315	0.359	0.554	0.062
$CFVOL_{it}$	27,834	0.006	0.033	0.275	0.061	0.128	19.748	1.895
$SALESVOL_{it}$	27,414	0.009	0.092	0.403	0.182	0.358	9.740	1.012
$EARNVOL_{it}$	27,832	0.005	0.026	0.641	0.055	0.127	31.980	3.992
HHI_{it}	28,379	0.057	0.087	0.208	0.153	0.252	0.899	0.178
$FILESIZE_{it}$	28,382	0.098	0.649	1.581	1.427	2.534	3.614	1.040
$WMOD_{it}$	28,382	0.000	0.003	0.004	0.004	0.005	0.015	0.002
$ADJROTA_{it}$	25,375	-2.990	-0.401	0.010	0.002	0.462	2.546	0.928
$PROP_COST_{it}$	28,382	0.000	0.000	0.088	0.005	0.087	1.529	0.204

This table presents the descriptive statistics of variables used in our additional analyses

Table 11 Equity intent and firm characteristics

	$EQUITY_INTENT_{it}$		$EQUITY_INTENT_{it}$
(Intercept)	0.311*** (32.057)	$SIGMA_{it}$	0.027 (1.62)
$LOGMV_{it}$	-0.005*** (19.17)	$CFVOL$	0.001** (2.50)
MTB_{it}	0.000*** (5.26)	$ADJROTA_{it}$	0.002*** (5.00)
LEV_{it}	-0.006*** (-6.79)	$PROP_COST_{it}$	0.063*** (19.50)
ROA_{it}	-0.022*** (-9.91)	AGE_{it}	-0.000*** (-15.02)
Year FE	YES		
Industry FE	YES		
Observations	26,584		
Adjusted R ²	0.190		

This table presents the results for the OLS regressions of $EQUITY_INTENT_{it+1}$ on firm characteristics for the time period from 1994 to 2018. The variable $EQUITY_INTENT_{it}$ is the cosine similarity from each firm's LIQ+CAP section to the average of all equity intent vectors in which managers explicitly express their equity intent in their LIQ+CAP section. The t-statistics reported in parentheses are based on robust standard errors. *, **, and *** represent significance at the 10%, 5%, and 1% levels, respectively

Funding Open Access funding enabled and organized by Projekt DEAL.

Declarations

Conflict of interest The authors have no relevant financial or non-financial interests to disclose.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Al Mamun MD, Balachandran B, Duong HN, Gul FA (2021) Are corporate general counsels in top management effective monitors? Evidence from stock price crash risk. *Eur Acc Rev* 30:405–437. <https://doi.org/10.1080/09638180.2020.1763819>
- Alti A (2003) How persistent is the impact of market timing on capital structure? *SSRN J*. <https://doi.org/10.2139/ssrn.458640>
- Baker M, Wurgler J (2002) Market timing and capital structure. *J Financ* 57:1–32. <https://doi.org/10.1111/1540-6261.00414>
- Ball RA (2009) Market and political/regulatory perspectives on the recent accounting scandals. *J Account Res* 47:277–323. <https://doi.org/10.1111/j.1475-679X.2009.00325.x>
- Bao D, Fung SYK, Su LN (2018) Can shareholders be at rest after adopting clawback provisions? Evidence from stock price crash risk. *Contemp Account Res* 35:1578–1615. <https://doi.org/10.1111/1911-3846.12326>
- Baron RM, Kenny DA (1986) The moderator–mediator variable distinction in social psychological research: conceptual, strategic, and statistical considerations. *J Pers Soc Psychol* 51:1173–1182. <https://doi.org/10.1037/0022-3514.51.6.1173>
- Bergmann NK, Roychowdhury S (2008) Investor sentiment and corporate disclosure. *J Account Res* 46:1057–1083. <https://doi.org/10.1111/j.1475-679X.2008.00305.x>
- Callen JL, Fang X (2013) Institutional investor stability and crash risk: monitoring versus short-termism? *J Bank Finance* 37:3047–3063. <https://doi.org/10.1016/j.jbankfin.2013.02.018>
- Callen JL, Fang X (2015a) Religion and stock price crash risk. *J Financ Quant Anal* 50:169–195. <https://doi.org/10.1017/S0022109015000046>
- Callen JL, Fang X (2015b) Short interest and stock price crash risk. *J Bank Finance* 60:181–194. <https://doi.org/10.1016/j.jbankfin.2015.08.009>
- Chen J, Hong H, Stein J (2001) Forecasting crashes: trading volume, past returns and conditional skewness in stock prices. *J Financ Econ* 61:345–381. <https://doi.org/10.3386/w7687>
- Chi J, Gupta M (2009) Overvaluation and earnings management. *J Bank Finance* 33:1652–1663. <https://doi.org/10.1016/j.jbankfin.2009.03.014>
- Clausen S, Hirth S (2016) Measuring the value of intangibles. *J Corp Finan* 40:110–127. <https://doi.org/10.1016/j.jcorpfin.2016.07.012>
- Cohen DA, Zarowin P (2010) Accrual-based and real earnings management activities around seasoned equity offerings. *J Account Econ* 50:2–19. <https://doi.org/10.1016/j.jacceco.2010.01.002>
- DeAngelo H, DeAngelo L, Stulz RM (2010) Seasoned equity offerings, market timing, and the corporate lifecycle. *J Financ Econ* 95:275–295. <https://doi.org/10.1016/j.jfineco.2009.11.002>
- Dechow P, Sloan R, Sweeney A (1995) Detecting earnings management. *Account Rev* 70:193–225
- Dimson E (1979) Risk measurement when shares are subject to infrequent trading. *J Financ Econ* 7:197–226

- DuCharme LL, Malatesta PH, Sefcik SE (2004) Earnings management, stock issues, and shareholder lawsuits. *J Financ Econ* 71:27–49. [https://doi.org/10.1016/S0304-405X\(03\)00182-X](https://doi.org/10.1016/S0304-405X(03)00182-X)
- Eaglesham J (2013) Accounting fraud targeted. *Wall Street J* 2013:C1
- El-Haj M, Rayson P, Walker M, Young S, Simaki V (2019) In search of meaning: lessons, resources and next steps for computational analysis of financial discourse. *J Bus Financ Acc* 46:265–306. <https://doi.org/10.1111/jbfa.12378>
- Ertugrul M, Lei J, Qiu J, Wan C (2017) Annual report readability, tone ambiguity, and the cost of borrowing. *J Financ Quant Anal* 52:811–836. <https://doi.org/10.1017/S0022109017000187>
- Franco-Santos M, Lucianetti L, Bourne M (2012) Contemporary performance measurement systems: a review of their consequences and a framework for research. *Manag Account Res* 23:79–119. <https://doi.org/10.1016/j.mar.2012.04.001>
- Giroud X, Mueller HM (2010) Does corporate governance matter in competitive industries? *J Financ Econ* 95:312–331. <https://doi.org/10.1016/j.jfineco.2009.10.008>
- Graham JR, Harvey CR (2001) The theory and practice of corporate finance—evidence from the field. *J Financ Econ* 60:187–243. [https://doi.org/10.1016/S0304-405X\(01\)00044-7](https://doi.org/10.1016/S0304-405X(01)00044-7)
- Graham JR, Harvey CR, Rajgopal S (2005) The economic implications of corporate financial reporting. *J Account Econ* 40:3–73. <https://doi.org/10.1016/j.jacceco.2005.01.002>
- Harris ZS (1954) Distributional structure. *Word* 10:146–162. <https://doi.org/10.1080/00437956.1954.11659520>
- Hoberg G, Maksimovic V (2014) Redefining financial constraints: a text-based analysis. *Rev Financ Stud* 28:1312–1352. <https://doi.org/10.1093/rfs/hhu089>
- Hong HA, Kim J-B, Welker M (2017) Divergence of cash flow and voting rights, opacity, and stock price crash risk: international evidence. *J Account Res* 55:1167–1212. <https://doi.org/10.1111/1475-679X.12185>
- Hung S, Qiao Z (2017) Shadows in the sun: crash risk behind earnings transparency. *J Bank Finance* 83:1–18. <https://doi.org/10.1016/j.jbankfin.2017.06.007>
- Hutton AP, Marcus AJ, Tehrani H (2009) Opaque financial reports, R2, and crash risk. *J Financ Econ* 94:67–86. <https://doi.org/10.1016/j.jfineco.2008.10.003>
- Jensen MC (2005) Agency costs of overvalued equity. *Financ Manage* 34:5–19. <https://doi.org/10.1111/j.1755-053X.2005.tb00090.x>
- Jensen MC, Meckling WH (1976) Theory of the firm: managerial behavior, agency costs, and ownership structure. *J Financ Econ* 3:305–360. <https://doi.org/10.1002/9780470752135.ch17>
- Jin L, Myers S (2006) R2 around the world: New theory and new tests. *J Financ Econ* 79:257–292. <https://doi.org/10.1016/j.jfineco.2004.11.003>
- Kim Y, Park MS (2005) Pricing of seasoned equity offers and earnings management. *J Financ Quant Anal* 40:435–463. <https://doi.org/10.1017/S0022109000002374>
- Kim J-B, Zhang L (2016) Accounting conservatism and stock price crash risk: firm-level evidence. *Contemp Account Res* 33:412–441. <https://doi.org/10.1111/1911-3846.12112>
- Kim J-B, Li Y, Zhang L (2011) Corporate tax avoidance and stock price crash risk: Firm-level analysis. *J Financ Econ* 100:639–662. <https://doi.org/10.1016/j.jfineco.2010.07.007>
- Kim J-B, Wang Z, Zhang L (2016) CEO overconfidence and stock price crash risk. *Contemp Account Res* 33:1720–1749. <https://doi.org/10.1111/1911-3846.12217>
- Kim C, Wang K, Zhang L (2019) Readability of 10-K reports and stock price crash risk. *Contemp Account Res* 36:1184–1216. <https://doi.org/10.1111/1911-3846.12452>
- Kim J-B, Si Y, Xia C, Zhang L (2021) Corporate derivatives usage, information environment, and stock price crash risk. *Eur Account Rev*. <https://doi.org/10.1080/09638180.2021.1918564>
- Kothari SP, Shu S, Wysocki PD (2009) Do managers withhold bad news? *J Account Res* 47:241–276. <https://doi.org/10.1111/j.1475-679X.2008.00318.x>
- Krasker WS (1986) Stock price movements in response to stock issues under asymmetric information. *J Financ* 41:93–105. <https://doi.org/10.1111/j.1540-6261.1986.tb04493.x>
- Lang MH, Lundholm RJ (2010) Voluntary disclosure and equity offerings: reducing information asymmetry or hyping the stock? *Contemp Account Res* 17:623–662. <https://doi.org/10.1506/9N45-F0JX-AXVW-LBWJ>
- Lau JH, Baldwin T (2016) An empirical evaluation of doc2vec with practical insights into document embedding generation. In: 1st Workshop on Representation Learning for NLP, pp 78–86
- Le Q, Mikolov T (2014) Distributed representations of sentences and documents. In: Proceedings of the 31st international conference on machine learning 32, pp 1188–1196
- Li K, Mai F, Shen R, Yan X (2021) Measuring corporate culture using machine learning. *Rev Financ Stud* 34:3265–3315. <https://doi.org/10.1093/rfs/hhaa079>

- Lo K, Ramos F, Rogo R (2017) Earnings management and annual report readability. *J Account Econ* 63:1–25. <https://doi.org/10.1016/j.jacceco.2016.09.002>
- Loughran T, McDonald B (2011) When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks. *J Financ* 66:35–65. <https://doi.org/10.1111/j.1540-6261.2010.01625.x>
- Luo H, Liu I, Tripathy N (2021) A study on firms with negative book value of equity. *Int Rev Financ* 21:145–182. <https://doi.org/10.1111/irfi.12260>
- MacKinnon JG, White H (1985) Some heteroskedasticity-consistent covariance matrix estimators with improved finite sample properties. *J Econometr* 29:305–325. [https://doi.org/10.1016/0304-4076\(85\)90158-7](https://doi.org/10.1016/0304-4076(85)90158-7)
- Matin R, Hansen C, Hansen C, Mølgaard P (2019) Predicting distresses using deep learning of text segments in annual reports. *Expert Syst Appl* 132:199–208. <https://doi.org/10.1016/j.eswa.2019.04.071>
- Mikolov T, Sutskever I, Chen K, Corrado G, Dean J (2013b) Distributed representations of words and phrases and their compositionality. In: Jordan MI, LeCun Y, Solla SA (eds) *Advances in neural information processing systems*. MIT, Cambridge, pp 3111–3119
- Mikolov T, Chen K, Corrado G, Dean J (2013a) Efficient estimation of word representations in vector space. In: 1st International conference on learning representations, ICLR 2013a, Scottsdale, Arizona, USA, May 2–4, 2013a, workshop track proceedings
- Pennington J, Socher R, Manning CD (2014) Glove: global vectors for word representation. In: Alessandro Moschitti QCRI, Bo Pang G, Walter Daelemans UoA (eds) *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*. Association for Computational Linguistics, Stroudsburg, pp 1532–1543
- Reichmann D, Reichmann M (2022) Predicting firm-specific stock price crashes. Working Paper
- Roychowdhury S (2006) Earnings management through real activities manipulation. *J Account Econ* 42:335–370. <https://doi.org/10.1016/j.jacceco.2006.01.002>
- Schmidt PS, von Arx U, Schrimpf A, Wagner AF, Ziegler A (2019) Common risk factors in international stock markets. *Fin Markets Portfolio Mgmt* 33:213–241. <https://doi.org/10.1007/s11408-019-00334-3>
- Sunder S (2010) Riding the accounting train: from crisis to crisis in eighty years. In: *Presentation at the conference on financial reporting, auditing and governance*
- Walker MD, Yost K (2008) Seasoned equity offerings: what firms say, do, and how the market reacts. *J Corp Financ* 14:376–386. <https://doi.org/10.1016/j.jcorpfin.2008.04.001>
- Wu K, Lai S (2020) Intangible intensity and stock price crash risk. *J Corp Financ* 64:101682. <https://doi.org/10.1016/j.jcorpfin.2020.101682>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.