# Speculative and Informative: Lessons from Market Reactions to Speculation Cues

J. Anthony Cookson S. Katie Moon Joonki Noh*

*Cookson and Moon are at University of Colorado at Boulder, Leeds School of Business, Campus Box 419, Boulder, CO 80309, USA, [tony.cookson@colorado.edu; katie.moon@colorado.edu]. Noh is at Case Western Reserve University, Weatherhead School of Management, 11119 Bellflower Rd, Cleveland, Ohio 44106 [joonki.noh@case.edu]. The authors are grateful to conference and seminar participants at the 2018 Midwest Finance Association Conference, the 2018 China International Conference in Finance, the 2018 International Industrial Organization Conference, the 2020 Future of Financial Information Conference, the SEC DERA group, the 2020 Asia-Pacific Association of Derivatives Conference, the 2020 Northern Finance Association Conference, the 2020 Financial Management Association Conference, the NBER Big Data and Securities Markets Conference Fall 2020, the 2022 American Finance Association Conference, the 2022 Summer Finance Roundtable, Arizona State University, Brigham Young University, Case Western Reserve University, Drexel University, KAIST, Korea University, SungKyunKwan University, Texas Christian University, University of Colorado brownbag, and the University of Utah, as well as Gustaf Bellstam, Asaf Bernstein, Steve Billings, Stefan Bueller, Brendan Daley, Naveen Daniel, Diego Garcia, Kathleen Weiss Hanley, Jerry Hoberg, Sahn-Wook Huh, Dave Ikenberry, Ryan Israelsen, Hogyu Jhang, Fuwei Jiang, Ron Kaniel, Hugh Hoikwang Kim, Jaeram Lee, Uday Rajan, Wendy Rotenberg, Kelly Shue, Noah Stoffman, Shawn Thomas, Ralph Walkling, Brian Waters, Michael Weisbach, Jaime Zender, and Dexin Zhou for helpful comments and suggestions. In addition, the authors are grateful to Jerry Hoberg and Bill McDonald for making their textual measures available on their respective websites. All remaining errors are our own.

# Speculative and Informative: Lessons from Market Reactions to Speculation Cues


**ABSTRACT**

Speculative language in corporate disclosures can convey valuable information on firms' fundamentals. We evaluate this idea by developing a measure for speculative statements based on sentences marked with the "weasel tag" on Wikipedia. In the 16-week test period after filing, greater use of speculative statements in 10-Ks predicts positive and non-reverting abnormal returns, improvements to stock liquidity, more insider and informed buying, and more positive news sentiment. These findings are driven by disclosures that are more forward-looking and contain more R&D terms. These findings imply that manager usage of speculative language tends to reflect voluntary disclosure of private information, which reveals positive but uncertain future events.



**Keywords**: text analysis, market reactions, voluntary disclosure, information, uncertainty, speculative statements

# 1 Introduction

Speculative language is often *necessary* to convey important information.[1] For example, in scientific writing, authors often use speculative language to present directions of future research (Kilicoglu and Bergler (2008)). In this paper, we ask whether speculative language in corporate disclosures could be similarly valuable. Conventional wisdom suggests a genuine tension. On one hand, speculative language could obfuscate since it is not entirely concrete. Such obfuscation would be concerning to investors and regulators who scrutinize corporate filings. On the other hand, speculative language could be employed by managers to signal positive prospects of their firms, just like a scholar outlining promising future directions for research.

These two opposite intuitions about the potential effects of speculative language in financial statements are exemplified in the following excerpt from an annual financial statement (10-K) filed by Moderna Inc., a firm that conducts biomedical research:

> The intrinsic advantages of using mRNA as a medicine: We believe mRNA possesses inherent characteristics that could serve as the foundation for a new class of medicines.··· mRNA is used by cells to vary the quantities of protein produced over time, in different locations, and in various combinations. Given the universal role of mRNA in protein production, we believe that mRNA medicines could have broad applicability across human disease.
>
> (Moderna, Inc.'s 10-K filing in February 2020)

This 10-K was filed with the SEC immediately before the declaration of the COVID-19 outbreak in the U.S. in March 2020. The underlined words are speculative cues from our textual classification. In this example, these words were necessary to describe early-stage mRNA prospects since the technology had not yet been reviewed by the FDA, which would approve of this technology's safety and efficacy nine months later. Are statements of belief, like Moderna's, useful to convey future prospects, or could they be used to obfuscate the underlying truth, and thus, a subject of concern? In this paper, we examine this research question systematically by constructing an explicit measure of speculative language and using it to understand the information content of speculative statements made in corporate disclosures.

[1]The Cambridge Dictionary defines speculation as "the activity of guessing possible answers to a question without having enough information to be certain. The Oxford Learner's Dictionary defines it as "the act of forming opinions about what has happened or what might happen without knowing all the facts." Therefore, by its nature, speculative language can have both components: information and the uncertainty of information.

Using our speculation measure, we find that during the 16 weeks after 10-K filings, greater speculative language in 10-Ks predicts permanent and positive abnormal returns, improvements to stock liquidity, greater intensities of insider and informed buying (but not those of insider and informed selling), and the arrival of positive-sentiment news about the firm. Together, these findings support the idea that the use of speculative language by managers can convey positive and value-relevant information about firm activities, leading to positive and delayed market reactions. This contrasts with the view that speculative language might reflect managers' obfuscation, which would lead to negative market reactions.

It is empirically challenging to measure the extent of speculative language in financial disclosures for at least two reasons. First, the selection of a list of speculation cues is likely to be fraught with subjectivity, leading to concerns about researcher degrees of freedom (Simmons, Nelson, and Simonsohn (2011)). Second, despite there being related concepts in the literature (e.g., uncertainty, weak modality, vagueness), there is not a pre-existing list of speculation cues. We address these dual challenges by constructing a new dictionary of speculation *cues* (not keywords) that draws on Wikipedia's crowdsourced solution to identify statements under the limits of evidence or with unverifiable information — "weasel tags." As a popular and widely-used encyclopedia, Wikipedia has a strong incentive to separate factual statements from speculative ones for its credibility by advising its users to attach weasel tags when they encounter sentences or phrases in Wikipedia articles that are unverifiable or speculative. By appealing to Wikipedia's crowdsourced solution, we provide a reliable basis for identifying speculative cues while "tying our hands" by eliminating researchers' subjective choices typically involved in building a language dictionary.[2] In our approach of defining a dictionary of speculative cues, which are words used when language is speculative, the entire sentence that contains a cue word has the *context* of speculation rather than each cue word itself having a direct *meaning* of speculation.

Using our dictionary of speculation cues, we generate a measure for the extent of speculative language at the firm-year level by computing the percentage of speculation cues in each firm's 10-K filing. Consistent with the idea that our measure captures managers' degree of speculation

[2]Our approach of appealing to an external source to ground our textual analysis of speculation cues is similar to that taken in Bellstam, Bhagat, and Cookson (2020), who use an innovation textbook as a benchmark to evaluate which topics discussed by analysts reflect the innovation activities of the firms they cover.

in 10-Ks, we find that 10-Ks with more speculative language tend to exhibit greater uncertainty and to contain more modal words that convey differing shades of meaning. Yet, the information contained in our speculative language measure is distinct from existing textual measures. We also find that 10-Ks with more speculative language tend to have more positive sentiment. These results are consistent with the idea that our speculation measure can capture not only positive information but also uncertainty, each of which can lead to very different (possibly opposite) market outcomes in the post-filing periods.

In our main tests, we examine how the market reacts to speculative language in 10-K disclosures. For a 16-week window following each 10-K filing (roughly a four-month interval), we find that more speculative language in 10-Ks predicts (i) positive and delayed price reactions (BHARs and CARs) that do not eventually revert, (ii) greater stock liquidity, evidenced in lower bid-ask spreads, (iii) greater probability of informed buying, based on Brennan, Huh, and Subrahmanyam (2018), and greater volume of insider purchases, with no difference in informed and insider selling, and (iv) more positive-sentiment news, using a news sentiment variable available from RavenPack. With the exception of insider purchases, which are likely informed before the 10-K disclosure, all these outcomes exhibit no pre-trends in the four-week period prior to the 10-K disclosure. Past eight weeks following the 10-K disclosure, the price reactions level off. These core empirical results are consistent with speculative language, as a form of voluntary disclosure, reflecting the private information of firm managers about positive but uncertain future events.

A simple conceptual framework guides our interpretation of this main result. In this framework, a manager chooses a type of disclosure (voluntary vs. mandatory) and a type of language (speculative vs. non-speculative). Given timely disclosure rules of material information, speculative language can be used to convey important positive information if it is sufficiently uncertain, or it could be used to obfuscate negative information, giving rise to our central empirical tension. Given that we find a delayed positive market reaction to speculative language, we conclude that the information component in speculative language dominates the uncertainty component, while the uncertainty component still plays a second-order role of delaying the market reactions.

As a complement to these core tests, we also evaluate the textual content of the disclosures that drives these results. Specifically, we detect two consistent features of the textual disclosures in the sentences that surround speculation cues both as words and bigrams: (1) forward-looking disclosures about the firm's unresolved, but potentially valuable, plans (e.g., "forward-looking" and "future cash") and (2) disclosures of product innovation and R&D terms (e.g., "clinical trial" or "product candidate"). This pattern is notable because both forward-looking disclosure and disclosure on R&D activities are almost entirely voluntary (Jones (2007) and Muslu, Radhakrishnan, Subramanyam, and Lim (2015)). Therefore, when information is voluntarily disclosed before the disclosure is mandated, managers have incentives to disclose relatively good news (Verrecchia (1983), Dye (1985), and Skinner (1994)). These characteristics of the textual content surrounding speculation cues, summarized as forward-looking disclosures of R&D with relatively good news, are consistent with the previous example of Moderna's speculative language on the early-stage mRNA prospects.

From our conceptual framework, we also make an auxiliary prediction: the reaction to speculative disclosure is stronger if the firm's information is about future outcomes, especially if the information is proprietary. Consistent with this prediction, we find that the price reaction to speculative language is stronger for firms that use more forward-looking terms and disclose more information about R&D activities.[3] These findings corroborate our core interpretation that the use of speculative language reflects the positive information about future firm prospects.

Our findings rule out several alternative interpretations. First, we observe no pre-trends in abnormal returns prior to 10-K disclosure dates and no subsequent reversal of returns. This pattern of returns is inconsistent with the possibility that investors overreact to the speculative language in 10-Ks, and it is difficult to explain via a standard risk-based explanation. Second, during the same time frame when there are significant positive abnormal returns, stock liquidity also improves, contrasting further with a risk-based explanation in which investors perceive the uncertainty in speculative language as a systematic risk. Third, consistent with the idea that speculative language can deliver positive information, we find that both the probability of informed buying and news sentiment about the firm with a high extent of speculation are greater. Fourth, during the entire test period around 10-K disclosures with high speculation, we find that insiders purchase significantly

[3]In Appendix Table A.3, we also find that our speculation measure is positively associated with the next year's R&D investments and significantly more so for firms with high growth opportunities proxied by Tobin's Q. This evidence is consistent with our economic story that a manager releases her private information about positive R&D investment opportunities using speculative language before making the real investments.

more of the firm's shares. Furthermore, our finding is not driven by the difference in systematic risks captured by the Fama-French three-factor model, and is also robust to controlling for the abnormal return on the 10-K filing-day window that proxies a quickly varying risk exposure in the spirit of Patton and Verardo (2012). All these results collectively are inconsistent with a risk-based explanation.

Our paper makes several contributions. First, our evidence on the use of speculative language in firm disclosures relates to the work on discretionary disclosure and persuasion through information revelation (e.g., Bloomfield (2002)). Discretionary disclosure leads to full disclosure in a perfect information environment, but not in the presence of proprietary costs or other market frictions (Ross (1979), Verrecchia (1983), Kamenica and Gentzkow (2011), and Ely (2017)). Following this line of research, recent empirical applications have focused on how the disclosure of bad news can signal a firm's quality (Gormley, Kim, and Martin (2012) and Gao, Liang, Merkley, and Pacelli (2017)). Also, Cohen, Malloy, and Nguyen (2016) has shown that there is information content in the minor changes from year to year in firm disclosures that slowly but eventually is capitalized into asset prices. Our results on the informational value of speculative language provide a novel and unique perspective on this research question. We show that our explicit measure of speculative language relates to providing early access to managers' private information. Our results suggest that managers act in their decisions to provide more voluntary information on immature but positive opportunities that their firms have.[4]

Second, our work is a part of a growing literature within finance and accounting that makes use of text descriptions to study important aspects of financial market reactions (Tetlock (2007), Hoberg and Phillips (2016), Hoberg and Lewis (2017), Hoberg and Moon (2019), and Bellstam, Bhagat, and Cookson (2020)). Within the broader literature on textual analysis in finance, our work is most closely related to applying textual analysis tools to analyze the financial information (Hanley and Hoberg (2010), Dougal, Engelberg, Garcia, and Parsons (2012), Loughran and McDonald (2013),

[4]Although speculative language has not been studied in the academic literature of finance and accounting, the uncertainty aspect of speculative language can be related to linguistic complexity or readability in the literature. Similar to our motivation, this literature disagrees regarding whether complex language contains more or less information. On one hand, the "obfuscation view" suggests that managers strategically increase the complexity of their disclosures, which increases information asymmetry (Li (2008)) and decreases valuations (Hwang and Kim (2017)). On the other hand, the "information view" suggests that complex language is needed to explain complex situations and thus can convey important information that cannot be disclosed in simple terms (e.g., Bushee, Gow, and Taylor (2018)).

Garcia (2013), and Jegadeesh and Wu (2017)). We show that our measure is sensibly related to, but distinct from the existing lexicon of measures — many of which are available in the master dictionary by Loughran and McDonald (2011). Relative to these other textual measures, our speculation measure provides a useful description of speculative language in financial disclosures, which is distinctive unto itself. Also, our textual analysis approach of using word cues instead of keywords is unique and effective in capturing a more nuanced concept like speculation, thereby offering a new approach that could be adapted in other contexts. In this respect, we anticipate fruitful applications of our speculative language measure and approach to understand better the information environment into which the speculative language can be injected.

## 2 Conceptual Framework

This section presents a structured discussion of hypotheses, which guides our empirical analyses and economic interpretations of their results. First, we describe the manager's disclosure decision and language choice in it, based on her private information. Second, we relate these choices to the stock market reaction and the timing thereof. This discussion builds upon existing theoretical and empirical studies to derive testable predictions for how stock prices react to speculative language in corporate disclosures (and the likely information content of the speculative language).

In our conceptual framework, a manager is endowed with a private information signal that varies along the following dimensions: i) *timing* of the event (current vs. future), ii) *direction* of the signal (positive vs. negative), and iii) *strength* of the signal (strong vs. weak). These three dimensions well match important aspects of firm's disclosure in reality. First, with respect to timing of an event, e.g., a manager can choose to focus more on current earnings or future earnings opportunities in 10-K. Second, the direction of a signal is an important dimension, particularly due to the information asymmetry between a manager and investors. For example, the manager's private information can be about the success or failure of a clinical trial that her firm has been conducting during a fiscal year, depending on which she can choose to disclose it or not. Lastly, based on the strength of the signal, the manager can choose what kinds of language to use in the disclosure.

We expect the manager to use more speculative language when her information is based on a weak signal than a strong signal. Using the three dimensions above, all in binary forms for simplicity, we present six possible scenarios of manager's disclosure decision and language choice, and the associated price reaction in Figure 1. Studying the price reactions to different disclosure and language choices of a manager is interesting by itself. However, it is particularly useful in our setting since we can separate out Scenario D, in which the manager uses more speculative language as an effort to deliver positive but uncertain future information, and Scenario E, in which the manager employs speculative language as an attempt to obfuscate investors and to reduce possible adverse consequences of releasing negative future information.

[Insert Figure 1 Here]

In Scenarios A and B, the manager has private information about current events. In this case, regardless of whether the signal is good or bad, the SEC's Regulation FD ("Fair Disclosure") mandates that the information should be disclosed as concretely as possible. We note that the signal for this type of information must be strong as the information is about current events. Therefore, the information in Scenarios A and B is disclosed by mandates without speculative language. In these two scenarios, the predictions of price reactions are straightforward and depend only on the directions of signal that the manager received. We expect a positive and immediate price reaction in Scenario A and a negative and immediate price reaction in Scenario B.

In Scenarios C and D, the manager obtains positive information about future events in the firm. The information is based on a strong signal in Scenario C while it is based on a weak signal in Scenario D. As the timing of events is future, the disclosure of this type of information is naturally voluntary. Earlier studies on voluntary disclosures show that managers have incentives to leak or reveal good news immediately to investors (Verrecchia (1983), Dye (1985), and Kothari, Shu, and Wysocki (2009)). We expect that such incentives for the immediate releases of good news will exist regardless of the strength of the manager's signal. Therefore, we predict that the positive information will be disclosed by the manger voluntarily in both Scenarios C and D.

However, the manager's language choice and the price reaction can be different between Scenarios C and D. In Scenario C, we expect the manager to use non-speculative language in the disclosure since she has a strong signal, which leads to a positive and immediate price reaction. In contrast, in Scenario D, since the manager has only a weak signal with the limits of concrete evidence for the positive news, we expect the manager to use more speculative language to avoid any perception of misleading investors in case the positive event might not be realized in the future. Since more speculative language is used, the disclosure in Scenario D inevitably contains both aspects of speculative language, i.e., information and uncertainty. To the extent that the market is not fully efficient but can understand the information in speculative language, we thus expect that the information content in speculation statements is not immediately processed by investors, leading to a *positive* and *delayed* price reaction. This prediction is consistent with Daniel, Hirshleifer, and Subrahmanyam (1998), Daniel, Hirshleifer, and Subrahmanyam (2001), Hirshleifer (2001), and Zhang (2006) who show that uncertainty in information can delay the digestion of value-relevant information by investors.[5]

In Scenarios E and F, the manager has negative information about future events in the firm. The information is based on a strong signal in Scenario E while it is based on a weak signal in Scenario F. Interestingly in Scenario E, although the disclosure of the information on future events is voluntary, the manager is more likely to disclose the negative information voluntarily because she receives a strong signal. Such disclosure of negative future information helps the manager avoid high legal and reputational costs as in Skinner (1994). The manager's language choice in Scenario E is unclear. Although we expect the manager to use non-speculative language in the disclosure since the negative signal is strong, she might have incentives to employ obfuscating language to avoid potential adverse consequences of disclosing the negative information.[6] Therefore, we expect the negative future information with a strong signal could contain speculative language. However, under Scenario E, we expect a negative price reaction, which contrasts with the positive price reaction in Scenario D.

When a signal about negative future events is sufficiently weak in Scenario F, the information is less likely to be disclosed unless the signal becomes more certain and concrete in the future, which is not against the SEC's regulations. Kothari, Shu, and Wysocki (2009) and Bao, Kim, Mian, and Su (2019) show empirical evidence on managers' incentives to withhold bad news up to

[5]Motivated by the theoretical predictions of Daniel, Hirshleifer, and Subrahmanyam (1998) and Hirshleifer (2001), Zhang (2006) empirically shows that investors' biases can be amplified by the uncertainty in information, which is proxied by cash-flow or return volatility, leading to stronger under-reactions to the information.

[6]Managers tend to employ various tactics to reduce the negative impacts of releasing bad news on firms or themselves, e.g., by spinning positive news through investor relations firms (Solomon (2012)), by using an excessively positive tone (Huang, Teoh, and Zhang (2013)), by hosting evasive shareholder meetings (Li and Yermack (2016)), by playing favorites for the bullish analysts (Cohen, Lou, and Malloy (2020)) or by blaming external factors (Noh and Zhou (2022)).

a certain threshold, which is supported by the theoretical models in Verrecchia (1983), Dye (1985), and Acharya, DeMarzo, and Kremer (2011).

These six scenarios are a stylized description of the complex nature of a manager's private information and disclosure and language choices. However, they effectively illustrate the type of private information to be disclosed voluntarily by the manager using speculative language and the expected price reaction. Importantly, only Scenario D clearly leads to more usage of speculative language with a positive and delayed price reaction. A more complete model would allow the amount of disclosure and the degree of speculative language to vary continuously with the signal strength that the manager receives, but this is unlikely to change the qualitative predictions we present nor the motivating intuition. Scenario E might also lead to the use of speculative language, but it predicts a negative price reaction, which is empirically distinguishable from the prediction of Scenario D.[7]

Lastly, aside from the direction and timing of price reactions, this discussion makes another notable prediction: the price reactions to speculative language ought to be greatest when the firm's information is about future states − e.g., forward-looking statements and forward-looking investments such as R&D. We test this auxiliary prediction via a series of heterogeneity tests after we present our main results in Section 5.

# 3 Speculative Language in 10-K Disclosures

## 3.1 Weasel Words and Phrases in Wikipedia Articles

We begin by constructing a dictionary of "speculation cues." To do this, we appeal to unique features of Wikipedia, which effectively crowdsources our identification of speculative language. Specifically, we take all Wikipedia articles as of April 20, 2017 as our text corpus, identify sentences with weasel tags attached, and compile a list of speculation cues from these weasel-tagged sentences. Ganter and Strube (2009) describe three broad categories of weasel words or phrases used in Wikipedia articles: 1) numerically vague expressions (e.g., "a number of"), 2) the passive voice (e.g., it is said), and 3) adverbs that weaken (e.g., probably). Examples of these weasel words or phrases directly given

[7] If investors perceive the uncertainty in speculative language as a systematic risk, the speculative statements in firm disclosures might be associated with higher expected return (e.g., see Barry and Brown (1985) and Coles and Loewenstein (1988)). In Section 5, we test this alternative risk-based explanation by using different proxies for systematic risks.

by Wikipedia as style guidelines include "People are saying...", "There is evidence that...", and "It has been mentioned that...". Wikipedia users are then advised to avoid using weasel words or phrases and at the same time to detect and mark excessive uses of such words or phrases by others using a special weasel tag, {{Weasel-inline—{{subst:DATE}}}}, for improvement. The following examples illustrate how the weasel tag is used in sentences of Wikipedia articles:

- "The Tic Tok Men"
  Many{{weasel inline—date=March 2009}} consider this album to be the quintessential Tic Tok sound.
- "Manu Parrotlet"
  It has been said{{weasel inline—date=January 2014}} that the Manu parrotlet can be seen along the Man on top of trees across from the Altamira beach about 25 minutes from the Manu Resort.
- "Nathaniel Mather"
  He finished his studies in England probably{{weasel inline—date=January 2014}} returning with his brother [[Samuel Mather (Independent minister)—Samuel]] in 1650.

We process a Wikipedia dump that is completed on April 20, 2017 and comprised of 17,483,910 articles and extract sentences that contain weasel tags.[8] We start by extracting all words in sentences that contain weasel tags. Because weasel tags are typically removed after the language is edited and improved, the tags are not frequently observed at any given snapshot of Wikipedia articles. Therefore, sentences containing weasel tags are rare despite the large number of Wikipedia articles that we process. We identify 433 sentences with weasel tags from 367 Wikipedia articles after removing corrupt or redundant sentences. Our number of weasel tags is slightly more than 328 weasel tags identified by Ganter and Strube (2009) who processed two Wikipedia dumps with different completion dates.[9]

The numbers of unique and total words in the extracted sentences containing weasel tags are approximately 6,000 and 16,000, respectively. Based on this textual corpus, we calculate the frequencies of each word and bigrams and trigrams as well to better identify potential weasel words and phrases. The bigrams and trigrams help to capture weasel phrases that use the passive voice and appeal to anonymous authority. We note that a raw sort of frequency of word usage will not

[8]Wikipedia dumps are available for downloading at `https://dumps.wikimedia.org/`. See Wikipedia's own article about weasel words at `https://en.wikipedia.org/wiki/Weasel_word`, which provides more context.

[9]Due to Wikipedia's open-source nature, sentences with weasel tags are removed in the crowd-sourced editing process. Therefore, we acknowledge the limitation of our methodology that the list of extracted speculation cues might not be complete if some words with weasel tags were already removed.

accurately capture the most distinct cues for speculation because common words tend to occur more frequently. In the sort of most frequent words (see Panel A(i) of Table 1), commonly used words tend to show up as most frequent, despite not being weasel words themselves (e.g., stop words like "the", "and", and "that"). This is a much larger issue with the unigrams than with the bigrams or trigrams.

[Insert Table 1 Here]

To ensure that we do not merely pick up commonly used words for our speculation cues, we extract and use only the speculation cues that are the most distinctive of the weasel-tagged sentences relative to control sentences. Specifically, for each weasel-tagged sentence, we extract a control sentence that occurs three sentences later in the same Wikipedia article. We note that these control sentences are free of weasel language and have the virtue that they are on the same set of topics as the weasel sentence. Using these control sentences together with the corresponding weasel-tagged sentences, we compute the saliency of the words in the weasel-tagged sentences relative to control sentences based on Goldsmith-Pinkham, Hirtle, and Lucca (2016).[10] This saliency measure captures the degree to which the words are overused relative to the common language, and is thus, appropriate for screening a list of the common language. Panel A(ii) of Table 1 shows how effective our saliency screening is in filtering out common language from the list of words. The most common words (e.g., "the" and "and") are least salient, and are filtered out of our dictionary of speculation cues.

After filtering out common language using the saliency screen, we compile our final list of speculation cues that consists of unigrams, bigrams, and trigrams. Further, we expand the list of speculation cues using variations on these words such as the singular and plural forms for nouns and the past, present, and future tenses for verbs. We also manually eliminate redundancy in bigrams and trigrams in cases where including both would count the same phrase twice. In addition, Wikipedia has published guidelines for weasel words with specific examples to help users understand and identify the weasel language. Our methodology captures the vast majority of the example

[10] Salience is computed by $Salience(word|weasel\ sentence) = p(word\&weasel\ sentence) \times log(\frac{p(weasel\ sentence|word)}{p(weasel\ sentence)})$. We also consider a re-weighted version of the frequencies for speculation cues using term frequency-inverse document frequency (tf-idf) weights that mirror the intuition of the Goldsmith-Pinkham, Hirtle, and Lucca (2016) saliency filter. We prefer to use the saliency filtered list for speculation cues since it is more transparent and involves fewer researcher choices.

phrases offered by Wikipedia, but several example phrases in the guidelines are not in the Wikipedia dump that we analyze. To maintain the most comprehensive list of speculation cues, we also include these guideline weasel words or phrases in our final list. The complete dictionary for speculation cues is provided in Appendix A.

Our dictionary contains speculation “cues” *signaling* that the context of the entire sentence including the cue words is speculative, not keywords where each keyword means speculation. Therefore, cues that confer the opposite meaning of speculation such as “clear” and “clearly” can be also included in the dictionary. This implies that people tend to add those cues to create an impression. Also, for the transparency of reporting, we fully list all words in the dictionary created using Wikipedia articles, but certain words in the dictionary such as “award-winning” are virtually never mentioned in 10-K filings and thus are safe to be ignored in the consideration (the fraction of the mentions of “award-winning” in total mentions of all speculation cues in 10-K flings is 0.0000327).

The dictionary for speculation cues is distinct from notable sets of keywords that exist in the literature. For example, Panel A(iii) of Table 1 presents the top 10 most frequently used words in 10-Ks based on our dictionary for speculation cues, and for comparison, on the dictionaries for uncertainty and weak and strong modality taken from the Loughran and McDonald (2011) master dictionary. The most frequently used words in each of these dictionaries have minimal overlap with one another, indicating that our speculation measure using the speculation cues is distinct from these related measures and thus can contain unique information. For example, numerically vague expressions such as “other”, “number of”, and “various” are uniquely included in the top 10 most frequently used speculation cues.[11] Also, several passive expressions such as “said”, “considered”, and “found” are frequently used speculation cues in 10-Ks, although those are not included in the top 10 list.

### 3.2 Speculative Language in 10-K Disclosures

The final step in our text processing procedure is to download all 10-K filings whose report dates range from 1997 to 2021 and to extract the raw counts of mentions of each of the speculation

[11]The most frequently used unigram, “other”, can be simply mentioned in 10-Ks to refer to an accounting item that contains “other”, e.g., as in “Other Comprehensive Income”, “Assets - Other”, “Liabilities - Other - Total”. Our findings discussed in the subsequent sections are robust to excluding “other” from our speculation dictionary. Also, the count of the unigram “common” excludes “common share(s)” and “common stock(s)” that can be frequently mentioned in 10-Ks.

cues by firm-year. This generates a full panel of speculation cue vectors with 60,982 firm-year observations after merging with the Compustat and CRSP databases (in Table 2). The number of firm-year observations decreases further to approximately 58,000 when the sample is merged with the product market threats data from Hoberg, Phillips, and Prabhala (2014) or the institutional ownership data from the Thomson Financial Institutional Holdings (13F) data.

We create our main speculation measure, *Speculation*, based on the vector of our speculation cues. *Speculation* is the count of speculation cues (i.e., the sum of all elements in the speculation cues vector) in a given firm's 10-K filing divided by the total word count in the filing (in percentage). Throughout the paper, we focus on *Speculation* as our main variable of interest.

To give context for the speculation measure, we examine neighboring unigrams and bigrams, which are those that occur in the same paragraph that contains at least one speculation cue. Panel B of Table 1 presents the lists of the frequently mentioned neighboring unigrams and bigrams. For unigrams in Panel B(i) of Table 1, we only include words in the Loughran and McDonald (2011) master dictionary that are considered to add financial information content. We also identify the part of speech for each of the unique neighboring words and sort them by frequency.

The three columns in Panel B(i) of Table 1 list verbs, nouns, and adjectives or adverbs, respectively. In the list of verbs, the most frequently mentioned neighboring words are "will", "require(d)", "expected", and "estimated." These words are associated with a firm's discussion on upcoming but uncertain situations. Besides, "anticipate(d)", "assumed", "intended", "achieve", "increasing", and "projected" in lower ranks of the verb list also suggest similar context around an speculation cue in a paragraph. The most frequently mentioned noun is "plan"; "future" is nearly as common. These two words are also associated with forward-looking disclosures. The most frequently mentioned adjective or adverb was "approximately." In addition, the adjective or adverb list includes neighboring words that imply positive attributes of circumstances, for example, "effective", "able", "greater", "beneficial", "successful", and "favorable."

Panel B(ii) of Table 1 lists meaningful neighboring bigrams in 10-K paragraphs that contain the speculation cues. We analyze 5,597,740 unique pairs of neighboring words and compute the saliency score of each bigram in the paragraphs with the speculation cues relative to the paragraphs without such cues. We then classify the top 100 most salient bigrams by their content. Out of the top 100 bigrams, 41 bigrams are classified into innovation terms, forward-looking terms, terms to describe market conditions, and terms to describe firm value. The remainder refers to individuals, days, time periods, or generic terminology. Internet Appendix Table IA.1 presents the complete list of the top 100 neighboring bigrams that are overused relative to common language with the saliency screen.

The use of speculation cues might depend on personal writing styles of managers and corporate lawyers.[12] However, we find a significant fraction of the variation comes from time-series changes within a firm, even within the same topic in a firm's 10-Ks. As an example, we present excerpts from the MD&A sections of Technical Communications Corp's 10-K filings, which were released in different years and discussed "Liquidity and Capital Resources - Cash Requirements", where our speculation cues are underlined.

> ***Technical Communications Corp's 10-K in 2000***
> Cash and cash equivalents increased by $783,000 or 33% to $3,122,000 as of September 30, 2000, from a balance of $2,339,000 at October 2, 1999. This increase was primarily due to the reduction of accounts receivable, which were partially offset by operating losses and a reduction in current liabilities.

> ***Technical Communications Corp's 10-K in 2010***
> It is anticipated that cash from operations will fund our near-term research and development and marketing activities. We also believe that, in the long term, based on current billable activities and the improvement in business prospects, cash from operations will be sufficient to meet the development goals of the Company, although we can give no assurances.

> ***Technical Communications Corp's 10-K in 2015***
> We believe that our overall financial condition remains strong. Our cash, cash equivalents and marketable securities at October 3, 2015 totaled $3,709,000 and we continue to have no long-term debt. It is anticipated that our cash balances and cash generated from operations will be sufficient to fund our near-term research and development and marketing activities. We believe that the combination of existing cash, cash equivalents, and highly liquid short-term investments, together with future cash to be generated by operations, will be sufficient to meet our ongoing operating and capital expenditure requirements for the foreseeable future and at least through the end of fiscal year 2016. We also believe that, in the long term, an anticipated improvement of business prospects, current billable activities and cash from operations will be sufficient to meet the Company's investment in product development, although we can give no assurances.

The company discusses its liquidity and cash requirements using no speculative language in 2000

[12] Dzieliński, Wagner, and Zeckhauser (2021) examine CEO communication with analysts and investors during quarterly earnings conference calls and show that CEO clarity is a matter of personal style.

but increases its usage of speculative language in 2010 and more significantly so in 2015. Sentences containing speculation cues tend to provide a company's anticipation of future positive prospects and to assure its shareholders that cash from operations will be sufficient. The company's expression of forward-looking information and positive possibility often accompanies by speculative language in 2010 and 2015 in contrast with its straightforward numerical description of the company's current cash situation in 2000.

We expect that the speculation measure based on 10-K disclosures — which are required by Regulation S-K to include any information with material effects on the firm's financial condition or results of operations, are carefully curated by the firm's legal team, and should be also audited by external auditors — is likely different from a similar measure based on other source texts that do not have the same degrees of difficulty of censoring and ex-ante scrutiny (e.g., the question and answer portion of earnings conference calls). Because of a high degree of care in preparing 10-Ks by the firm, the speculative language in 10-Ks is more deliberate than in other source texts. Thus, we expect our speculation measure based on 10-K disclosures to contain genuine information that is not possible to be made certain at the time of the disclosure because of market conditions or timing. This information can be distinctively useful from the standpoint of investors in evaluating the likely consequences of conditions that the firm faces.

Furthermore, the 1995 Private Securities Reform Act provides a safe harbor provision that protects companies against frivolous and abusive lawsuits when disclosing forward-looking information in their financial statements. Under this safe harbor provision, carefully chosen cautionary statements are more likely to be added around forward-looking statements. To mitigate a potential concern that the speculation measure merely picks up boilerplate warnings that are not sufficient as meaningful disclosures of value-relevant information, we conduct a robustness check in Section 5.2 by excluding paragraphs deemed to be boilerplate warnings in 10-Ks.

# 4 Validation and Relation to Firm Characteristics

In this section, we evaluate which firms employ more speculative language in their 10-Ks than others under what situations. These analyses provide useful validation of our speculation measure before we examine various market reactions to it in Section 5.

## 4.1 Summary Statistics

Table 2 presents the summary statistics for various textual tonal variables and non-tonal firm characteristics. Each variable is winsorized at the top and bottom 1% of its distribution and their detailed definitions are provided in Appendix B.[13]

[Insert Table 2 Here]

For the textual tonal variables, we consider our speculation measure (*Speculation*), existing textual tonal variables based on the master dictionary by Loughran and McDonald (2011) (Sentiment, Uncertain, Modal, Constraining, Litigious, Superfluous, and Interesting), and Fog words initially proposed by Robert Gunning in 1952 and used extensively in the literature to quantify the lack of plain English (e.g., Li (2008)). All textual tonal variables are expressed in percentages. The mean and median of *Speculation* are 1.311% and 1.366%, respectively. The average of Sentiment, defined as the difference between the percentages of positive words and negative words (out of total words), is -0.679%, indicating that negative sentiment dominates positive one in our sample of 10-K filings. On average, 32% of words are considered complex (Fog) words, and uncertain or litigious words are mentioned as many times as our speculation cues.

As for non-tonal firm characteristics, the average market value of assets is approximately $1.483 billion (Size in logarithm), and the average of firm age (Age in logarithm) is roughly 13 years. We include two growth opportunities proxies, Tobin's Q and Sales growth, whose means are 1.993 and 9.5%, respectively. We also consider two proxies for the economic conditions that firms face. Those are product market competition measured by Hoberg, Phillips, and Prabhala (2014) (Product market fluidity) and financial constraints measured by Kaplan and Zingales (1997) (Financial constraints (KZ)). Alternatively, using a textual measure of financial constraints from Hoberg and Maksimovic (2015) that covers the sample period until 2016, we find similar results. For our main analysis of how markets react to the speculation measure in Section 5, we include share turnover (Turnover), book-to-market ratio (Book-to-market), percentage of institutional investors' holdings

[13]We include common stocks (share code of 10 or 11) listed on the NYSE/AMEX and Nasdaq with average daily prices between $3 and $1000 and with at least 60 observations of daily return and volume from CRSP over the past 12 months before 10-K filing dates. We exclude stock days with trading volumes below 100 shares or with returns below -1.0, or stock months with less than 15 days of valid return and volume data, and stock days with prices less than $3 to mitigate potential concerns of the bid-ask bounce. These filters follow those in Jegadeesh and Wu (2013) and Amihud and Noh (2020). Our findings are insensitive to employing other price cutoffs ranging from $1 to $5.

(Institutional ownership), risk-adjusted return before 10-K filing (Fama-French alpha), and Filing-day abnormal return as control variables.

## 4.2 Relations to Other Textual Variables and Firm Characteristics

Although the speculative language is distinct from uncertainty, we expect *Speculation* to be positively related to textual indicators of uncertainty (e.g., uncertainty and weak modal terms). We validate this intuition by using uncertainty keywords, and weak and strong modal keywords from the master dictionary by Loughran and McDonald (2011). Portraying a series of univariate comparisons, Figure 2 presents sets of side-by-side box plots for the usage of speculative language in 10-Ks by whether uncertainty and modality are above versus below the median.

[Insert Figure 2 Here]

These side-by-side boxplots in Figure 2 indicate that speculative language in 10-Ks is more commonly used with more uncertainty words and more modality words. In addition, they show that there is substantial overlap in the distributions of the speculation cues for high and low uncertainty or modality, implying that there is useful residual variation in our speculation measure when holding the other textual tonal measures constant.

Next, we regress *Speculation* on a set of existing textual tonal measures, where all variables are contemporaneous. The first two columns of Table 3 report the estimation results of this regression model where we also control for firm and year fixed effects. To account for potential serial and cross-sectional correlations in the speculation measure, the standard errors are clustered by firm and year.

[Insert Table 3 Here]

In Column (1) of Table 3, we include as independent variables Fog to quantify the complexity of 10-K disclosures and Uncertain and Modal from Loughran and McDonald (2011) to quantify the uncertainty. Column (1) shows that complexity, uncertainty, and modality are all positively associated with *Speculation*, as expected, even when controlling for unobserved firm characteristics by including firm fixed effects. In Column (2), we additionally examine the relations of speculative language to Sentiment and four other textual tonal variables, Constraining, Litigious (which capture the firm's discussion on constraining and litigious situations in particular), Superfluous, and Interesting from Loughran and McDonald (2011). We find that *Speculation* is positively associated with Sentiment and all four other textual measures. In particular, the positive relation between speculation and sentiment, albeit statistically weak, confirms our contextual analysis that firms are more likely to have positive tones when they insert speculative cues into their 10-Ks. Taken together, this evidence is consistent with the idea that speculative language in 10-Ks captures a relatively positive tone with high uncertainty and high modality. Although positivity, uncertainty, and modality aspects of the text are, to a large degree, parts of the content of speculative language, we show later in our subsequent price-reaction tests of Section 5 that our findings are robust to controlling for all these measures plus two other proxies for complexity and readability. This shows that *Speculation* is distinct from existing textual measures.

Next, we relate *Speculation* to lagged non-tonal firm characteristics. For example, speculative language may be more frequently used by firms when they face greater growth opportunities that are difficult to quantify at the moment of disclosure. We first illustrate graphically which non-tonal firm characteristics (among notable ones) are related to the use of speculative language in 10-Ks. Figure 3 presents the 95% confidence intervals for the means of Size, Age, and two proxies for growth opportunities (Tobin's Q and Sales growth) by each quartile of the distribution of our speculation measure. From Figure 3, we find strong patterns that smaller and younger firms (Figures 3(a) and 3(b)) and firms that are likely to have more growth opportunities (Figures 3(c) and 3(d)) tend to use more speculative language in their 10-Ks.

[Insert Figure 3 Here]

We then investigate the associations with those firm characteristics by estimating the regression models in the last three columns of Table 3. As in Columns (1) and (2), all regression models include firm and year fixed effects, and standard errors are clustered by firm and year to account for potential serial and cross-sectional correlations. In Column (3), we observe a strong negative association between *Speculation* and Age in Figure 3 is also present in the regression analysis. This supports the idea that young firms have incentives to disclose positive but immature information early by hedging the action of disclosing it with speculative language. This finding is robust to controlling for one-year lagged product market threats and financial constraints (Column (4)) and Turnover, Institutional ownership, Fama-French alpha, and Filling-day abnormal return (Column (5)), which are all pre-filing variables and known to affect firms' returns on the event days of 10-K releases.[14] Overall, the various test results in this section validate our speculation measure, which sharpens the interpretation of our market reaction test results in the next section.

# 5 Reactions to Speculative Language

## 5.1 Price Reaction

This section investigates the relation between speculative language in 10-Ks and subsequent stock returns. Specifically, for each 10-K release, we compute the buy and hold abnormal returns (BHARs) over various estimation windows based on four-week intervals around the filing date and test whether Speculation predicts abnormal returns using the following regression specification. For stock $i$, over the $n$-week period around its 10-K filing in year $t$,

$$BHAR_{itn} \quad = \quad \alpha_n + \beta_n Speculation_{it} + \eta_n^{'} \mathbf{X}_{it} + \gamma_n RMkt_{tn} + \epsilon_{itn}. \tag{1}$$

$BHAR_{itn}$ is defined as the cumulative return difference between stock $i$ and the CRSP value-weighted index relative to Week 0 (including the 10-K filing day) up to the $n$th-week, where $n = \{-4, 0, 4, 8, 12, 16\}$.[15] $Speculation_{it}$ is the percentage of our speculation cues (out of the total words) in the 10-K disclosure, and $\mathbf{X}_{it}$ is a column vector of our baseline control variables based on prior studies (e.g., Loughran and McDonald (2011)) including Sentiment, Market value, Book-to-market, Turnover, Institutional ownership, Fama-French alpha, and Filing-day abnormal return. $RMkt_{tn}$ is the cumulative market return to capture common economic shocks that affect all individual stocks over the same $n$-week period as $BHAR_{itn}$. All of the baseline explanatory variables are constructed based on the information available as of the 10-K filing date, and their

[14] We do not include Market value and Book-to-market in Column (5) because their inclusion is redundant due to Size and Tobin's Q, but they are included in the subsequent market reaction tests.

[15] We use Week 0 (i.e., four days between the 10-K filing day and three days later) as the reference period to define the following cumulative BHARs before and after Week 0. For the post-filing period, each cumulative BHAR is computed over the period from the start of the 1st week (i.e., the fourth day after the 10-K filing) to the end of the $n$th week, where $n = 4, 8, 12, 16$. For the pre-filing period, each "reverse" cumulative BHAR is computed over the period from the end of Week -1 (i.e., one day before the 10-K filing date) to the start of the $n$th week, where $n = -4$. That is, the reverse cumulative BHAR for Week[-4,-1] is obtained by solving $(1 + BHAR_{[-4,-1]}) = \frac{1}{(1+BHAR_{-4})*(1+BHAR_{-3})*(1+BHAR_{-2})*(1+BHAR_{-1})}$ for $BHAR_{[-4,-1]}$, where $BHAR_{-k}$ is the weekly BHAR over Week $-k$.

detailed definitions are provided in Appendix B. For ease of interpretation, explanatory variables are standardized to have the mean of zero and the standard deviation of one. We estimate Equation (1) separately over the four-week period prior to each 10-K release date (i.e., Week[-4,-1]), the four-day period starting from each 10-K filing date (i.e., Week 0), and the expanding 16-week period by four weeks after each 10-K release date (i.e., Week[1,4], Week[1,8], $\cdots$, Week[1,16]). We cluster standard errors by firm and filing year-month to account for potential serial and cross-sectional correlations of abnormal returns, respectively.[16]

The coefficient of interest in Equation (1) is $\beta_n$ which captures how each stock's price in the $n$th week cumulatively reacts to the speculative language used in its 10-K disclosure. As predicted by Scenario D in Figure 1, if the speculative language in 10-Ks contains positive value-relevant information that takes time for investors to digest due to the uncertainty in it, we expect positive and delayed price reactions to it after 10-K release dates. In contrast, as in Scenario E of Figure 1, if the speculative language serves managers' obfuscation purposes when releasing negative value-related information, we expect negative and delayed price reactions.

[Insert Table 4 Here]

In Table 4, we present the estimation results of Equation (1). We find a significant and positive price reaction between Week 1 and Week 8 weeks after 10-K filings. After Week 8, the positive price reaction wanes and stabilizes (i.e., no return reversal). In the pre-period, there is no detectable price reaction to Speculation (i.e., no pre-trend or anticipation). The economic magnitude of the positive return effect in Column (6) translates into about 1% increase in BHAR over the 16-week period after 10-K releases with a one standard deviation increase in Speculation (=0.47%). In Appendix Table A.1, we present the estimation results of Equation (1) using BHAR for each four-week window as the dependent variable instead of the cumulative BHAR. These results are consistent with those in Table 4 and show more clearly that the positive price reaction wanes after Week 8 and there is no reversal afterward.

To evaluate whether there is an immediate price reaction to speculative language, we employ a slightly different specification for the price reaction during Week 0. In this regression, we drop

[16]Consistent with the conventional wisdom that serial correlation is typically low for returns, we find that the statistical significance without clustering by firm is nearly identical to that with the double clustering by firm and filing year-month.

the control variable of *Filing-day abnormal return* because this term is virtually the same as the dependent variable, but we keep other aspects of the specification in Equation (1) the same. The results are reported in Column (2) of the table and show no immediate reaction to speculative language. In subsequent tables where we perform robustness and heterogeneity analysis in addition to this main result, we adopt this same convention to drop the *Filing-day abnormal return* control variable.

As a complement to this main test, we show the results graphically using weekly estimation intervals in Figure 4, which plots the coefficient estimates of Speculation for cumulative BHARs up to the 16th week together with 95% confidence intervals to indicate statistical significance. See also Internet Appendix Table IA.2 for a tabular presentation of these results. Figure 4 shows explicitly that the positive speculation effect is not preceded by any pre-trend for five weeks prior to the 10-K filing, nor followed by a reversal through 16 weeks after Week 8.

These results all together support the idea that speculative language in 10-Ks provides positive value-relevant information about firms and investors digest the information over the subsequent eight-week period, which is delayed due to the embedded uncertainty in Speculation. They also imply that the information channel of Speculation dominates its uncertainty channel as predicted by Scenario D in our conceptual framework, and contrast with the view that managers strategically employ speculative language in 10-Ks to obfuscate market participants.

[Insert Figure 4 Here]

## 5.2 Decomposition of Speculation Measure and Robustness

In this section, we distinguish the price reaction to Speculation from the reactions to other textual measures. Specifically, we decompose our measure of speculation into components related to uncertainty, modality, and complexity, control for the forward-looking measure, and examine the unique aspects of our speculation measure beyond those existing textual measures. Table 5 shows the results.

[Insert Table 5 Here]

In Panel A of Table 5, we create a version of Speculation that is purged of all words in the dictionaries for uncertainty and modality and all words identified as complex according to the Fog index. Then, we recount how many times those purged speculation cues are mentioned in a given firm's 10-K filing in a given year scaled by the total word count in the filing (in percentage). The results in Panel A show that our main findings are similar but have smaller magnitudes using the purged speculation measure. For example, the magnitude of the positive cumulative return effect up to Week 16 (i.e., Column (6)) is about a half of that in Table 4. The results in Panel A of Table 5 well support our conjecture that our measure of speculation language captures unique variation that the existing uncertainty, modality, and complexity measures do not capture.

In Panel B, we take a different approach to show the robustness to other textual measures. Instead of purging the dictionaries of common terms, we use the original speculation measure but additionally control for the measures of uncertainty, modality, 10-K readability (Loughran and McDonald (2014)), and complexity. We find that the coefficient estimates for speculation stay positive and significant with larger magnitudes than what we found in Panel A. Interestingly, we find that the modality and 10-K readability are negatively associated with cumulative BHARs up to Week 8, which is consistent with the obfuscation view in Scenario E of our conceptual framework. These results highlight the empirical relevance of our speculation measure and provide its sharp economic distinction from related concepts.

We further decompose the speculation dictionary into unigrams and non-unigrams (bigrams and trigrams) and reconstruct our speculation measure only based on bigrams and trigrams. In this way, we mitigate the concern that our measure can be recreated with some combinations of existing textual measures that mainly consist of unigrams and also confirm that the bigrams and trigrams in our speculation dictionary are particularly useful. Panel C shows the results. We again find that the speculative language in 10-Ks leads to a positive and delayed price reaction over the post-filing periods.

Next, in Panel D, we address the concern that our measure simply picks up forward-looking statements. Although the timing of the information (i.e., whether the event is either current or future) is one dimension of our conceptual framework in Section 2, a manager's decision to use speculative language is not entirely driven by the timing, and thus, should not be fully captured by forward-looking statements. To evaluate the possibility that forward-looking statements might drive our findings, we estimate a regression model that additionally controls for the forward-looking measure based on Muslu, Radhakrishnan, Subramanyam, and Lim (2015). This measure is defined as how many times the forward-looking disclosure keywords from their paper are mentioned in a given firm's 10-K filing in a given year scaled by the total word count in the filing (in percentage). The results in Panel D show that, apart from separating out the effects of Speculation and Forward-looking, the non-forward-looking component of Speculation captures the disclosures that are speculative for other reasons; for example, disclosures of proprietary information about R&D. We further investigate this point in the next section.

Extending the analysis on forward-looking disclosure, we next take the sentiment in forward-looking statements into account. In the motivating framework, both timing and direction of the information (i.e., whether the signal is either positive or negative) play critical roles in the manager's use of speculative language. We therefore examine whether the sentiment in forward-looking statements drives our findings. First, we classify each paragraph into forward-looking and non-forward-looking groups by checking the existence of forward-looking disclosure keywords from Muslu, Radhakrishnan, Subramanyam, and Lim (2015). Then, we calculate the average sentiment score of the forward-looking group in a given firm's 10-K filing in a given year. Specifically, FW-Sentiment is the difference between the numbers of positive words and negative words in all paragraphs that belong to the forward-looking group scaled by the total word count in the filing (in percentage). Panel E shows the estimation results when FW-Sentiment is included additionally in the regression. We find similar results even with FW-Sentiment controlled. Lastly in Panel F, we decompose FW-Sentiment into its positive and negative components and obtain similar results.

We conduct several additional robustness checks, which are reported in Appendix Table A.2. First, to ensure that the significant positive and delayed price reactions to the speculative language in 10-Ks are not driven by future earnings announcements near 10-K filing dates, we re-estimate Equation (1) based on a refined subsample after we exclude all weekly BHAR observations having earnings announcements in the same week. The test results are presented in Panel A and show that the positive speculation effect is not mechanically related to upcoming future earnings announcements near 10-K filings. Second, we reconstruct Speculation by excluding the count of speculation cues in the paragraphs that contain boilerplate warnings related to safe harbor provisions, which lack meaningful information but can contain some repeated hedging language over years. We then repeat the estimation of Equation (1) using this alternative speculation measure. In Panel B, we find that our findings are robust to excluding the safe harbor paragraphs.

Third, to ensure that the positive speculation effect is not driven by exposure to systematic risks, we estimate Equation (1) after adding the sensitivities to Fama-French three factors estimated over the one-year period preceding each 10-K filing date as in calculating Fama-French alpha. In Panel C, we find that our findings are not explained by the risk exposures to Fama-French three factors. In addition, to control for a quickly varying market risk as in Patton and Verardo (2012), we include Filing-day abnormal return as a control variable in Equation (1) for all tests that we conduct. All these tests confirm that the positive speculation effect is not driven by proxies for systematic risks. Fourth, to ensure that our results are not sensitive to how to compute abnormal returns, we estimate Equation (1) with cumulative abnormal return ($CAR_{itn}$) as the dependent variable instead of $BHAR_{itn}$.[17] In Panel D, we obtain a similar positive speculation effect using CAR.

Fifth, we demean Speculation by firm and re-estimate Equation (1) with the firm-demeaned measure of Speculation to examine whether our measure has sufficient time-varying information. Panel E shows that positive and significant return effects of speculation are not merely cross-sectional. Lastly, in Panel F, we decompose the sentiment into its positive and negative components, and find that positive words indeed show positive return reactions, but our results are robust to the separate control of the positive and negative words.

### 5.3 Heterogeneity in Price Reactions

In this section, we shed additional light into mechanisms underlying the positive and delayed price reactions to Speculation via several heterogeneity tests. Specifically, we evaluate heterogeneity in the price reactions to Speculation by forward-looking statements, disclosure of proprietary information, or information environment of firms. To do so, we estimate Equation (1) using above-median vs. below-median sample splits for each of the following four firm-specific variables: Forward-looking disclosure, R&D disclosure, pre-filing idiosyncratic return volatility, and pre-filing analyst coverage. These sample splits are constructed such that they do not depend on any future informa-

[17] Fama (1998) advocates CAR and argues that BHAR exacerbates the "bad-model problems" by compounding an expected-return model's problem in explaining short-term returns. In contrast, Barber and Lyon (1997) advocates BHAR.

tion by determining the two groups of each variable based on the information available as of 10-K filing dates. For the pre-filing analyst coverage, we use the absence vs. presence of analysts. We also control for uncertainty, modality, 10-K readability, and complexity as in Panel B of Table 5. Table 6 reports the subgroup estimation results, and Figure 5 provides their graphical representations.

[Insert Table 6 and Figure 5 Here]

First, we evaluate whether the positive price reaction is stronger for firms that release more forward-looking disclosures based on the measure from Muslu, Radhakrishnan, Subramanyam, and Lim (2015). In Panel A of Table 6, we find that the price reaction to speculative language is much stronger for firms that employ more forward-looking terms. The positive coefficient estimates of Speculation are significantly present after 10-K filing dates only in the high forward-looking disclosure group, while those in the low forward-looking disclosure group are insignificant over any estimation windows. Reading Column (6), a one standard deviation increase in Speculation for the high forward-looking disclosure group is associated with about 1.72% higher BHAR over the 16-week period after 10-K filings. This estimated magnitude is roughly twice the magnitude for the whole sample at 0.83% (Panel B of Table 5).

Second, we examine the role of proprietary information using the R&D disclosure measure based on Merkley (2014). This measure is the percentage of narrative R&D disclosure keywords (out of total words) in each firm's 10-K. In Panel B of Table 6, the estimated price reaction to Speculation is strongest for firms that disclose above-median R&D-related terms. For example, in Column (6), a one standard deviation increase in the use of speculative language for the high R&D disclosure group is associated with nearly 1.50% higher BHAR over the 16-week period after 10-K filings.

Furthermore, we investigate how the speculative language in 10-Ks relates to subsequent R&D investments and present the results in Appendix Table A.3. Using overall investments including both R&D and capital expenditures in Column (1) and only R&D expenditures in Column (2), we first find that our speculation measure is significantly and positively associated with the next year's investments. We further note that this positive association between speculation and subsequent investments strengthens significantly for firms with high growth opportunities proxied by Tobin's Q. In contrast, we find no such association for fixed asset investments, which is proxied by capital expenditures or total assets. These results suggest that the manager of a firm knows about the upcoming positive R&D investment opportunities and releases her private information about them to the market using speculative language before making the real investments.[18] These findings corroborate our bigram analysis in Panel B(ii) of Table 1, which show that many neighboring words near speculation cues are connected to the disclosure of proprietary information and also are consistent with the motivating example of Moderna's early-stage mRNA prospects. Overall, the above results based on forward-looking and R&D disclosures are consistent with the auxiliary prediction that we made in Section 2.

Third, we investigate whether the positive price reaction to speculative language is affected by the difficulty in processing new information. Specifically, we split the sample based on the pre-filing idiosyncratic return volatility (Panel C) and the pre-filing absence vs. presence of analyst coverage (Panel D). The pre-filing idiosyncratic return volatility is the standard deviation of daily residual returns under the Fama-French three-factor model estimated over the period from the beginning of the prior month to six days (inclusive) before each 10-K filing. The pre-filing absence or presence of analyst coverage is determined by the availability of analysts over the period from the beginning of the previous two quarters to six days (inclusive) before each 10-K filing. In these tests, we observe the significant price reaction only for firms with a poor information environment (i.e., with above-median pre-filing idiosyncratic return volatility or without analyst coverage). For example, from Column (6) in Panel D, firms with no analyst coverage have about 1.59% higher BHAR over the 16-week period after 10-K filings. The estimated magnitudes for the no pre-filing analyst group are comparable to those for the high R&D disclosure group in Panel B.

## 5.4 Liquidity Reaction

In this section, we examine whether and how the use of speculative language in 10-Ks relates to stock illiquidity over the same post-filing event window as that for returns. If the positive price reaction to Speculation is primarily driven by the release of value-relevant information and investors incorporate it gradually, we expect a negative relation between our speculation measure and firms' illiquidity after 10-K filings with no pre-trend. This prediction is consistent with a traditional view

[18] An alternative interpretation of the positive coefficient of the interaction term of Speculation and Tobin's Q follows Chen, Goldstein, and Jiang (2007) and Goldstein and Yang (2019) on feedback effects. That is, when the manager has "uncertain" private information on new R&D investment opportunities, she releases the information to the market early using speculative language to seek for feedback from the market, which can be potentially useful to lower the degree of uncertainty and thus to make better subsequent real investment decisions.

that more disclosures improve stock liquidity, e.g., Goldstein and Yang (2017).[19] To test it, we employ the following regression specification: For stock $i$, over the four-week window ending in the $n$th week around its 10-K filing in year $t$,

$$Spread_{itn} = \alpha_n + \beta_n Speculation_{it} + \eta_n^{'} \mathbf{X}_{it} + \epsilon_{itn}. \tag{2}$$

$Spread_{itn}$ is the level of the quoted relative bid-ask spread, which is the average of daily ratios of quoted bid-ask spread to the bid-ask midpoint from the CRSP database in logarithm, over the four-week window ending in the $n$th week where $n = 4, 8, 12, 16$ for the post-filing period and over the four-week window from Week -4 to Week -1 (i.e., Week[-4,-1]) for the pre-filing period. For Week 0, we employ the same definition used for the BHAR regression in Equation (1). $Speculation_{it}$ is the percentage of speculation cues (out of total words) in firm $i$'s 10-K, and $\mathbf{X}_{it}$ is a column vector of our baseline control variables as in Table 4 and two additional variables, Nasdaq dummy and Pre-filing spread. All explanatory variables in Equation (2) are constructed based on the information available as of the 10-K filing date, and their detailed definitions are provided in Appendix B. For ease of interpretation, explanatory variables are standardized to have the mean of zero and the standard deviation of one. In estimating Equation (2), we include firm and filing year-month fixed effects to control for unobserved heterogeneity in illiquidity level across firms and a secular reduction in market illiquidity over time, respectively. We cluster standard errors by firm and filing year-month to account for the serial and cross-sectional correlations of illiquidity levels. This is importance since illiquidity is persistent over time (Amihud (2002) and Bali, Peng, Shen, and Tang (2014)) and also has a commonality across assets (Chordia, Roll, and Subrahmanyam (2000)).

The coefficient of interest in Equation (2) is $\beta_n$, which captures how speculative language in 10-Ks relates to illiquidity over each four-week window around 10K filing. The estimation results of Equation (2) are presented in Table 7. We also present a graphical representation of the results in Figure 6.

[Insert Table 7 and Figure 6 Here]

[19] Alternatively, if the speculative language in 10-Ks reflects uncertainty more than information, we expect a positive relation between our speculation measure and subsequent illiquidity, which can also explain the positive price reaction in Sections 5.1 to 5.3. This alternative explanation is consistent with the literature that views the information uncertainty as information asymmetry or information processing costs. Therefore, the liquidity reaction test helps us evaluate the relative importance of information vs. uncertainty channels of speculative language.

We find that the coefficient for Speculation is negative and significant starting in Week 0, it stays so in the first eight weeks (with the most negative value in the first four weeks), and its magnitude decreases afterward. We also find no evidence of the pre-trend as the coefficient for Speculation in Week[-4,-1] is statistically insignificant. The largest improvements to liquidity occur in the first eight-week period after 10-K filings, which coincides with the time period when the stock price reactions are the strongest in Table 4. These results are consistent with the interpretation that the greater use of speculative language in 10-Ks is indeed associated with more value-relevant information about firms and that it is digested by investors in financial markets, leading to an improvement in liquidity after 10-K filings. They also reconfirm that the information channel of speculative language in 10-Ks plays a more important role than its uncertainty channel.

### 5.5 Informed Traders' Reactions

If the speculative language contains positive value-relevant information that takes time to digest, Speculation might predict informed purchases since the informed investors are likely to understand its positive value implications more quickly than the uninformed investors. To test this prediction, we employ the following specification. For stock $i$, over the four-week window ending in the $n$th week around 10-K filing in year $t$,

$$\textit{Probability of informed buying}_{itn} = \alpha_n + \beta_n \textit{Speculation}_{it} + \eta_n^{'} \mathbf{X}_{it} + \epsilon_{itn}. \tag{3}$$

$\textit{Probability of informed buying}_{itn}$ proxies the trading activity of informed buyers, which is the average of daily (posterior) probabilities of informed buying proposed by Brennan, Huh, and Subrahmanyam (2018)[20] over each four-week window following the same event window definitions as our earlier tests. $\textit{Speculation}_{it}$ is the percentage of speculation cues (out of total words) in firm $i$'s 10-K, and $\mathbf{X}_{it}$ is a column vector of our baseline control variables as in Table 4 and two additional variables, Nasdaq dummy and pre-filing probability of informed buying. All of the explanatory

[20]Based on the structural model of the probability of informed trading (PIN) developed by Easley, Kiefer, O'hara, and Paperman (1996), Brennan, Huh, and Subrahmanyam (2018) propose daily proxies for informed buying and selling activities at the stock level. For each stock and month, Brennan, Huh, and Subrahmanyam (2018) estimate the five parameters of the PIN model using a three-month rolling window and then calculate the daily posterior probability that a given trading day has good news or bad news by conditioning on the number of buyer-initiated and seller-initiated trades on each day. In that model, the informed encompasses not only traders who own private information but also those who possess superior information processing skills of publicly available information to others. For more details, see Section 1 of Brennan, Huh, and Subrahmanyam (2018).

variables in Equation (3) are constructed based on the information available as of the 10-K filing date, and their detailed definitions are provided in Appendix B. For ease of interpretation, explanatory variables are standardized to have the mean of zero and the standard deviation of one. We also include filing year-month fixed effects and cluster standard errors by firm and filing year-month. The coefficient of interest in Equation (3) is $\beta_n$, which captures how the speculative language in 10-Ks affects the informed buying activity of corporate outsiders over each four-week window after 10-K filings. The results from estimating Equation (3) are presented in Table 8 with a graphical representation in Figure 7.

[Insert Table 8 and Figure 7 Here]

We find that the estimated coefficient for our speculation measure becomes positive and significant from Week 0 (including the 10-K filing day), has its highest value in Week[1,4], and then decreases gradually afterward. In these results, we note that the informed buyers' reactions to the speculative language in 10-Ks emerge earlier than the rise of stock prices as shown in Tables 4. This timing is consistent with the idea that the informed investors can understand the positive value implications of speculative statements in 10-Ks more quickly than the uninformed investors. We also find no evidence of the pre-trend as the coefficient of Speculation before 10-K filings in Week[-4,-1] is not statistically different from zero. These results of the informed buyers and Table 7 together imply that the speculative language in 10-Ks attracts more noise traders than the informed traders.

Separately, we use *Probability of informed selling*$_{itn}$, also proposed by Brennan, Huh, and Subrahmanyam (2018), as the dependent variable in Equation (3) and repeat the estimation. We present the results in Appendix Table A.4 with no significant association between Speculation and informed selling activities. These results support the interpretation that speculative language in 10-K filings contains good news that triggers buying activities among informed investors. Figure 7(a) and (b) graphically confirm the stark difference between informed buying vs. informed selling activities. We observe significant increases in informed buying activities in the first eight weeks, while the informed selling activities are muted throughout the estimation period. Taken together, these results support the interpretation that positive value-relevant information disseminates into markets through speculative language in 10-Ks and is digested by informed traders more quickly than other investors.

Next, we investigate how the informed buying activities by corporate insiders are associated with speculative language in 10-Ks by replacing the dependent variable in Equation (3) with *Dollar volume of insider buying*$_{itn}$, the dollar volume of insider buying in logarithm.[21] Different from the earlier tests on the probability of informed buying, we expect insiders' buying to be greater throughout all estimation windows because insiders have the access to the private information prior to 10-K filing dates. The test results are presented in Table 9 with a graphical representation in Figure 8.

[Insert Table 9 and Figure 8 Here]

We find that the estimated coefficients on Speculation are positive and significant throughout all estimation windows in Table 9. Interestingly, the coefficient on Speculation is the largest in magnitude in Week[1,4], over which the largest liquidity improvement also occurs in Table 7. This is suggestive evidence that informed corporate insiders time their buying trades when stock liquidity becomes most in favor of them after 10-K releases, consistent with Collin-Dufresne and Fos (2015) and Collin-Dufresne and Fos (2016).

In Appendix Table A.5, using *Dollar volume of insider selling*$_{itn}$, we conduct analogous tests and present the results of insider selling. Similar to informed selling activities, we find no significant association between our speculation measure and insider selling activities. Interestingly, in Week 0, we rather find the coefficient estimate for Speculation is negative and significant at the 10% level. Finding that corporate insiders are less likely to sell their shares in the period when they release corporate disclosures with more speculative language to the public indeed indicates that they aim to convey positive value-relevant information through speculative statements. From Figure 8(a) and (b) that graphically compare insider buying vs. selling activities, we can easily observe that insider selling related to speculative language is completely nonexistent around 10-K filing dates relative to insider buying related to speculative language. Taken together, these findings on insider trading activities reinforce our earlier conclusion that speculative language used in 10-K disclosures

[21] When analyzing this dependent variable, we drop Pre-filing informed buying from the vector of control variables. For firm $i$ and year $t$, we construct *Dollar volume of insider buying*$_{itn}$ by summing all records of insider purchases in dollar volume for each week, averaging the weekly quantity over the four-week window ending in the $n$th week, and taking its logarithm. The insider trading data are from Thomson Reuters. Results are consistent when using the turnover of insiders' buying as the dependent variable.

contains *positive* value-relevant information.

### 5.6 News Sentiment in the Subsequent Periods

We now relate Speculation to subsequent news sentiment, which can provide more direct evidence on whether the information embedded in speculative statements indeed leads to realized positive events. Specifically, we employ the following specification. For stock $i$, over the four-week window ending in the $n$th week around 10-K filing in year $t$,

$$News\ sentiment_{itn} = \alpha_n + \beta_n Speculation_{it} + \eta_n^{'} \mathbf{X}_{it} + \gamma_n NSMkt_{tn} + \epsilon_{itn}. \tag{4}$$

$News\ sentiment_{itn}$ is the average news sentiment score over each four-week window,[22] $Speculation_{it}$ is the percentage of speculation cues (out of total words) in firm $i$'s 10-K, and $\mathbf{X}_{it}$ is a column vector of our baseline control variables as in Table 4 and Pre-filing news sentiment. $NSMkt_{tn}$ is the four-week market news sentiment to proxy the common economic shocks that potentially affect all individual firms' news sentiment scores over the estimation window.[23] All explanatory variables in Equation (4) are standardized to have the mean of zero and the standard deviation of one, and their detailed definitions are provided in Appendix B. We cluster standard errors by firm and filing year-month to account for the serial and cross-sectional correlations in News sentiment, respectively. The results are presented in Table 10 with a graphical representation in Figure 9.

[Insert Table 10 and Figure 9 Here]

We find that the coefficient for Speculation is positive and significant in Week[1,4] and Week[5,8], indicating the arrivals of good news about the firm during the subsequent eight weeks after 10-K filings. We find no evidence of the pre-trend since the coefficient of Speculation before 10-K filings is not statistically different from zero. The evidence from news media sentiment implies that positive value-relevant information in Speculation is followed by the arrival of good news in the form of the positive sentiment from media, which supports our information-based explanation of the positive speculation effect.

[22]We use the same definition of estimation windows as those for the earlier regression models in Section 5. The daily news-related sentiment score is defined as (ESS-50)/50, where the ESS variable comes from the RavenPack News Analytics database for the period starting in January 2000. The news stories with Relevance=100 and Novelty≥75 are used for our analysis.

[23]The value of $NSMkt_{tn}$ depends on the actual dates of the estimation window around each 10-K filing. Our market-wide news sentiment is value-weighted, but our findings are robust to equal-weighting.

# 6 Conclusions

In this paper, we introduce a textual measure to the finance and accounting literature, which uniquely quantifies the degree of speculation in firms' disclosures with a minimum level of researcher subjectivity. Our measure of speculative language is distinct from existing textual measures such as sentiment, uncertainty, modality, complexity, readability, and forward-looking statements and has the ability to identify the unique qualitative information in firm disclosures beyond quantitative information.

We find strong evidence that firms tend to use more speculative language in their 10-Ks to communicate valuable positive information. Indeed, our heterogeneity and content analysis uncover that this speculative language is most valuable when the information is difficult to make precise at the time of disclosure, for example, because it pertains to the future or requires making statements about proprietary information. Thus, it is natural for firms to use speculative language in this information environment to deliver new information on positive yet uncertain prospects of future events. Market participants initially under-react to the information in speculative language due to the embedded uncertainty but they incorporate the information into prices eventually, which do not revert. Collectively, our findings and approach suggest that there is much more to learn from the qualitative content of textual disclosures beyond what sentiment, complexity, and readability can capture.

## Appendix B. Variable Definitions

| | |
|---|---|
| Speculation | is the number of speculation cues scaled by the total word count in the 10-K filing (in percentage). |
| Purged speculation | is the number of speculation cues, after taking out all words that overlap with the dictionaries for uncertainty, modality, and complex (Fog) words from our list of speculation cues, scaled by the total word count in the 10-K filing (in percentage). |
| Positive | is the number of positive words from the master dictionary by Loughran and McDonald (2011) scaled by the total word count in the 10-K filing (in percentage). |
| Negative | is the number of negative words from the master dictionary by Loughran and McDonald (2011) scaled by the total word count in the 10-K filing (in percentage). |
| Sentiment | is Positive minus Negative. |
| Uncertain | is the number of uncertain words from the master dictionary by Loughran and McDonald (2011) scaled by the total word count in the 10-K filing (in percentage). |
| Modal | is the number of (weak and strong) modal words from the master dictionary by Loughran and McDonald (2011) scaled by the total word count in the 10-K filing (in percentage). |
| Constraining | is the number of constraining words from the master dictionary by Loughran and McDonald (2011) scaled by the total word count in the 10-K filing (in percentage). |
| Litigious | is the number of litigious words from the master dictionary by Loughran and McDonald (2011) scaled by the total word count in the 10-K filing (in percentage). |
| Superfluous | is the number of superfluous words from the master dictionary by Loughran and McDonald (2011) scaled by the total word count in the 10-K filing (in percentage). |
| Interesting | is the number of interesting words from the master dictionary by Loughran and McDonald (2011) scaled by the total word count in the 10-K filing (in percentage). |
| Fog | is the number of words of three or more syllables that are not hyphenated words or two-syllable verbs made into three with -es and -ed endings, scaled by the total word count in the 10-K filing (in percentage). |
| LM-Readability | is the log of the file size in megabytes of the "complete submission text file" for the 10-K filing in the SEC EDGAR, proposed by Loughran and McDonald (2014). |
| Forward-looking | is the number of forward-looking disclosure keywords from Muslu, Radhakrishnan, Subramanyam, and Lim (2015) scaled by the total word count in the 10-K filing (in percentage). |
| FW-Sentiment | is the difference between the numbers of positive words and negative words in all forward-looking paragraphs scaled by the total word count in the 10-K filing (in percentage). Forward-looking paragraphs are all paragraphs that contain any forward-looking disclosure keyword from Muslu, Radhakrishnan, Subramanyam, and Lim (2015). |
| FW-Positive | is the number of positive words in all forward-looking paragraphs scaled by the total word count in the 10-K filing (in percentage). |
| FW-Negative | is the number of negative words in all forward-looking paragraphs scaled by the total word count in the 10-K filing (in percentage). |
| R&D disclosure | is the number of narrative R&D disclosure keywords from Merkley (2014) scaled by the total word count in the 10-K filing (in percentage). |

| | |
|---|---|
| Product market fluidity | is a 10-K based textual measure for the competitive threats faced by a firm in its product markets that captures the changes in rival firms' products relative to the firm, from Hoberg, Phillips and Prabhala (2014). |
| Financial constraints (KZ) | is the measure of financial constraints proposed by Kaplan and Zingales (1997). |
| Size | is the log of market value of total assets (market value of common equity plus book value of preferred stock, long-term and short-term debt, and minority interest) in a given year. |
| Age | is the log of one plus firm age in a given year based on its first appearance in Compustat. |
| Tobin's $Q$ | is the market value of assets divided by the book value of assets in a given year. |
| Sales growth | is the log of sales in a given year divided by sales in the prior year. |
| Market value | is the log of market value of equity, which is the number of shares outstanding times the price of the stock on the day before 10-K filing date. |
| Book-to-market | is the log of the book-to-market ratio using the book value from firm's annual report known as of the end of the previous fiscal year and the market value known as of December of the year before the year of the analysis. |
| Turnover | is the log of the volume of shares traded over the period from the beginning of the prior month to one day before the 10-K filing, divided by the number of shares outstanding at the end of the period. For each stock, at least 60 observations of daily volumes over the one-year period before the 10-K filing are required to be included in the sample. |
| Institutional ownership | is the percentage of institutional investors' holdings from the CDA/Spectrum database for the most recent quarter before the 10-K filing. The variable is treated as missing for negative values and winsorized to 100% for values above 100%. |
| Fama-French alpha | is the intercept estimated by regressing daily excess returns of a stock on daily Fama-French's three factors over the one-year period before the 10-K filing. For each stock, at least 60 observations of daily returns over the one-year period are required to be included in the sample. |
| Filing-day abnormal return | is either the BHAR or CAR over the 10-K filing-day window, *i.e.,* Week 0 that covers the four days between the 10-K filing day (inclusive) and three days later (inclusive). |
| Cumulative market return | is the cumulative re turn of the CRSP value-weighted index over the same $n$-week period around the 10-K filing as in BHAR. |
| Pre-filing idiosyncratic return volatility | is the standard deviation of daily residual returns by regressing daily excess returns of a stock on daily Fama-French's three factors over the period from the beginning of the prior month to six days (inclusive) before the 10-K filing. For each stock, at least 60 observations of daily returns over the one-year period ending at six days (inclusive) before the 10-K filing are required to be included in the sample. |
| Pre-filing analyst coverage | is the availability (absence or presence) of analysts following the firm over the period from the beginning of the previous two quarters to six days (inclusive) before the 10-K filing. |
| Pre-filing spread | is the log of the average of daily quoted relative bid-ask spread over the period from the beginning of the prior month to one day before the 10-K filing. |
| Nasdaq dummy | is one if the stock is listed in the Nasdaq on the day before the 10-K filing and zero otherwise. |
| Pre-filing informed buying (selling) | is the average of daily posterior probabilities of informed buying (selling) based on Brennan, Huh, and Subrahmanyam (2018) over the period from the beginning of the prior month to one day before the 10-K filing. |

| | |
|---|---|
| Pre-filing news sentiment | is the average of daily news sentiment scores over the period from the beginning of the prior month to one day before the 10-K filing, where the daily news sentiment score is defined as (ESS-50)/50 based on the RavenPack News Analytics database. |
| Market-wide news sentiment | is the average of daily market-wide news sentiment score over each four-week estimation window around the 10-K filing, where the daily market-wide news sentiment score is a value-weighted cross-stock average of individual stocks' news sentiment scores using their market capitalizations in the preceding month. |

## References


Acharya, V. V., DeMarzo, P., Kremer, I., 2011. Endogenous information flows and the clustering of announcements. *American Economic Review* 101, 2955–79.

Amihud, Y., 2002. Illiquidity and stock returns: cross-section and time-series effects. *Journal of financial markets* 5, 31–56.

Amihud, Y., Noh, J., 2020. Illiquidity and Stock Returns II: Cross-section and Time-series Effects. *The Review of Financial Studies* forthcoming.

Bali, T. G., Peng, L., Shen, Y., Tang, Y., 2014. Liquidity Shocks and Stock Market Reactions. *The Review of Financial Studies* 27, 1434–1485.

Bao, D., Kim, Y., Mian, G. M., Su, L., 2019. Do managers disclose or withhold bad news? evidence from short interest. *The Accounting Review* 94, 1–26.

Barber, B. M., Lyon, J. D., 1997. Detecting long-horizon abnormal stock returns: the empirical power and specification of test statistics. *Journal of financial economics* 43, 341–372.

Barry, C. B., Brown, S. J., 1985. Differential Information and Security Market Equilibrium. *Journal of Financial and Quantitative Analysis* 20, 407–422.

Bellstam, G., Bhagat, S., Cookson, J. A., 2020. A Text-Based Analysis of Corporate Innovation. *Management Science, Forthcoming.*

Bloomfield, R. J., 2002. The "incomplete revelation hypothesis" and financial reporting. *Accounting Horizons* 16, 233–243.

Brennan, M. J., Huh, S.-W., Subrahmanyam, A., 2018. High-frequency measures of informed trading and corporate announcements. *The Review of Financial Studies* 31, 2326–2376.

Bushee, B. J., Gow, I. D., Taylor, D. J., 2018. Linguistic complexity in firm disclosures: Obfuscation or information? *Journal of Accounting Research* 56, 85–121.

Chen, Q., Goldstein, I., Jiang, W., 2007. Price informativeness and investment sensitivity to stock price. *The Review of Financial Studies* 20, 619–650.

Chordia, T., Roll, R., Subrahmanyam, A., 2000. Commonality in liquidity. *Journal of financial economics* 43, 341–372.

Cohen, L., Lou, D., Malloy, C. J., 2020. Casting Conference Calls. *Management Science* 66, 5015–5039.

Cohen, L., Malloy, C., Nguyen, Q. H., 2016. Lazy prices. *Working Paper SSRN 1658471.*

Coles, J. L., Loewenstein, U., 1988. Equilibrium pricing and portfolio composition in the presence of uncertain parameters. *Journal of Financial Economics* 22, 279–303.

Collin-Dufresne, P., Fos, V., 2015. Do Prices Reveal the Presence of Informed Trading? *The Journal of Finance* 70, 1555–1582.

Collin-Dufresne, P., Fos, V., 2016. Insider Trading, Stochastic Liquidity, and Equilibrium Prices. *Econometrica* 84, 1441–1475.

Daniel, K. D., Hirshleifer, D., Subrahmanyam, A., 1998. Investor Psychology and Security Market Under- and Overreactions. *The Journal of Finance* 53, 1839–1885.

Daniel, K. D., Hirshleifer, D., Subrahmanyam, A., 2001. Overconfidence, Arbitrage, and Equilibrium Asset Pricing. *The Journal of Finance* 56, 921–965.

Dougal, C., Engelberg, J., Garcia, D., Parsons, C. A., 2012. Journalists and the stock market. *Review of Financial Studies* 25, 639–679.

Dye, R. A., 1985. Disclosure of nonproprietary information. *Journal of accounting research* pp. 123–145.

Dzieliński, M., Wagner, A. F., Zeckhauser, R. J., 2021. Ceo clarity. *Working paper.*

Easley, D., Kiefer, N., O'hara, M., Paperman, J., 1996. Liquidity, information, and infrequently traded stocks. *Journal of finance* 51, 1405–1436.

Ely, J. C., 2017. Beeps. *American Economic Review* 107, 31–53.

Fama, E. F., 1998. Market efficiency, long-term returns, and behavioral finance. *Journal of financial economics* 49, 283–306.

Ganter, V., Strube, M., 2009. Finding Hedges by Chasing Weasels: Hedge Detection Using Wikipedia Tags and Shallow Linguistic Features. In: *Proceedings of the ACL-IJCNLP 2009 Conference Short Papers* no. 2044. Association for Computational Linguistics pp. 173–176.

Gao, J., Liang, C., Merkley, K., Pacelli, J., 2017. Do lenders promote the revelation of bad news? Evidence from Lender-side loan defaults. *Working Paper.*

Garcia, D., 2013. Sentiment during recessions. *Journal of Finance* 68, 1267–1300.

Goldsmith-Pinkham, P., Hirtle, B., Lucca, D., 2016. Parsing the content of bank supervision. *Working Paper FRBNY Staff Report No. 770.*

Goldstein, I., Yang, L., 2017. Information Disclosure in Financial Markets. *Annual Review of Financial Economics* 9, 101–125.

Goldstein, I., Yang, L., 2019. Good disclosure, bad disclosure. *Journal of Financial Economics* 131, 118–138.

Gormley, T. A., Kim, B. H., Martin, X., 2012. Do firms adjust their timely loss recognition in response to changes in the banking industry? *Journal of Accounting Research* 50, 159–196.

Hanley, K. W., Hoberg, G., 2010. The information content of IPO prospectuses. *Review of Financial Studies* 23, 2821–2864.

Hirshleifer, D., 2001. Investor Psychology and Asset Pricing. *The Journal of Finance* 56, 1533–1597.

Hoberg, G., Lewis, C., 2017. Do fraudulent firms produce abnormal disclosure? *Journal of Corporate Finance, Forthcoming.*

Hoberg, G., Maksimovic, V., 2015. Redefining financial constraints: A text-based analysis. *Review of Financial Studies* 28, 1312–1352.

Hoberg, G., Moon, S. K., 2019. The offshoring return premium. *Management Science* 65, 2876–2899.

Hoberg, G., Phillips, G., Prabhala, N., 2014. Product market threats, payouts, and financial flexibility. *Journal of Finance* 69, 293–324.

Hoberg, G., Phillips, G. M., 2016. Text-based network industries and endogenous product differentiation. *Journal of Political Economy* 124, 1423–1465.

Huang, X., Teoh, S. H., Zhang, Y., 2013. Tone Management. *The Accounting Review* 89, 1083–1113.

Hwang, B.-H., Kim, H. H., 2017. It pays to write well. *Journal of Financial Economics, Forthcoming.*

Jegadeesh, N., Wu, D., 2013. Word power: A new approach for content analysis. *Journal of financial economics* 110, 712–729.

Jegadeesh, N., Wu, D., 2017. Deciphering Fedspeak: The information content of FOMC meetings. *Working Paper SSRN 2939937.*

Jones, D. A., 2007. Voluntary disclosure in r&d-intensive industries. *Contemporary Accounting Research* 24, 489–522.

Kamenica, E., Gentzkow, M., 2011. Bayesian Persuasion. *American Economic Review* 101, 2590–2615.

Kaplan, S. N., Zingales, L., 1997. Do investment-cash flow sensitivities provide useful measures of financing constraints? *The quarterly journal of economics* 112, 169–215.

Kilicoglu, H., Bergler, S., 2008. Recognizing speculative language in biomedical research articles: a linguistically motivated perspective. *BMC bioinformatics* 9, 1–10.

Kothari, S. P., Shu, S., Wysocki, P. D., 2009. Do managers withhold bad news? *Journal of Accounting research* 47, 241–276.

Li, F., 2008. Annual report readability, current earnings, and earnings persistence. *Journal of Accounting and Economics* 45, 221–247.

Li, Y., Yermack, D., 2016. Evasive shareholder meetings. *Journal of Corporate Finance* 38, 318–334.

Loughran, T., McDonald, B., 2011. When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks. *Journal of Finance* 66, 35–65.

Loughran, T., McDonald, B., 2013. IPO first-day returns, offer price revisions, volatility, and form S-1 language. *Journal of Financial Economics* 109, 307–326.

Loughran, T., McDonald, B., 2014. Measuring Readability in Financial Disclosures. *Journal of Finance* 69, 1643–1671.

Merkley, K. J., 2014. Narrative disclosure and earnings performance: Evidence from r&d disclosures. *The Accounting Review* 89, 725–757.

Muslu, V., Radhakrishnan, S., Subramanyam, K., Lim, D., 2015. Forward-looking MD&A disclosures and the information environment. *Management Science* 61, 931–948.

Noh, J., Zhou, D., 2022. Executives' Blaming external factors and market reactions: Evidence from earnings conference calls. *Journal of Banking & Finance* 134, 106358.

Patton, A. J., Verardo, M., 2012. Does Beta Move with News? Firm-Specific Information Flows and Learning about Profitability. *The Review of Financial Studies* 25, 2789–2839.

Ross, S. A., 1979. *Disclosure Regulation in Financial Markets: Implications of Modern Finance Theory and Signaling Theory* . Issues in Financial Regulation (McGraw-Hill New York).

Simmons, J. P., Nelson, L. D., Simonsohn, U., 2011. False-Positive Psychology: Undisclosed Flexibility in Data Collection and Analysis Allows Presenting Anything as Significant. *Pscyhological Science* 22, 1359–1366.

Skinner, D. J., 1994. Why firms voluntarily disclose bad news. *Journal of accounting research* 32, 38–60.

Solomon, D. H., 2012. Selective Publicity and Stock Prices. *The Journal of Finance* 67, 599–638.

Tetlock, P. C., 2007. Giving Content to Investor Sentiment: The Role of Media in the Stock Market. *Journal of Finance* 62, 1139–1168.

Verrecchia, R. E., 1983. Discretionary disclosure. *Journal of Accounting and Economics* 5, 179–194.

Zhang, X. F., 2006. Information Uncertainty and Stock Returns. *The Journal of Finance* 61, 105–137.

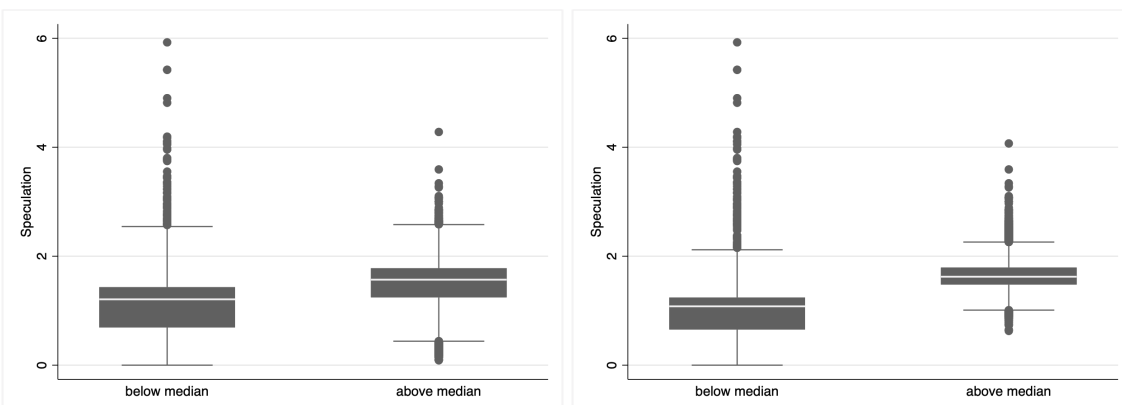
Figure 1: Manager's Disclosure and Language Choices and Price Reaction by Different Information Characteristics

This figure shows a conceptual framework that analyzes a manager's decision to disclose a certain type of private information, her choice of language in the disclosure, and a prediction on how the stock price reacts to the released information with the chosen language. The manager's decisions are jointly derived from the following three dimensions of the private information: i) timing of the event, ii) direction of the signal, and iii) strength of the signal. First, regarding the timing of the event, the manager can choose to discuss current or future events. Second, regarding the direction of the signal, the manager's private information can be on either positive or negative news of the firm. Third, regarding the strength of the signal, her private information can be based on either a strong or weak signal. Six possible disclosure scenarios (from Scenarios A to F) are provided based on the SEC's disclosure regulations and implications from existing theoretical and empirical studies.

| Scenario | Timing of Event | Direction of Signal | Strength of Signal | Disclosure Choice by Management | Language Choice by Management | Price Reaction by Market |
|---|---|---|---|---|---|---|
| A | Current | Positive | Strong | Mandatory: Yes | Non-speculative | Positive and immediate |
| B | Current | Negative | Strong | Mandatory: Yes | Non-speculative | Negative and immediate |
| C | Future | Positive | Strong | Voluntary: Yes<br>Verrecchia (1983) and Dye (1985)<br>Kothari, Shu, Wysocki (2008) | Non-speculative | Positive and immediate |
| D | Future | Positive | Weak | Voluntary: Yes<br>Verrecchia (1983) and Dye (1985)<br>Kothari, Shu, Wysocki (2008) | More speculative | Positive and delayed |
| E | Future | Negative | Strong | Voluntary: Yes<br>Skinner (1994) | Non-speculative or<br>Less speculative | Negative and immediate or<br>Negative and somewhat delayed |
| F | Future | Negative | Weak | Voluntary: No<br>Verrecchia (1983) and Dye (1985)<br>Kothari, Shu, Wysocki (2008)<br>Bao, Kim, Mian, Su (2019) | — | — |

Figure 2: Speculation versus Uncertainty and Modality

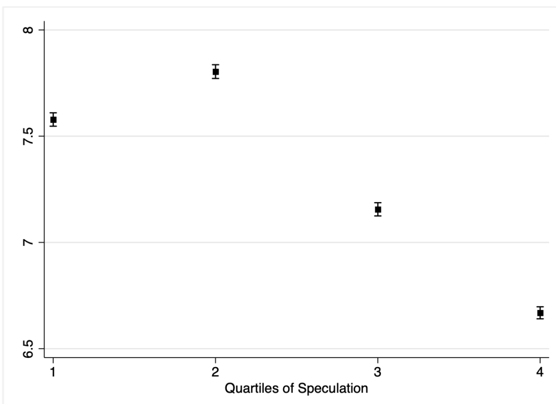
This figure shows the relation between (a) uncertainty words or (b) modality words (from the master dictionary by Loughran and McDonald (2011)) and our speculation cues. Each panel presents two side-by-side box plots for the distribution of the speculation measure by above and below the median of uncertainty or modality measure. Each box displays the interquartile range between the 25th to 75th percentiles of the distribution of the speculation measure, where the thick solid line inside the box displays the median. The top and bottom solid lines outside the box display the maximum and minimum, respectively, where the maximum and minimum are defined as the 75th percentile+1.5×the interquartile range and the 25th percentile−1.5×the interquartile range. Circles above and below those two solid lines represent outliers. The difference in medians for each panel is statistically significant at the 1% level.



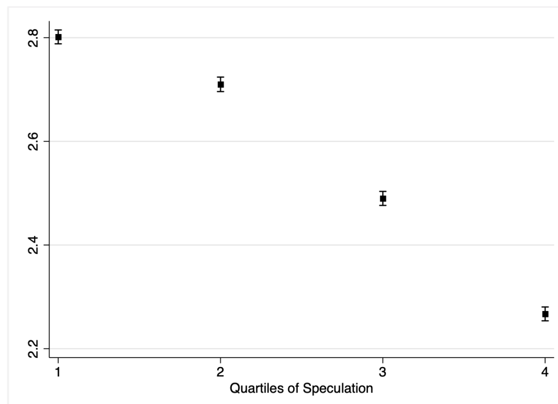

(a) Percentage of Uncertainty Words (b) Percentage of Modal Words

Figure 3: Speculation and Firm Characteristics

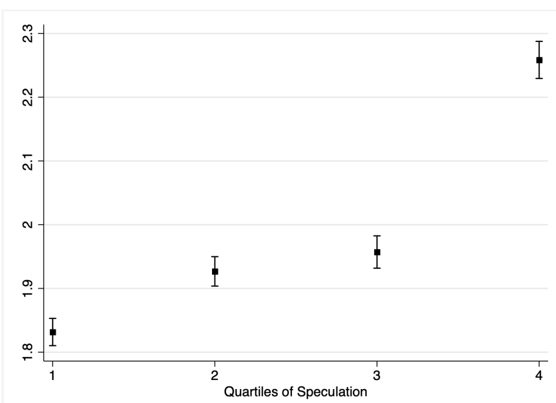
This figure shows the relation between each of notable firm characteristics and the propensity to use speculative language in 10-K disclosures. Each panel presents the 95% confidence interval for the mean of each firm characteristic for the first, second, third, and fourth quartile of the distribution of our speculation measure. We examine four firm characteristics including Size, Age, Tobin's Q, and Sales growth.



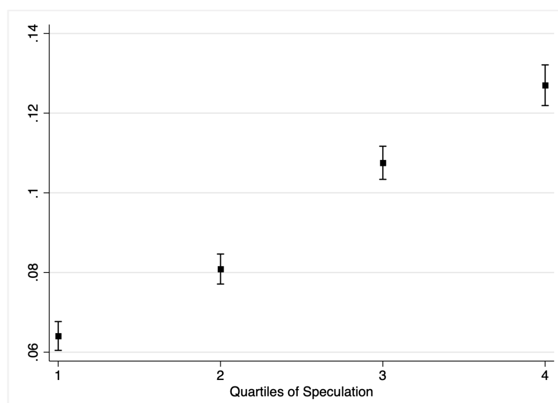

(a) Size




(b) Age




(c) Tobin's Q




(d) Sales growth

Figure 4: Speculation and Cumulative BHARs

This figure presents the coefficient estimates of our speculation measure (Speculation) in the regressions of cumulative buy and hold abnormal returns (BHARs) on Speculation over the $n$-week period around Week 0 (including the 10-K filing day), which are based on Internet Appendix Table IA.2. The shaded area represents the 95% confidence interval for each coefficient estimate (circle marker). For the post-filing weeks, each cumulative BHAR is computed over the period from the start of the 1st week to the end of the $n$th week ($n = 1, \ldots, 16$). For the pre-filing weeks, the "reverse" cumulative BHAR is computed over the period from the end of the $-1$th week to the start of the $n$th week ($n = -1, \ldots, -5$). The vertical solid line is located at Week 0 that covers the four days between the 10-K filing day and three days later. The coefficient estimates and confidence intervals are presented in percentage.

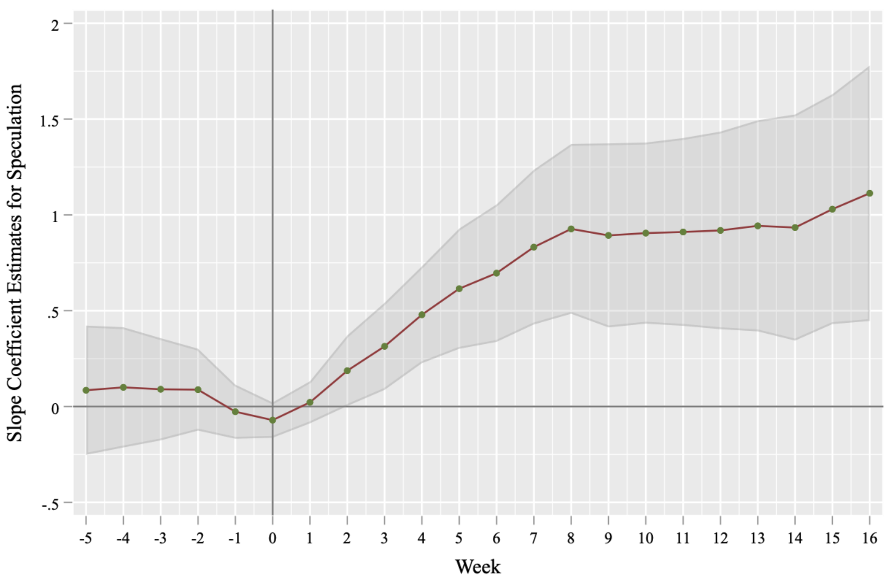
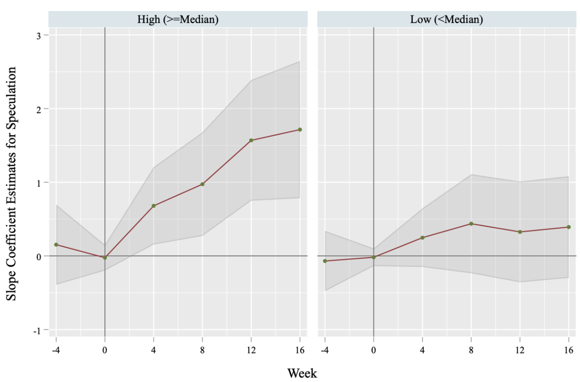
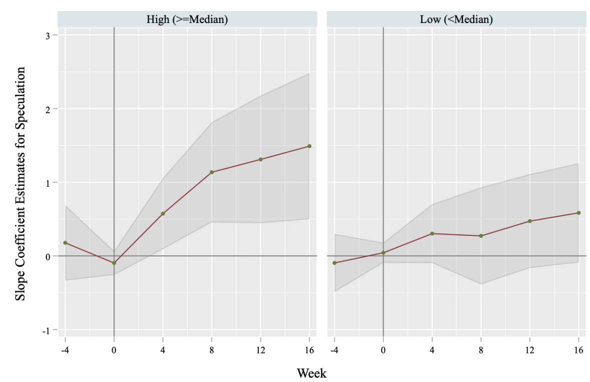


Figure 5: Speculation and Cumulative BHARs by Heterogeneous Variables

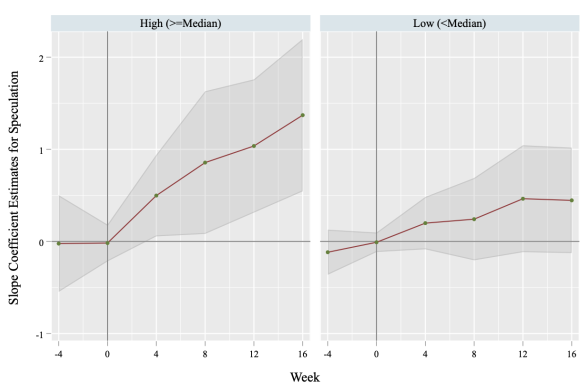
This figure presents the coefficient estimates of our speculation measure (Speculation) in the regressions of cumulative buy and hold abnormal returns (BHARs) on Speculation over various estimation windows based on four-week (roughly monthly) intervals by heterogeneous firm variables. The shaded area represents the 95% confidence interval for each coefficient estimate (circle marker). For the post-filing periods, each four-week cumulative BHAR is computed over the period from the start of the 1st week to the end of the $n$th week ($n = 4, 8, 12, 16$). For the pre-filing period, the "reverse" four-week cumulative BHAR is computed over the period from the end of the $-1$th week to the start of the $n$th week ($n = -4$). The vertical solid line is located at Week 0 that covers the four days between the 10-K filing day and three days later. We consider the following four firm-specific variables: forward-looking disclosure, R&D disclosure, pre-filing idiosyncratic return volatility, and pre-filing analyst coverage, which are constructed based on the information available as of 10-K filing dates. High and Low of each of the first three variables refer to firms with the variable above and below its median, respectively. For pre-filing analyst coverage, Absence and Presence are firms with no analyst following and those with at least one analyst following, respectively, which are determined by the availability of analysts before each 10-K filing date. The coefficient estimates and confidence intervals are presented in percentage.



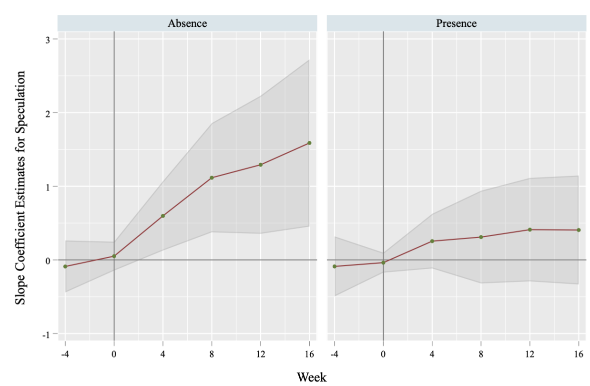

(a) Forward-looking disclosure




(b) R&D disclosure




(c) Pre-filing idiosyncratic return volatility




(d) Pre-filing analyst coverage

Figure 6: Speculation and Illiquidity

This figure presents the coefficient estimates of our speculation measure (Speculation) in the regressions of the quoted relative bid-ask spread averaged over each of four-week (roughly monthly) windows on Speculation. The shaded area represents the 95% confidence interval for each slope coefficient estimate (circle marker). For the post-filing periods, each relative bid-ask spread is computed over the four-week window from the start of the $n-3$th week to the end of the $n$th week ($n = 4, 8, 12, 16$). For the pre-filing period, the relative bid-ask spread is computed from Week -4 to Week -1. The vertical solid line is located at Week 0 that covers the four days between the 10-K filing day and three days later. The coefficient estimates and confidence intervals are presented in percentage.

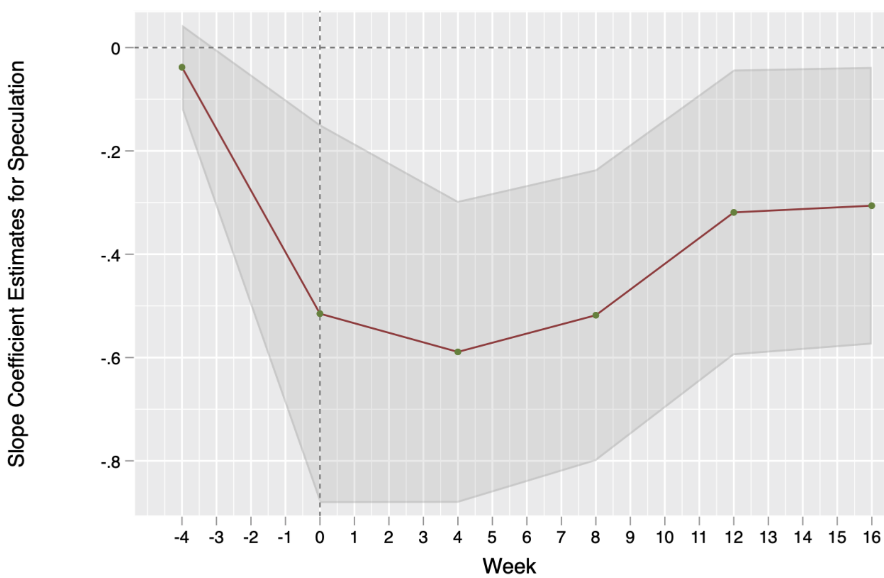
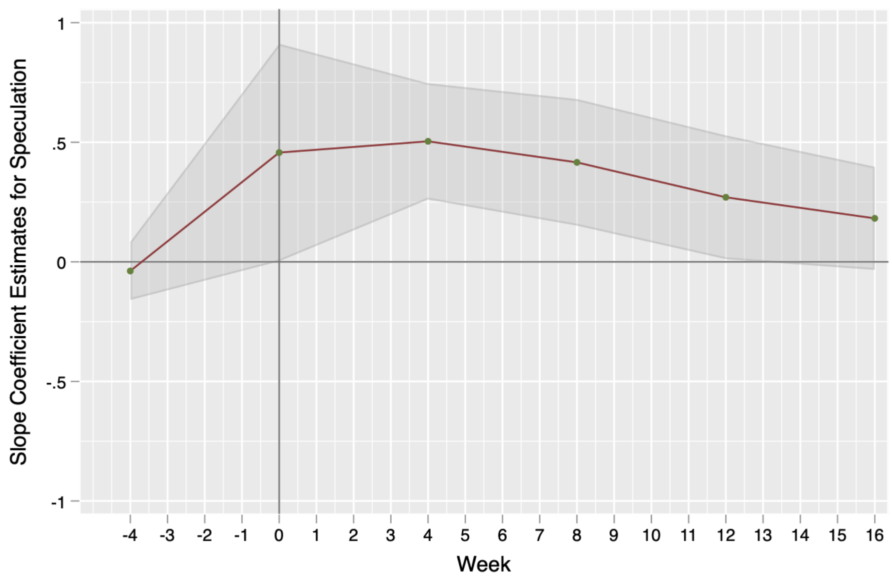
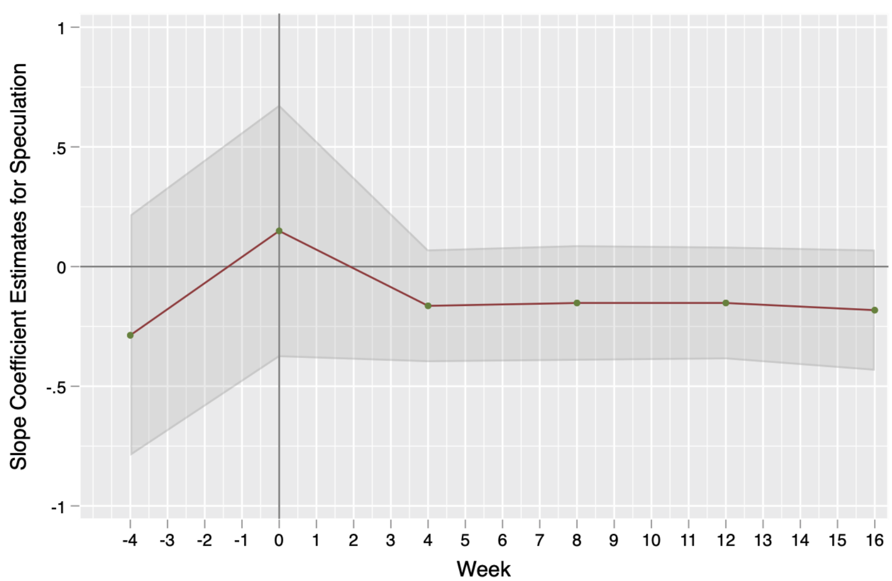


Figure 7: Speculation and Informed Trading Activities

This figure presents the coefficient estimates of our speculation measure (Speculation) in the regressions of (a) Probability of informed buying or (b) Probability of informed selling over each of four-week (roughly monthly) windows on Speculation. These posterior probabilities of informed buying and selling are based on Brennan, Huh, and Subrahmanyam (2018). The shaded area represents the 95% confidence interval for each slope coefficient estimate (circle marker). For the post-filing periods, the probability of informed buying or selling is computed over the four-week window from the start of the $n-3$th week to the end of the $n$th week ($n = 4, 8, 12, 16$). For the pre-filing period, the probability of informed buying or selling is computed from Week -4 to Week -1. The vertical solid line is located at Week 0 that covers the four days between the 10-K filing day and three days later. The coefficient estimates and confidence intervals are presented in percentage.



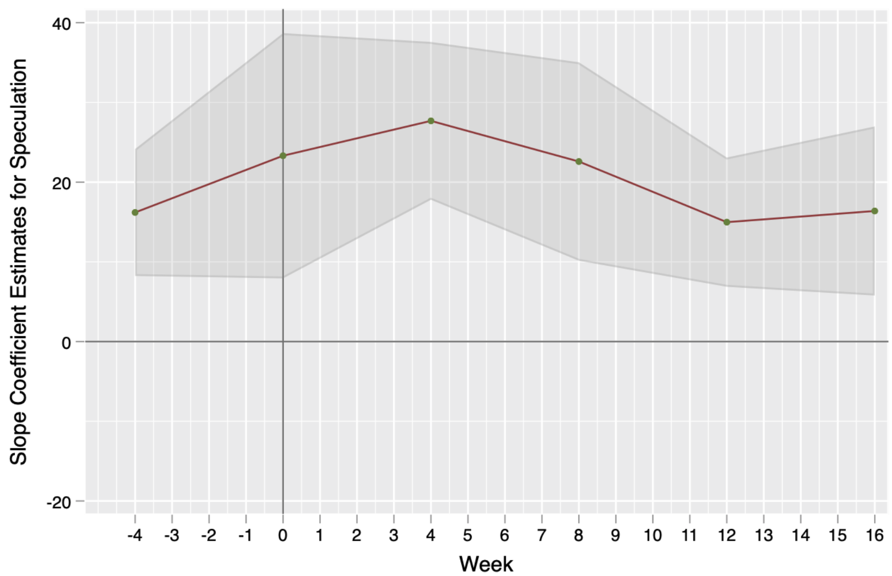

(a) Probability of informed buying



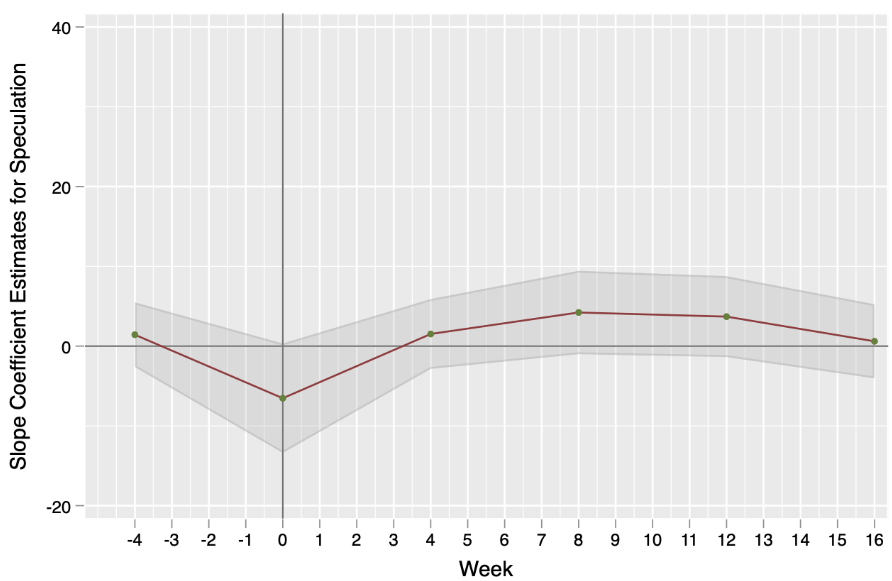

(b) Probability of informed selling

Figure 8: Speculation and Insider Trading Activities

This figure presents the coefficient estimates of our speculation measure (Speculation) in the regressions of (a) Dollar volume of insider buying or (b) Dollar volume of insider selling over each of four-week (roughly monthly) windows on Speculation. The shaded area represents the 95% confidence interval for each slope coefficient estimate (circle marker). For the post-filing periods, the dollar volume of insider buying or selling is computed over the four-week window from the start of the $n-3$th week to the end of the $n$th week ($n = 4, 8, 12, 16$). For the pre-filing period, the dollar volume of insider buying or selling is computed from Week -4 to Week -1. The vertical solid line is located at Week 0 that covers the four days between the 10-K filing day and three days later. The coefficient estimates and confidence intervals are presented in percentage.




(a) Dollar volume of insider buying




(b) Dollar volume of insider selling

Figure 9: Speculation and News Sentiment

This figure presents the coefficient estimates of our speculation measure (Speculation) in the regressions of the news sentiment score averaged over each of four-week (roughly monthly) windows on Speculation. The shaded area represents the 95% confidence interval for each slope coefficient estimate (circle marker). For the post-filing periods, the news sentiment score is computed over the four-week window from the start of the $n-3$th week to the end of the $n$th week ($n = 4, 8, 12, 16$). For the pre-filing period, the news sentiment score is computed from Week -4 to Week -1. The vertical solid line is located at Week 0 that covers the four days between the 10-K filing day and three days later. The coefficient estimates and confidence intervals are presented in percentage.

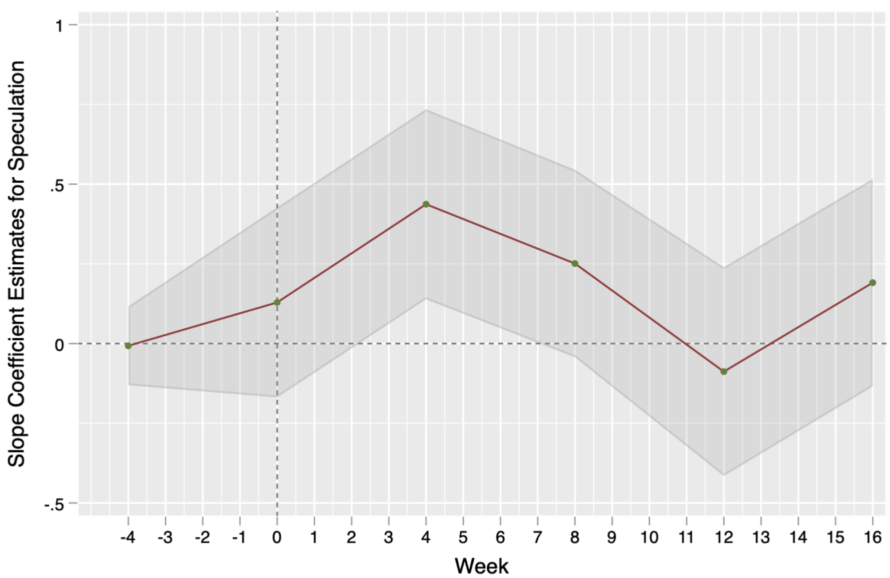


Table 1: Frequently Used Words in Wikipedia Sentences with Weasel Tags and in 10-K Paragraphs with Speculation Cues

Panel A presents frequently used words in Wikipedia Sentences with weasel tags ({{Weasel-inline—{{subst:DATE}}}}) from a Wikipedia dump completed on April 20, 2017. The Wikipedia dump contains 17,483,910 articles. Panel A(i) lists the top 10 most frequently mentioned unigrams, bigrams, and trigrams in the sentences with weasel tags. Panel A(ii) lists the top 10 unigrams and the bottom 10 unigrams sorted on the saliency score of Goldsmith-Pinkham, Hirtle, and Lucca (2016) to illustrate the effectiveness of our saliency screen. Panel A(iii) lists the top 10 most frequently mentioned speculation cues, uncertainty words, and weak and strong modal words used in 10-Ks. The keyword lists of uncertainty, weak modality, and strong modality words come from the master dictionary by Loughran and McDonald (2011). Panel B presents frequently used words in 10-K paragraphs with our speculation cues. Panel B(i) lists the top 30 most frequently mentioned nouns, verbs, and adjectives/adverbs. We only include the words in the Loughran and McDonald (2011) master dictionary that are considered to add financial information content. Panel B(ii) lists meaningful neighboring bigrams in 10-K paragraphs with the speculation cues by information contents. We analyze 5,597,740 pairs of neighboring words in all 10-K paragraphs and compute the saliency score of each bigram in paragraphs with the speculation cues relative to paragraphs without such keywords. (adverse, material, effect)†represents 13 distinct bigrams that capture the same idea. Appendix Table IA.1 shows the complete list of the top 100 neighboring bigrams that are overused relative to common language by the saliency screen.

***Panel A: Frequently Used Words in Wikipedia Sentences with Weasel Tags***

(i) Top 10 Unigrams, Bigrams, and Trigrams

| Rank | Unigrams | Bigrams | Trigrams |
|---|---|---|---|
| 1 | the | of the | one of the |
| 2 | and | in the | it has been |
| 3 | some | it is | considered by many |
| 4 | that | to be | is considered by |
| 5 | was | has been | of the most |
| 6 | many | to the | is one of |
| 7 | for | for the | it can be |
| 8 | with | one of | may have been |
| 9 | has | and the | according to some |
| 10 | have | that the | be one of |

(ii) Top and Bottom 10 Unigrams, Sorted on Saliency

| Rank | Top 10 Unigrams | Bottom 10 Unigrams |
|---|---|---|
| 1 | some | the |
| 2 | many | and |
| 3 | although | for |
| 4 | considered | was |
| 5 | may | from |
| 6 | said | their |
| 7 | have | new |
| 8 | argued | united |
| 9 | believed | also |
| 10 | often | first |

(iii) Top 10 Speculation Cues, Uncertainty, and Modal Words used in 10-Ks

| Rank | Speculation Cues | Uncertainty Words | Weak Modal Words | Strong Modal Words |
|---|---|---|---|---|
| 1 | other | may | may | will |
| 2 | may | could | could | must |
| 3 | could | approximately | possible | best |
| 4 | number of | risk | might | highest |
| 5 | would | intangible | depend | never |
| 6 | clear | believe | uncertain | lowest |
| 7 | can | assumptions | depending | always |
| 8 | well | risks | depends | clearly |
| 9 | various | believes | appears | strongly |
| 10 | however | anticipated | appearing | undisputed |

***Panel B: Frequently Used Words in 10-K Paragraphs with Speculation Cues***

(i) Top 30 Unigrams by Parts of Speech

| Verb | | Noun | | Adjective/Adverb | |
|---|---|---|---|---|---|
| will | anticipated | plan | law | approximately | unpaid |
| required | impaired | future | regulations | effective | favorable |
| expected | assumed | loss | contract | generally | statutory |
| estimated | restructuring | losses | assumptions | regulatory | difficult |
| require | intended | obligations | risks | adverse | successfully |
| restricted | discontinued | risk | default | legal | duly |
| amended | intend | benefit | decrease | adversely | critical |
| requires | restated | requirements | obligation | able | uncertain |
| permitted | anticipate | estimates | collapse | greater | strong |
| expect | prevent | impairment | court | unable | hazardous |
| comply | achieve | plans | closing | contractual | doubtful |
| terminated | increasing | contracts | intangible | beneficial | negatively |
| disclosed | projected | termination | amendment | notwithstanding | satisfactory |
| terminate | depend | laws | failure | pending | furthermore |
| differ | satisfy | claims | gains | successful | beneficially |

(ii) Meaningful Neighboring Bigrams by Information Contents

| Innovation (6) | Forward-looking (5) | Market Conditions (11) | Firm Value (19) |
|---|---|---|---|
| (intellectual, property) | (forward-looking, statement) | (market, value) | (adverse, material, effect)† |
| (clinical, trial) | (company, belief) | (market, price) | (comprehensive, income) |
| (product, candidate) | (management, belief) | (economic, condition) | (financial, condition) |
| (property, right) | (future, cash) | (market, condition) | (operating, result) |
| (new, product) | (future, period) | (stock, price) | (significant, deficiency) |
| (trade, secret) | | (equity, instrument) | (financial, result) |
| | | (public, offering) | (actual, result) |
| | | (closing, price) | |
| | | (overall, financial) | |
| | | (market, participant) | |
| | | (fair, market) | |

Table 2: Summary Statistics

This table presents the summary statistics for variables used in our empirical analyses. The sample period is from 1997 to 2021. The summary statistics are on our speculation measure (Speculation), existing textual tonal variables based on the master dictionary by Loughran and McDonald (2011), Fog words initially proposed by Robert Gunning in 1952 and used in the literature (e.g., Li (2008)), where all tonal variables are in percentage, and non-tonal firm characteristics that are used in our analyses. The detailed definitions of all variables are provided in the Appendix B. Each variable is winsorized at the top and bottom 1% of its distribution.

| | Mean | Std.Dev | Min | Median | Max | Num. of Obs. |
|---|---|---|---|---|---|---|
| ***Textual Tonal Variables*** | | | | | | |
| Speculation | 1.311 | 0.466 | 0.000 | 1.366 | 5.925 | 60982 |
| Sentiment | -0.679 | 0.429 | -4.362 | -0.624 | 1.670 | 60982 |
| Fog | 31.768 | 5.152 | 13.675 | 30.782 | 53.947 | 60982 |
| Uncertain | 0.982 | 0.368 | 0.000 | 1.009 | 3.365 | 60982 |
| Modal | 0.710 | 0.372 | 0.000 | 0.728 | 2.607 | 60982 |
| Litigious | 1.126 | 0.842 | 0.037 | 0.925 | 6.819 | 60982 |
| Constraining | 0.531 | 0.246 | 0.000 | 0.547 | 2.116 | 60982 |
| Superfluous | 0.008 | 0.011 | 0.000 | 0.005 | 0.253 | 60982 |
| Interesting | 0.111 | 0.075 | 0.000 | 0.103 | 1.666 | 60982 |
| ***Non-tonal Firm Characteristics*** | | | | | | |
| Size | 7.302 | 1.992 | 0.515 | 7.217 | 15.084 | 60982 |
| Age | 2.567 | 0.878 | 0.000 | 2.708 | 4.094 | 60982 |
| Tobin's Q | 1.993 | 1.577 | 0.708 | 1.431 | 10.774 | 60982 |
| Sales Growth | 0.095 | 0.266 | -0.814 | 0.073 | 1.226 | 60982 |
| Product market fluidity | 6.982 | 3.551 | 1.461 | 6.327 | 17.956 | 58038 |
| Financial constraints (KZ) | 0.466 | 1.135 | -5.556 | 0.429 | 3.669 | 60562 |
| Turnover | -1.816 | 1.085 | -4.791 | -1.695 | 0.535 | 60969 |
| Institutional ownership | 61.089 | 29.185 | 0.644 | 67.429 | 100.000 | 58764 |
| Fama-French alpha | 0.001 | 0.002 | -0.004 | 0.000 | 0.010 | 60982 |
| Filing-day abnormal return | 0.001 | 0.054 | -0.180 | -0.000 | 0.203 | 60982 |

Table 3: Relations between Speculation and Other Existing Variables

This table presents the estimation results from the regressions of our speculation measure (Speculation) on various textual tonal measures and non-tonal firm characteristics. In Columns (1) and (2), Fog is based on Robert Gunning in 1952, and the other tonal measures including Uncertainty, Modal, Positive and Negative (for Sentiment), Constraining, Litigious, Superfluous, and Interesting are based on the master dictionary by Loughran and McDonald (2011). Speculation and these textual tonal variables are contemporaneous in the sense that all of them are constructed based on the same 10-Ks. Columns (3) to (5) consider non-tonal firm characteristics, which are lagged by one year relative to the year of 10-K filing. The detailed definitions of all variables are provided in the Appendix B. Each variable is winsorized at the top and bottom 1% of its distribution. (Z) indicates that the variable is standardized to have the mean of zero and the standard deviation of one. Firm and year fixed effects are also included in the estimation of regressions. Standard errors reported in parentheses are clustered by firm and year. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% level, respectively.

| | Dependent variable = Speculation | | | | |
|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) |
| ***Textual Tonal Variables*** | | | | | |
| Fog (Z) | 0.063*** | 0.070*** | | | |
| | (0.008) | (0.007) | | | |
| Uncertain (Z) | 0.114*** | 0.132*** | | | |
| | (0.011) | (0.013) | | | |
| Modal (Z) | 0.323*** | 0.255*** | | | |
| | (0.022) | (0.013) | | | |
| Sentiment (Z) | | 0.005* | | | |
| | | (0.003) | | | |
| Constraining (Z) | | 0.061*** | | | |
| | | (0.005) | | | |
| Litigious (Z) | | 0.067*** | | | |
| | | (0.005) | | | |
| Superfluous (Z) | | 0.009*** | | | |
| | | (0.002) | | | |
| Interesting (Z) | | 0.017*** | | | |
| | | (0.003) | | | |
| ***Non-tonal Firm Characteristics*** | | | | | |
| Size (Z) | | | -0.004 | -0.008 | -0.002 |
| | | | (0.018) | (0.019) | (0.019) |
| Age (Z) | | | -0.075*** | -0.075*** | -0.079*** |
| | | | (0.008) | (0.009) | (0.010) |
| Tobin's Q (Z) | | | 0.006 | 0.007 | 0.005 |
| | | | (0.005) | (0.005) | (0.006) |
| Sales Growth (Z) | | | -0.001 | -0.002 | -0.001 |
| | | | (0.003) | (0.003) | (0.003) |
| Product market fluidity (Z) | | | | 0.008 | 0.008 |
| | | | | (0.005) | (0.005) |
| Financial constraints (KZ Z) | | | | 0.005 | 0.003 |
| | | | | (0.004) | (0.005) |
| Turnover (Z) | | | | | -0.006 |
| | | | | | (0.007) |
| Institutional ownership (Z) | | | | | -0.003 |
| | | | | | (0.006) |
| Fama-French alpha (Z) | | | | | -0.001 |
| | | | | | (0.002) |
| Filing-day abnormal return (Z) | | | | | -0.002 |
| | | | | | (0.002) |
| Observations | 59575 | 59575 | 59575 | 56215 | 54112 |
| Adjusted $R^2$ | 0.849 | 0.867 | 0.638 | 0.642 | 0.645 |

## Table 4: Speculative Language in Disclosure and Cumulative BHARs

This table presents the estimation results of Equation (1) as follows: For stock $i$ in year $t$,

$$BHAR_{itn} \quad = \quad \alpha_n + \beta_n Speculation_{it} + \eta_n^{'} \mathbf{X}_{it} + \gamma_n RMkt_{tn} + \epsilon_{itn}.$$

$BHAR_{itn}$ is defined as the cumulative return difference between stock $i$ and the CRSP value-weighted index relative to Week 0 up to the $n$th-week, where $n = -4, 0, 4, 8, 12, 16$ and Week 0 has the four days between the 10-K filing day and three days later. $Speculation_{it}$ is the percentage of our speculation cues (out of the total words) in the 10-K disclosure, and $\mathbf{X}_{it}$ is a column vector of our baseline control variables based on prior studies including Sentiment, Market value, Book-to-market, Turnover, Institutional ownership, Fama-French alpha, and Filing-day abnormal return. $RMkt_{tn}$ is the cumulative market return to capture common economic shocks that affect all individual stocks over the same $n$-week period as $BHAR_{itn}$. The detailed definitions of these explanatory variables are given in the Appendix B. We estimate Equation (1) with cumulative BHARs over various estimation windows based on four-week (roughly monthly) intervals as the dependent variables. For the post-filing period, each cumulative BHAR is computed over the period from the start of the 1st week (i.e., the fourth day after 10-K filing) to the end of the $n$th week, where $n = 4, 8, 12, 16$. For the pre-filing period, the "reverse" cumulative BHAR is computed over the period from the end of Week -1 to the start of the $n$th week, where $n = -4$. That is, the reverse cumulative BHAR for Week[-4,-1] is obtained by solving $(1 + BHAR_{[-4,-1]}) = \frac{1}{(1+BHAR_{-4})*(1+BHAR_{-3})*(1+BHAR_{-2})*(1+BHAR_{-1})}$ for $BHAR_{[-4,-1]}$, where $BHAR_{-k}$ is the weekly BHAR for Week $-k$. (Z) indicates that the variable is standardized to have the mean of zero and the standard deviation of one. Standard errors (SEs) reported in parentheses are clustered by firm and filing year-month. Estimated coefficients and SEs are reported in percentage. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% level, respectively.

| | Dependent variable = Cumulative BHARs | | | | | |
|---|---|---|---|---|---|---|
| | (1) Week[-4,-1] | (2) Week 0 | (3) Week[1,4] | (4) Week[1,8] | (5) Week[1,12] | (6) Week[1,16] |
| Speculation (Z) | 0.108 | -0.071 | 0.479*** | 0.927*** | 0.919*** | 1.113*** |
| | (0.162) | (0.047) | (0.129) | (0.226) | (0.263) | (0.340) |
| Sentiment (Z) | 0.127 | 0.021 | 0.127 | 0.185 | -0.089 | 0.010 |
| | (0.128) | (0.031) | (0.115) | (0.193) | (0.207) | (0.275) |
| Market value (Z) | -1.796*** | -0.098** | -0.541*** | -1.183*** | -1.511*** | -2.060*** |
| | (0.186) | (0.050) | (0.147) | (0.301) | (0.391) | (0.447) |
| Book-to-market (Z) | -0.453*** | -0.028 | 0.139 | 0.163 | -0.072 | -0.220 |
| | (0.170) | (0.037) | (0.166) | (0.254) | (0.305) | (0.309) |
| Turnover (Z) | 1.629*** | -0.144** | -0.050 | 0.198 | 0.144 | 0.055 |
| | (0.310) | (0.065) | (0.345) | (0.520) | (0.571) | (0.625) |
| Institutional ownership (Z) | -0.642*** | 0.203*** | 0.082 | 0.049 | 0.125 | 0.195 |
| | (0.245) | (0.054) | (0.194) | (0.283) | (0.354) | (0.443) |
| Fama-French alpha (Z) | -3.300*** | -0.116** | -0.068 | -0.258 | 0.264 | 0.523 |
| | (0.312) | (0.051) | (0.250) | (0.394) | (0.385) | (0.432) |
| Contemporaneous market return (Z) | 0.890** | 0.065 | 0.701** | 1.207** | 1.700*** | 1.654*** |
| | (0.409) | (0.081) | (0.315) | (0.488) | (0.462) | (0.603) |
| Filing-day abnormal return (Z) | 0.324*** | - | -0.224** | -0.265 | -0.375 | -0.397* |
| | (0.117) | - | (0.111) | (0.177) | (0.232) | (0.234) |
| Observations | 63195 | 63199 | 61621 | 60842 | 60061 | 59297 |
| Adjusted R2 | 0.065 | 0.002 | 0.007 | 0.011 | 0.012 | 0.012 |

Table 5: Speculative Language in Disclosure and Cumulative BHARs: Decomposition

This table presents the estimation results of the following extended versions of Equation (1). In Panel A, we use a more refined version of our speculation measure, where we purge all words in the dictionaries for uncertainty and modality and all words identified as complex according to the Fog index. After taking out all words that overlap with the dictionaries for uncertainty, modality, and complex (Fog) words from our list of speculation cues, we recount how many times those purged speculation cues are mentioned in a given firm's 10-K filing in a given year scaled by the total word count in the filing (in percentage). In Panel B, we use the original speculation measure but additionally control for the measures of uncertainty, modality, 10-K readability (Loughran and McDonald (2014)), and complexity. In Panel C, we decompose our dictionary of speculation cues into unigrams and non-unigrams (bigrams or trigrams) and reconstruct our speculation measure only based on bigrams and trigrams. In Panel D, we additionally control for the forward-looking measure based on Muslu, Radhakrishnan, Subramanyam, and Lim (2015). In Panel E, we additionally control for the sentiment in forward-looking statements. In Panel F, we decompose the sentiment in forward-looking statements into its positive and negative components, and then include them as additional controls. All other regression specifications are the same as those in Table 4. (Z) indicates that the variable is standardized to have the mean of zero and the standard deviation of one. Standard errors (SEs) reported in parentheses are clustered by firm and filing year-month. Estimated coefficients and SEs are reported in percentage. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% level, respectively.

| | Dependent variable = Cumulative BHARs | | | | | |
|---|---|---|---|---|---|---|
| | (1) Week[-4,-1] | (2) Week 0 | (3) Week[1,4] | (4) Week[1,8] | (5) Week[1,12] | (6) Week[1,16] |
| ***Panel A: Purging Uncertainty, Modality, and Fog measures from Speculation Cues*** | | | | | | |
| Purged speculation (Z) | 0.134 | -0.013 | 0.268*** | 0.501*** | 0.495*** | 0.552*** |
| | (0.094) | (0.031) | (0.079) | (0.136) | (0.148) | (0.185) |
| ***Panel B: Controlling for Uncertainty, Modality, LM-readability, and Fog measures*** | | | | | | |
| Speculation (Z) | -0.089 | -0.013 | 0.364** | 0.573** | 0.723*** | 0.833*** |
| | (0.163) | (0.059) | (0.162) | (0.255) | (0.267) | (0.283) |
| Uncertain (Z) | -0.057 | 0.010 | 0.226 | 0.279 | 0.235 | 0.416 |
| | (0.147) | (0.042) | (0.144) | (0.205) | (0.235) | (0.266) |
| Modal (Z) | 0.088 | -0.095 | -0.608** | -0.880** | -0.661 | -0.587 |
| | (0.229) | (0.063) | (0.247) | (0.396) | (0.434) | (0.515) |
| LM-Readability (Z) | -0.323 | -0.000 | -0.821*** | -1.597*** | -1.052* | -0.898 |
| | (0.307) | (0.068) | (0.283) | (0.493) | (0.590) | (0.744) |
| Fog (Z) | 0.184 | 0.017 | -0.291* | -0.274 | -0.346 | -0.548* |
| | (0.151) | (0.052) | (0.149) | (0.208) | (0.241) | (0.301) |
| ***Panel C: Using only Bigram and Trigram Speculation Cues*** | | | | | | |
| Speculation (Z) | 0.076 | -0.034 | 0.396*** | 0.717*** | 0.609** | 0.542* |
| | (0.157) | (0.044) | (0.129) | (0.240) | (0.282) | (0.309) |
| ***Panel D: Controlling for Forward-looking Disclosure*** | | | | | | |
| Speculation (Z) | 0.021 | -0.043 | 0.239** | 0.519*** | 0.527** | 0.745*** |
| | (0.118) | (0.044) | (0.102) | (0.187) | (0.224) | (0.280) |
| Forward-looking (Z) | 0.140 | -0.045 | 0.384** | 0.653** | 0.625** | 0.588* |
| | (0.165) | (0.044) | (0.156) | (0.257) | (0.293) | (0.349) |

| | Dependent variable = Cumulative BHARs | | | | | |
|---|---|---|---|---|---|---|
| | (1)<br>Week[-4,-1] | (2)<br>Week 0 | (3)<br>Week[1,4] | (4)<br>Week[1,8] | (5)<br>Week[1,12] | (6)<br>Week[1,16] |
| ***Panel E: Controlling for Sentiment in Forward-looking Disclosure*** | | | | | | |
| Speculation (Z) | 0.134 | -0.081 | 0.503*** | 0.970*** | 0.986*** | 1.173*** |
| | (0.173) | (0.049) | (0.142) | (0.252) | (0.284) | (0.359) |
| FW-Sentiment (Z) | 0.211 | 0.003 | 0.205 | 0.317 | 0.048 | 0.150 |
| | (0.146) | (0.030) | (0.136) | (0.205) | (0.214) | (0.293) |
| ***Panel F: Controlling for Positive and Negative Words in Forward-looking Disclosure*** | | | | | | |
| Speculation (Z) | 0.149 | -0.039 | 0.487*** | 0.970*** | 0.949*** | 1.131*** |
| | (0.168) | (0.050) | (0.143) | (0.239) | (0.281) | (0.336) |
| FW-Positive (Z) | 0.123 | -0.090 | 0.178 | 0.236 | 0.169 | 0.235 |
| | (0.190) | (0.065) | (0.178) | (0.289) | (0.379) | (0.430) |
| FW-Negative (Z) | -0.461 | 0.095 | -0.461 | -0.735* | -0.525 | -0.640 |
| | (0.310) | (0.088) | (0.284) | (0.441) | (0.503) | (0.588) |

Table 6: Speculative Language in Disclosure and Cumulative BHARs by Heterogeneous Variables

We consider the following four firm-specific variables for heterogeneity tests: forward-looking disclosure, R&D disclosure, pre-filing idiosyncratic return volatility, and pre-filing analyst coverage, which are constructed based on the information available as of 10-K filing dates. High and Low of each of the first three variables refer to firms with the variable above versus below its median, respectively. For pre-filing analyst coverage, Absence and Presence are firms with no analyst following and those with at least one analyst following, respectively, which are determined by the availability of analysts before each 10-K filing. For each subsample based on these variables, this table presents the estimation results of Equation (1) with additionally controlling for the measures of uncertainty, modality, 10-K readability (Loughran and McDonald (2014)), and complexity as in Panel B of Table 5. All other regression specifications are the same as those in Table 4. (Z) indicates that the variable is standardized to have the mean of zero and the standard deviation of one. Standard errors (SEs) reported in parentheses are clustered by firm and filing year-month. Estimated coefficients and SEs are reported in percentage. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% level, respectively.

| | | Dependent variable = Cumulative BHARs | | | | | |
|---|---|---|---|---|---|---|---|
| | | (1) | (2) | (3) | (4) | (5) | (6) |
| Subsamples by | | Week[-4,-1] | Week 0 | Week[1,4] | Week[1,8] | Week[1,12] | Week[1,16] |
| ***Panel A: Forward-looking disclosure*** | | | | | | | |
| High | Speculation (Z) | 0.153 | -0.024 | 0.680** | 0.974*** | 1.569*** | 1.715*** |
| | | (0.278) | (0.089) | (0.268) | (0.359) | (0.418) | (0.475) |
| Low | Speculation (Z) | -0.069 | -0.018 | 0.248 | 0.437 | 0.326 | 0.392 |
| | | (0.209) | (0.061) | (0.203) | (0.343) | (0.349) | (0.352) |
| ***Panel B: R&D disclosure*** | | | | | | | |
| High | Speculation (Z) | 0.178 | -0.096 | 0.574** | 1.136*** | 1.310*** | 1.490*** |
| | | (0.261) | (0.083) | (0.246) | (0.347) | (0.442) | (0.506) |
| Low | Speculation (Z) | -0.095 | 0.043 | 0.303 | 0.272 | 0.473 | 0.585* |
| | | (0.201) | (0.070) | (0.204) | (0.336) | (0.325) | (0.344) |
| ***Panel C: Pre-filing idiosyncratic return volatility*** | | | | | | | |
| High | Speculation (Z) | -0.023 | -0.017 | 0.498** | 0.856** | 1.036*** | 1.371*** |
| | | (0.268) | (0.102) | (0.226) | (0.395) | (0.369) | (0.422) |
| Low | Speculation (Z) | -0.117 | -0.009 | 0.199 | 0.242 | 0.464 | 0.446 |
| | | (0.125) | (0.054) | (0.145) | (0.228) | (0.296) | (0.292) |
| ***Panel D: Pre-filing analyst coverage*** | | | | | | | |
| Absence | Speculation (Z) | -0.089 | 0.051 | 0.598** | 1.116*** | 1.292*** | 1.588*** |
| | | (0.180) | (0.100) | (0.240) | (0.378) | (0.478) | (0.579) |
| Presence | Speculation (Z) | -0.088 | -0.037 | 0.255 | 0.310 | 0.411 | 0.406 |
| | | (0.208) | (0.069) | (0.189) | (0.321) | (0.358) | (0.377) |

Table 7: Speculative Language in Disclosure and Illiquidity

This table presents the estimation results of Equation (2) as follows: For stock $i$ in year $t$,

$$Spread_{itn} \quad = \quad \alpha_n + \beta_n Speculation_{it} + \eta_n^{'} \mathbf{X}_{it} + \epsilon_{itn}.$$

$Spread_{itn}$ is the average quoted relative bid-ask spread (in logarithm) over the four-week window ending in the $n$th week where $n = 4, 8, 12, 16$ for the post-filing period and over the four-week window from Week -4 to Week -1, i.e., Week[-4,-1], for the pre-filing period. Week 0 has the four days between the 10-K filing day and three days later. $Speculation_{it}$ is the percentage of our speculation cues (out of the total words) in the 10-K disclosure, and $\mathbf{X}_{it}$ is a column vector of our baseline control variables: Sentiment, Market value, Book-to-market, Turnover, Institutional ownership, Fama-French alpha, and Filing-day abnormal return as in Table 4, and two additional variables: Nasdaq dummy and Pre-filing spread. The detailed definitions of these explanatory variables are given in the Appendix B. (Z) indicates that the variable is standardized to have the mean of zero and the standard deviation of one. When estimating Equation (2), we include firm and filing year-month fixed effects. Standard errors (SEs) reported in parentheses are clustered by firm and filing year-month. Estimated coefficients and SEs are reported in percentage. $^{***}$, $^{**}$, and $^{*}$ indicate statistical significance at the 1%, 5%, and 10% level, respectively.

| | Dependent variable = Quoted relative bid-ask spread | | | | | |
|---|---|---|---|---|---|---|
| | (1) Week[-4,-1] | (2) Week 0 | (3) Week[1,4] | (4) Week[5,8] | (5) Week[9,12] | (6) Week[13,16] |
| Speculation (Z) | -0.038 | -0.515*** | -0.589*** | -0.518*** | -0.319** | -0.306** |
| | (0.042) | (0.187) | (0.149) | (0.144) | (0.141) | (0.137) |
| Sentiment (Z) | -0.004 | 0.045 | 0.101 | -0.084 | 0.089 | -0.001 |
| | (0.044) | (0.158) | (0.129) | (0.126) | (0.136) | (0.140) |
| Market value (Z) | -0.346*** | -3.689*** | -4.695*** | -5.842*** | -6.742*** | -7.181*** |
| | (0.110) | (0.436) | (0.497) | (0.675) | (0.669) | (0.777) |
| Book-to-market (Z) | -0.115** | 0.032 | 0.231* | 0.389** | 0.467** | 0.658*** |
| | (0.051) | (0.160) | (0.121) | (0.156) | (0.184) | (0.163) |
| Turnover (Z) | -0.384*** | -1.584*** | -1.581*** | -1.923*** | -1.902*** | -1.988*** |
| | (0.075) | (0.256) | (0.279) | (0.313) | (0.337) | (0.325) |
| Institutional ownership (Z) | 0.126** | -0.134 | -0.507*** | -0.663*** | -0.875*** | -0.902*** |
| | (0.049) | (0.184) | (0.171) | (0.209) | (0.232) | (0.267) |
| Fama-French alpha (Z) | -0.244*** | -0.711*** | -0.811*** | -0.760*** | -0.795*** | -1.106*** |
| | (0.050) | (0.139) | (0.161) | (0.170) | (0.194) | (0.194) |
| Pre-filing spread (Z) | 47.762*** | 40.085*** | 37.678*** | 34.203*** | 31.304*** | 29.802*** |
| | (0.139) | (0.617) | (0.683) | (1.064) | (1.180) | (1.059) |
| Filing-day abnormal return (Z) | 0.014 | -0.437*** | -1.030*** | -0.959*** | -1.050*** | -1.004*** |
| | (0.032) | (0.130) | (0.112) | (0.093) | (0.100) | (0.118) |
| Nasdaq dummy | 0.091 | -1.731* | -0.983* | -1.485* | -1.134 | -1.652* |
| | (0.181) | (0.915) | (0.585) | (0.769) | (0.958) | (0.969) |
| Observations | 61527 | 61531 | 60747 | 59379 | 58814 | 58303 |
| Adjusted $R^2$ | 0.988 | 0.873 | 0.932 | 0.913 | 0.897 | 0.883 |

Table 8: Speculative Language in Disclosure and Informed Buying Activity

This table presents the estimation results of Equation (3) as follows: For stock $i$ in year $t$,

$$Probability\ of\ informed\ buying_{itn} = \alpha_n + \beta_n Speculation_{it} + \eta_n^{'} \mathbf{X}_{it} + \epsilon_{itn}.$$

*Probability of informed buying*$_{itn}$ proxies the trading activity of informed buyers based on Brennan, Huh, and Subrahmanyam (2018) over the four-week window ending in the $n$th week where $n = 4, 8, 12, 16$ for the post-filing period and over the four-week window from Week -4 to Week -1, i.e., Week[-4,-1], for the pre-filing period. Week 0 has the four days between the 10-K filing day and three days later. *Speculation*$_{it}$ is the percentage of our speculation cues (out of the total words) in the 10-K disclosure, and $\mathbf{X}_{it}$ is a column vector of our baseline control variables: Sentiment, Market value, Book-to-market, Turnover, Institutional ownership, Fama-French alpha, and Filing-day abnormal return as in Table 4, and two additional variables: Nasdaq dummy and Pre-filing informed buying. The detailed definitions of these explanatory variables are given in the Appendix B. (Z) indicates that the variable is standardized to have the mean of zero and the standard deviation of one. When estimating Equation (3), we also include filing year-month fixed effects. Standard errors (SEs) reported in parentheses are clustered by firm and filing year-month. Estimated coefficients and SEs are reported in percentage. $^{***}$, $^{**}$, and $^{*}$ indicate statistical significance at the 1%, 5%, and 10% level, respectively.

| | Dependent variable = Probability of informed buying | | | | | |
|---|---|---|---|---|---|---|
| | (1) Week[-4,-1] | (2) Week 0 | (3) Week[1,4] | (4) Week[5,8] | (5) Week[9,12] | (6) Week[13,16] |
| Speculation (Z) | -0.038 | 0.457** | 0.504*** | 0.416*** | 0.270** | 0.182* |
| | (0.062) | (0.232) | (0.124) | (0.135) | (0.132) | (0.110) |
| Sentiment (Z) | -0.049 | -0.349** | -0.137 | -0.039 | 0.008 | 0.025 |
| | (0.045) | (0.172) | (0.096) | (0.110) | (0.092) | (0.102) |
| Market value (Z) | -0.041 | -0.430 | 0.481* | 0.391 | 0.691* | 0.928** |
| | (0.104) | (0.439) | (0.260) | (0.318) | (0.360) | (0.369) |
| Book-to-market (Z) | -0.133** | -0.278 | -0.302** | -0.278* | -0.458*** | -0.176 |
| | (0.056) | (0.201) | (0.127) | (0.153) | (0.138) | (0.120) |
| Turnover (Z) | 0.070 | -0.048 | -1.204*** | -1.446*** | -1.101*** | -1.111*** |
| | (0.085) | (0.260) | (0.198) | (0.193) | (0.173) | (0.216) |
| Institutional ownership (Z) | 0.121* | 0.780*** | 1.250*** | 1.614*** | 1.616*** | 0.861*** |
| | (0.066) | (0.271) | (0.193) | (0.211) | (0.171) | (0.188) |
| Fama-French alpha (Z) | 0.204*** | 0.219 | 0.510*** | 0.573*** | 0.616*** | 0.388*** |
| | (0.056) | (0.175) | (0.174) | (0.154) | (0.131) | (0.139) |
| Pre-filing informed buying (Z) | 16.204*** | 8.693*** | 4.815*** | 1.547*** | 0.318** | 0.634*** |
| | (0.074) | (0.252) | (0.199) | (0.144) | (0.127) | (0.120) |
| Filing-day abnormal return (Z) | 0.036 | 5.202*** | 1.025*** | 0.622*** | 0.501*** | 0.069 |
| | (0.050) | (0.385) | (0.126) | (0.089) | (0.089) | (0.103) |
| Nasdaq dummy | 0.001 | 0.457 | 1.421** | 1.412** | 1.690*** | 3.188*** |
| | (0.145) | (0.608) | (0.585) | (0.554) | (0.572) | (0.602) |
| Observations | 45349 | 42871 | 43822 | 42986 | 42476 | 42133 |
| Adjusted $R^2$ | 0.782 | 0.125 | 0.104 | 0.077 | 0.065 | 0.050 |

## Table 9: Speculative Language in Disclosure and Insider Buying Activity

This table presents the estimation results of Equation (3) using *Dollar volume of insider buying* as the dependent variable in place of *Probability of informed buying* as follows: For stock $i$ in year $t$,

$$\textit{Dollar volume of insider buying}_{itn} = \alpha_n + \beta_n \textit{Speculation}_{it} + \eta_n^{'} \mathbf{X}_{it} + \epsilon_{itn}.$$

$\textit{Dollar volume of insider buying}_{itn}$ captures the insiders' buying activity in dollar volume (in logarithm) over the four-week window ending in the $n$th week where $n = 4, 8, 12, 16$ for the post-filing period and over the four-week window from Week -4 to Week -1, i.e., Week[-4,-1], for the pre-filing period. Week 0 has the four days between the 10-K filing day and three days later. $\textit{Speculation}_{it}$ is the percentage of our speculation cues (out of the total words) in the 10-K disclosure, and $\mathbf{X}_{it}$ is a column vector of our baseline control variables: Sentiment, Market value, Book-to-market, Turnover, Institutional ownership, Fama-French alpha, and Filing-day abnormal return as in Table 4, and Nasdaq dummy. The detailed definitions of these explanatory variables are given in the Appendix B. (Z) indicates that the variable is standardized to have the mean of zero and the standard deviation of one. When estimating the regression model, we also include filing year-month fixed effects. Standard errors (SEs) reported in parentheses are clustered by firm and filing year-month. Estimated coefficients and SEs are reported in percentage. $^{***}$, $^{**}$, and $^{*}$ indicate statistical significance at the 1%, 5%, and 10% level, respectively.

| | Dependent variable = Dollar volume of insider buying | | | | | |
|---|---|---|---|---|---|---|
| | (1) Week[-4,-1] | (2) Week 0 | (3) Week[1,4] | (4) Week[5,8] | (5) Week[9,12] | (6) Week[13,16] |
| Speculation (Z) | 16.193*** | 23.310*** | 27.687*** | 22.585*** | 14.980*** | 16.381*** |
| | (4.062) | (7.848) | (5.048) | (6.350) | (4.135) | (5.410) |
| Sentiment (Z) | -8.778** | -2.234 | -3.455 | -4.419 | -4.078 | -10.840** |
| | (3.667) | (7.714) | (3.835) | (3.781) | (3.584) | (4.821) |
| Market value (Z) | 21.004*** | 33.608*** | 28.593*** | 26.966*** | 32.504*** | 12.535** |
| | (4.484) | (7.746) | (5.475) | (5.278) | (4.182) | (6.145) |
| Book-to-market (Z) | -20.171*** | -9.861* | -10.507*** | -4.231 | -11.499*** | -17.802*** |
| | (3.802) | (5.559) | (3.862) | (4.900) | (3.215) | (4.854) |
| Turnover (Z) | 47.660*** | 42.245*** | 37.974*** | 36.944*** | 30.063*** | 33.073*** |
| | (3.869) | (6.092) | (4.215) | (4.089) | (3.072) | (4.703) |
| Institutional ownership (Z) | 4.169 | 2.636 | 2.922 | 6.933 | 9.306** | 13.487** |
| | (4.689) | (8.166) | (5.708) | (5.234) | (4.292) | (5.649) |
| Fama-French alpha (Z) | 1.809 | -2.680 | -2.642 | 2.199 | 1.517 | 3.505 |
| | (3.739) | (4.810) | (4.338) | (3.815) | (3.131) | (4.101) |
| Filing-day abnormal return (Z) | 8.369*** | 0.713 | -4.453 | 1.898 | 2.174 | -0.454 |
| | (2.310) | (3.993) | (2.995) | (3.291) | (2.487) | (3.018) |
| Nasdaq dummy | -29.571*** | -17.290 | -8.884 | -14.686 | -18.918*** | -5.246 |
| | (6.870) | (12.255) | (7.756) | (9.013) | (7.036) | (10.124) |
| Observations | 7169 | 2161 | 6409 | 6266 | 8050 | 5236 |
| Adjusted $R^2$ | 0.151 | 0.155 | 0.144 | 0.100 | 0.129 | 0.092 |

Table 10: Speculative Language in Disclosure and News Sentiment

This table presents the estimation results of Equation (4) as follows: For stock $i$ in year $t$,

$$News\ sentiment_{itn} = \alpha_n + \beta_n Speculation_{it} + \eta_n^{'} \mathbf{X}_{it} + \gamma_n NSMkt_{tn} + \epsilon_{itn}.$$

$News\ sentiment_{itn}$ is the average news sentiment score over the four-week window ending in the $n$th week where $n = 4, 8, 12, 16$ for the post-filing period and over the four-week window from Week -4 to Week -1, i.e., Week[-4,-1], for the pre-filing period. Week 0 has the four days between the 10-K filing day and three days later. $Speculation_{it}$ is the percentage of our speculation cues (out of the total words) in the 10-K disclosure, and $\mathbf{X}_{it}$ is a column vector of our baseline control variables: Sentiment, Market value, Book-to-market, Turnover, Institutional ownership, Fama-French alpha, and Filing-day abnormal return as in Table 4, and Pre-filing news sentiment. $NSMkt_{tn}$ is the four-week market-wide news sentiment to proxy for the common economic shocks that potentially affect all individual firms' news sentiment scores over the estimation window. The detailed definitions of these explanatory variables are given in the Appendix B. The sample period for this analysis starts from January, 2000 due to the data availability in the RavenPack News Analytics database. (Z) indicates that the variable is standardized to have the mean of zero and the standard deviation of one. Standard errors (SEs) reported in parentheses are clustered by firm and filing year-month. Estimated coefficients and SEs are reported in percentage. $^{***}$, $^{**}$, and $^{*}$ indicate statistical significance at the 1%, 5%, and 10% level, respectively.

| | Dependent variable = Average news sentiment score | | | | | |
|---|---|---|---|---|---|---|
| | (1)<br>Week[-4,-1] | (2)<br>Week 0 | (3)<br>Week[1,4] | (4)<br>Week[5,8] | (5)<br>Week[9,12] | (6)<br>Week[13,16] |
| Speculation (Z) | -0.007 | 0.129 | 0.437*** | 0.251* | -0.088 | 0.191 |
| | (0.063) | (0.152) | (0.152) | (0.150) | (0.167) | (0.166) |
| Sentiment (Z) | -0.003 | 0.266 | -0.114 | 0.031 | 0.028 | -0.033 |
| | (0.056) | (0.220) | (0.138) | (0.168) | (0.153) | (0.172) |
| Market value (Z) | 0.060 | 1.087*** | 0.253 | 0.810*** | 0.399** | -0.143 |
| | (0.062) | (0.256) | (0.157) | (0.219) | (0.168) | (0.208) |
| Book-to-market (Z) | -0.026 | -0.215 | -0.337*** | -0.115 | -0.208* | -0.278** |
| | (0.054) | (0.155) | (0.094) | (0.112) | (0.106) | (0.123) |
| Turnover (Z) | -0.063 | -1.289*** | 0.051 | -0.468*** | 0.182 | 0.207 |
| | (0.054) | (0.256) | (0.148) | (0.142) | (0.218) | (0.194) |
| Institutional ownership (Z) | -0.031 | -0.384 | -1.486*** | -1.005*** | -1.148*** | -1.131*** |
| | (0.060) | (0.266) | (0.265) | (0.168) | (0.208) | (0.212) |
| Fama-French alpha (Z) | -0.157** | 0.465* | 0.221 | 0.705*** | 0.358** | -0.146 |
| | (0.070) | (0.253) | (0.204) | (0.196) | (0.156) | (0.190) |
| Filing-day abnormal return (Z) | 0.016 | 2.967*** | 0.310** | 0.249** | -0.038 | 0.150 |
| | (0.043) | (0.276) | (0.142) | (0.107) | (0.099) | (0.109) |
| Pre-filing news sentiment (Z) | 13.302*** | 2.626*** | 2.231*** | 1.872*** | 2.041*** | 1.602*** |
| | (0.058) | (0.250) | (0.167) | (0.197) | (0.116) | (0.152) |
| Market-wide news sentiment (Z) | -0.032 | 0.450 | -0.306* | 0.315 | 0.353 | -0.425* |
| | (0.057) | (0.289) | (0.177) | (0.299) | (0.269) | (0.228) |
| Observations | 39840 | 26139 | 37298 | 39738 | 37334 | 35751 |
| Adjusted $R^2$ | 0.692 | 0.031 | 0.029 | 0.023 | 0.023 | 0.015 |

Appendix Tables to:

# Speculative and Informative: Lessons from Market Reactions to Speculative Disclosure

Table A.1: Speculative Language in Disclosure and Four-week BHARs

This table presents the estimation results of a modified Equation (1) as follows: For stock $i$ in year $t$,

$$BHAR_{itn} = \alpha_n + \beta_n Speculation_{it} + \eta_n^{'} \mathbf{X}_{it} + \gamma_n RMkt_{tn} + \epsilon_{itn}.$$

$BHAR_{itn}$ is the cumulative return difference between stock $i$ and the CRSP value-weighted index over the four-week window ending in the $n$th week where $n = 4, 8, 12, 16$ for the post-filing period and over the four-week window from Week -4 to Week -1, i.e., Week[-4,-1], for the pre-filing period. Week 0 has the four days between the 10-K filing day and three days later. $Speculation_{it}$ is the percentage of our speculation cues (out of the total words) in the 10-K disclosure, and $\mathbf{X}_{it}$ is a column vector of our baseline control variables based on prior studies including Sentiment, Market value, Book-to-market, Turnover, Institutional ownership, Fama-French alpha, and Filing-day abnormal return. $RMkt_{tn}$ is the market return to capture common economic shocks that affect all individual stocks over the same estimation window as $BHAR_{itn}$. The detailed definitions of these explanatory variables are given in the Appendix B. (Z) indicates that the variable is standardized to have the mean of zero and the standard deviation of one. Standard errors (SEs) reported in parentheses are clustered by firm and filing year-month to account for the serial and cross-sectional correlations of BHARs, respectively. Estimated coefficients and SEs are reported in percentage. $^{***}$, $^{**}$, and $^{*}$ indicate statistical significance at the 1%, 5%, and 10% level, respectively.

| | Dependent variable = Four-week BHAR | | | | | |
|---|---|---|---|---|---|---|
| | (1) Week[-4,-1] | (2) Week 0 | (3) Week[1,4] | (4) Week[5,8] | (5) Week[9,12] | (6) Week[13,16] |
| Speculation (Z) | 0.108 | -0.071 | 0.479*** | 0.380** | -0.033 | 0.149 |
| | (0.162) | (0.047) | (0.129) | (0.157) | (0.115) | (0.132) |
| Sentiment (Z) | 0.127 | 0.021 | 0.127 | -0.010 | -0.241*** | 0.122 |
| | (0.128) | (0.031) | (0.115) | (0.119) | (0.088) | (0.111) |
| Market value (Z) | -1.796*** | -0.098** | -0.541*** | -0.549*** | -0.318** | -0.462*** |
| | (0.186) | (0.050) | (0.147) | (0.173) | (0.153) | (0.119) |
| Book-to-market (Z) | -0.453*** | -0.028 | 0.139 | 0.041 | -0.250* | -0.075 |
| | (0.170) | (0.037) | (0.166) | (0.147) | (0.140) | (0.130) |
| Turnover (Z) | 1.629*** | -0.144** | -0.050 | 0.121 | 0.002 | -0.093 |
| | (0.310) | (0.065) | (0.345) | (0.241) | (0.243) | (0.180) |
| Institutional ownership (Z) | -0.642*** | 0.203*** | 0.082 | 0.072 | 0.080 | 0.107 |
| | (0.245) | (0.054) | (0.194) | (0.148) | (0.163) | (0.147) |
| Fama-French alpha (Z) | -3.300*** | -0.116** | -0.068 | -0.156 | 0.644** | 0.224 |
| | (0.312) | (0.051) | (0.250) | (0.176) | (0.262) | (0.151) |
| Contemporaneous market return (Z) | 0.890** | 0.065 | 0.701** | 0.251 | 0.250* | 0.282 |
| | (0.409) | (0.081) | (0.315) | (0.225) | (0.142) | (0.222) |
| Filing-day abnormal return (Z) | 0.324*** | - | -0.224** | -0.011 | -0.083 | -0.037 |
| | (0.117) | - | (0.111) | (0.107) | (0.130) | (0.084) |
| Observations | 63195 | 63199 | 61621 | 60987 | 60406 | 59832 |
| Adjusted R2 | 0.065 | 0.002 | 0.007 | 0.003 | 0.004 | 0.002 |

Table A.2: Speculative Language in Disclosure and Cumulative BHARs: Robustness Tests

This table presents the estimation results of the following extended versions of Equation (1) as robustness checks. In Panel A, we use a refined subsample after excluding all weekly BHAR observations having earnings announcements in the same week. In Panel B, we use Speculation adjusted for safe-harbor boilerplate paragraphs. Specifically, we subtract the number of speculation cues found in safe-harbor boilerplate paragraphs of each 10-K filing from the original total number of speculation cues in the 10-K filing. In Panel C, we additionally control for the exposures to the Fama-French three factors. In Panel D, we employ cumulative abnormal returns (CARs) as the dependent variable instead of BHARs for Equation (1). In Panel E, for each firm, we demean Speculation with the firm average of Speculation. In Panel F, we decompose the sentiment score into its positive and negative components. All other aspects of the regressions are the same as those in Table 4. (Z) indicates that the variable is standardized to have the mean of zero and the standard deviation of one. Standard errors (SEs) reported in parentheses are clustered by firm and filing year-month to account for the serial and cross-sectional correlations of BHARs, respectively. Estimated coefficients and SEs are reported in percentage. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% level, respectively.

| | Dependent variable = Cumulative BHARs | | | | | |
|---|---|---|---|---|---|---|
| | (1) Week[-4,-1] | (2) Week 0 | (3) Week[1,4] | (4) Week[5,8] | (5) Week[9,12] | (6) Week[13,16] |
| ***Panel A: Excluding Observations with Future Earnings Announcements*** | | | | | | |
| Speculation (Z) | 0.166 | -0.023 | 0.456*** | 0.891*** | 0.803*** | 0.990*** |
| | (0.144) | (0.049) | (0.146) | (0.235) | (0.247) | (0.313) |
| ***Panel B: Removing Safe-harbor Boilerplate Paragraphs*** | | | | | | |
| Speculation (Z) | 0.131 | -0.070 | 0.477*** | 0.908*** | 0.905*** | 1.095*** |
| | (0.161) | (0.047) | (0.128) | (0.220) | (0.253) | (0.328) |
| ***Panel C: Controlling for Fama-French Three Factors' Loadings*** | | | | | | |
| Speculation (Z) | 0.085 | -0.071 | 0.477*** | 0.915*** | 0.904*** | 1.086*** |
| | (0.164) | (0.047) | (0.130) | (0.225) | (0.258) | (0.334) |
| Market Beta (Z) | 0.389* | 0.029 | 0.262 | 0.599* | 0.537 | 0.664 |
| | (0.226) | (0.040) | (0.192) | (0.329) | (0.475) | (0.529) |
| SMB Beta (Z) | 0.251 | -0.042 | -0.152 | -0.440 | -0.462 | -0.225 |
| | (0.199) | (0.053) | (0.227) | (0.352) | (0.328) | (0.369) |
| HML Beta (Z) | -0.308 | 0.046 | 0.299 | 0.478 | 0.324 | 0.034 |
| | (0.217) | (0.061) | (0.215) | (0.341) | (0.388) | (0.449) |
| ***Panel D: Using CAR*** | | | | | | |
| Speculation (Z) | 0.083 | -0.065 | 0.461*** | 0.823*** | 0.824*** | 0.945*** |
| | (0.122) | (0.047) | (0.123) | (0.208) | (0.244) | (0.304) |
| ***Panel E: Using Demeaned Speculation*** | | | | | | |
| Speculation (Z) | -0.077 | -0.045 | 0.340*** | 0.710*** | 0.651*** | 0.574** |
| | (0.128) | (0.037) | (0.123) | (0.221) | (0.243) | (0.274) |
| ***Panel F: Decomposing Sentiment into Positive and Negative*** | | | | | | |
| Speculation (Z) | 0.133 | -0.051 | 0.325** | 0.695*** | 0.745*** | 0.998*** |
| | (0.143) | (0.047) | (0.128) | (0.204) | (0.267) | (0.310) |
| Sentiment Positive (Z) | 0.040 | -0.014 | 0.277** | 0.413** | 0.179 | 0.157 |
| | (0.153) | (0.035) | (0.125) | (0.183) | (0.247) | (0.313) |
| Sentiment Negative (Z) | -0.167 | -0.030 | -0.139 | -0.202 | 0.140 | 0.004 |
| | (0.165) | (0.042) | (0.155) | (0.259) | (0.279) | (0.364) |

Table A.3: Speculative Language in Disclosure and Real Effects

This table investigates how the speculative language in 10-Ks relates to the real investments in the subsequent year and presents the estimation results of the following equation: For firm $i$ in year $t$,

$$Investment_{it+1} = \alpha + \beta Speculation_{it} + \gamma Q_{it} + \delta Speculation_{it} \times Q_{it} + \eta^{'}\mathbf{X}_{it} + \epsilon_{it+1},$$

where $Investment_{it+1}$ is one of real investment measures in year $t + 1$ that include R&D expenditures, capital expenditures, both R&D and capital expenditures, and total assets, $Speculation_{it}$ is the percentage of our speculation cues (out of the total words) in the 10-K disclosure in year $t$, and $\mathbf{X}_{it}$ is a column vector of control variables including Sentiment, interaction of Sentiment and Q, Cash flow, interaction of Speculation and Cash Flow, one-year lagged inverse value of Total assets, and value-weighted market return over the next three years as in Chen, Goldstein, and Jiang (2007). The detailed definitions of these explanatory variables are given in the Appendix B. When estimating the regression model, we also include firm and year fixed effects. (Z) indicates that the variable is standardized to have the mean of zero and the standard deviation of one. Standard errors (SEs) reported in parentheses are clustered by firm and year. Estimated coefficients and SEs are reported in percentage. $^{***}$, $^{**}$, and $^{*}$ indicate statistical significance at the 1%, 5%, and 10% level, respectively.

| | R&D+CAPX | R&D | CAPX | Total asset |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| Q (Z) | 2.840*** | 1.612*** | 1.132*** | 16.169*** |
| | (0.248) | (0.198) | (0.082) | (1.046) |
| Speculation (Z) | 0.291*** | 0.246*** | 0.040 | 0.102 |
| | (0.092) | (0.070) | (0.038) | (0.451) |
| Speculation (Z) x Q (Z) | 0.375** | 0.358** | -0.003 | -0.231 |
| | (0.146) | (0.134) | (0.046) | (0.716) |
| Sentiment (Z) | 0.187** | 0.040 | 0.160*** | 0.846** |
| | (0.068) | (0.054) | (0.029) | (0.384) |
| Sentiment (Z) x Q (Z) | 0.089 | 0.118 | 0.011 | -0.844 |
| | (0.100) | (0.069) | (0.056) | (0.566) |
| Cash flow (Z) | 0.439** | -0.098 | 0.532*** | 7.669*** |
| | (0.201) | (0.151) | (0.085) | (1.052) |
| Speculation (Z) x Cash flow (Z) | -0.077 | 0.001 | -0.034 | 1.265* |
| | (0.127) | (0.103) | (0.046) | (0.634) |
| Total asset inverse lagged (Z) | 4.092*** | 3.403*** | 0.450*** | 16.529*** |
| | (0.380) | (0.337) | (0.116) | (1.677) |
| Next 3-year market returns (Z) | -0.284*** | -0.026 | -0.254*** | -5.136*** |
| | (0.089) | (0.052) | (0.057) | (0.594) |
| Observations | 69297 | 69297 | 69297 | 69297 |
| Adjusted $R^2$ | 0.745 | 0.819 | 0.619 | 0.253 |

## Table A.4: Speculative Language in Disclosure and Informed Selling Activity

This table presents the estimation results of Equation (3) using *Probability of informed selling* as the dependent variable in place of *Probability of informed buying* as follows: For stock $i$ in year $t$,

$$Probability\ of\ informed\ selling_{itn} = \alpha_n + \beta_n Speculation_{it} + \eta_n^{'} \mathbf{X}_{it} + \epsilon_{itn}.$$

$Probability\ of\ informed\ selling_{itn}$ proxies the trading activity of informed sellers based on Brennan, Huh, and Subrahmanyam (2018) over the four-week window ending in the $n$th week where $n = 4, 8, 12, 16$ for the post-filing period and over the four-week window from Week -4 to Week -1, i.e., Week[-4,-1], for the pre-filing period. Week 0 has the four days between the 10-K filing day and three days later. $Speculation_{it}$ is the percentage of our speculation cues (out of the total words) in the 10-K disclosure, and $\mathbf{X}_{it}$ is a column vector of our baseline control variables: Sentiment, Market value, Book-to-market, Turnover, Institutional ownership, Fama-French alpha, and Filing-day abnormal return as in Table 4, and two additional variables: Nasdaq dummy and Pre-filing informed selling. The detailed definitions of these explanatory variables are given in the Appendix B. (Z) indicates that the variable is standardized to have the mean of zero and the standard deviation of one. When estimating the regression model, we also include filing year-month fixed effects. Standard errors (SEs) reported in parentheses are clustered by firm and filing year-month. Estimated coefficients and SEs are reported in percentage. $^{***}$, $^{**}$, and $^{*}$ indicate statistical significance at the 1%, 5%, and 10% level, respectively.

| | Dependent variable = Probability of informed selling | | | | | |
|---|---|---|---|---|---|---|
| | (1) Week[-4,-1] | (2) Week 0 | (3) Week[1,4] | (4) Week[5,8] | (5) Week[9,12] | (6) Week[13,16] |
| Speculation (Z) | -0.287 | 0.149 | -0.164 | -0.152 | -0.152 | -0.182 |
| | (0.257) | (0.269) | (0.120) | (0.123) | (0.120) | (0.129) |
| Sentiment (Z) | 0.462** | -0.054 | -0.042 | -0.051 | -0.211** | 0.013 |
| | (0.194) | (0.132) | (0.078) | (0.089) | (0.105) | (0.068) |
| Market value (Z) | -0.715 | 0.318 | 1.401*** | 1.503*** | 1.723*** | 1.418*** |
| | (0.516) | (0.415) | (0.239) | (0.233) | (0.321) | (0.365) |
| Book-to-market (Z) | -0.160 | 0.467*** | 0.293*** | 0.196* | 0.319*** | 0.183 |
| | (0.237) | (0.177) | (0.107) | (0.105) | (0.111) | (0.116) |
| Turnover (Z) | 0.774** | -0.880*** | -2.060*** | -2.214*** | -2.090*** | -1.915*** |
| | (0.305) | (0.228) | (0.147) | (0.149) | (0.174) | (0.147) |
| Institutional ownership (Z) | 0.167 | 0.784*** | 0.937*** | 1.124*** | 1.196*** | 0.669*** |
| | (0.362) | (0.258) | (0.123) | (0.131) | (0.148) | (0.152) |
| Fama-French alpha (Z) | -0.594** | 0.370** | 0.277*** | 0.048 | -0.178 | -0.277** |
| | (0.257) | (0.158) | (0.089) | (0.100) | (0.123) | (0.110) |
| Pre-filing informed selling (Z) | 54.765*** | 8.156*** | 4.099*** | 1.407*** | 0.371** | 0.248* |
| | (1.124) | (0.290) | (0.233) | (0.184) | (0.184) | (0.135) |
| Filing-day abnormal return (Z) | -0.122 | -2.733*** | 0.339*** | 0.244** | 0.151* | 0.007 |
| | (0.130) | (0.283) | (0.097) | (0.096) | (0.086) | (0.080) |
| Nasdaq dummy | -1.074 | 0.055 | 1.006*** | 0.773** | 0.882* | 1.135*** |
| | (0.669) | (0.528) | (0.379) | (0.392) | (0.498) | (0.411) |
| Observations | 45330 | 42850 | 43795 | 42964 | 42449 | 42114 |
| Adjusted $R^2$ | 0.735 | 0.123 | 0.130 | 0.091 | 0.124 | 0.091 |

## Table A.5: Speculative Language in Disclosure and Insider Selling Activity

This table presents the estimation results of Equation (3) using *Dollar volume of insider selling* as the dependent variable in place of *Probability of informed buying* as follows: For stock $i$ in year $t$,

$$\textit{Dollar volume of insider selling}_{itn} = \alpha_n + \beta_n \textit{Speculation}_{it} + \eta_n^{'} \mathbf{X}_{it} + \epsilon_{itn}.$$

*Dollar volume of insider selling*$_{itn}$ captures the insiders' selling activity in dollar volume (in logarithm) over the four-week window ending in the $n$th week where $n = 4, 8, 12, 16$ for the post-filing period and over the four-week window from Week -4 to Week -1, i.e., Week[-4,-1], for the pre-filing period. Week 0 has the four days between the 10-K filing day and three days later. *Speculation*$_{it}$ is the percentage of our speculation cues (out of the total words) in the 10-K disclosure, and $\mathbf{X}_{it}$ is a column vector of our baseline control variables: Sentiment, Market value, Book-to-market, Turnover, Institutional ownership, Fama-French alpha, and Filing-day abnormal return as in Table 4, and Nasdaq dummy. The detailed definitions of these explanatory variables are given in the Appendix B. (Z) indicates that the variable is standardized to have the mean of zero and the standard deviation of one. When estimating the regression model, we also include filing year-month fixed effects. Standard errors (SEs) reported in parentheses are clustered by firm and filing year-month. Estimated coefficients and SEs are reported in percentage. $^{***}$, $^{**}$, and $^{*}$ indicate statistical significance at the 1%, 5%, and 10% level, respectively.

| | Dependent variable = Dollar volume of insider selling | | | | | |
|---|---|---|---|---|---|---|
| | (1) Week[-4,-1] | (2) Week 0 | (3) Week[1,4] | (4) Week[5,8] | (5) Week[9,12] | (6) Week[13,16] |
| Speculation (Z) | 1.429 | -6.532* | 1.521 | 4.216 | 3.697 | 0.602 |
| | (2.068) | (3.500) | (2.232) | (2.664) | (2.586) | (2.369) |
| Sentiment (Z) | 6.170*** | 0.109 | -2.001 | 2.768 | 2.767 | 0.201 |
| | (1.954) | (2.597) | (2.303) | (2.161) | (2.063) | (2.220) |
| Market value (Z) | 94.223*** | 85.008*** | 89.734*** | 86.596*** | 88.550*** | 81.472*** |
| | (2.857) | (3.792) | (2.576) | (2.870) | (2.672) | (3.394) |
| Book-to-market (Z) | -11.762*** | -12.944*** | -14.366*** | -11.119*** | -13.068*** | -13.533*** |
| | (1.880) | (2.915) | (2.119) | (2.173) | (1.879) | (2.094) |
| Turnover (Z) | 29.167*** | 10.172*** | 14.272*** | 17.525*** | 11.799*** | 11.082*** |
| | (3.043) | (3.765) | (2.893) | (3.386) | (2.969) | (2.952) |
| Institutional ownership (Z) | 11.800*** | 16.964*** | 14.243*** | 17.011*** | 17.706*** | 13.568*** |
| | (2.387) | (3.931) | (2.873) | (2.843) | (2.456) | (2.915) |
| Fama-French alpha (Z) | 24.459*** | 33.360*** | 30.563*** | 22.555*** | 23.174*** | 23.924*** |
| | (4.066) | (4.845) | (3.163) | (2.969) | (3.214) | (2.685) |
| Filing-day abnormal return (Z) | -0.002 | 9.128*** | 14.356*** | 8.038*** | 11.113*** | 8.814*** |
| | (1.550) | (3.055) | (1.848) | (1.987) | (1.574) | (2.065) |
| Nasdaq dummy | -6.281* | -1.168 | -7.979** | -7.753* | -8.554** | -7.395* |
| | (3.719) | (4.625) | (3.701) | (4.321) | (3.447) | (4.316) |
| Observations | 19029 | 7209 | 18126 | 15442 | 20213 | 15108 |
| Adjusted $R^2$ | 0.315 | 0.254 | 0.266 | 0.261 | 0.279 | 0.232 |

Internet Appendix to:

# Speculative and Informative: Lessons from Market Reactions to Speculative Cues

Table IA.1: Top 100 Frequently Used Neighboring Bigrams by Saliency Scores

This table presents the top 100 meaningful (after removing all stop words) neighboring bigrams in the 10-K paragraphs that contain our speculation cues by saliency scores. We analyze 5,597,740 unique pairs of neighboring words in 10-K paragraphs and compute the saliency score of each bigram in paragraphs with the speculation cues relative to paragraphs without such cues.

| Rank | Bigram | Rank | Bigram |
|---|---|---|---|
| 1 | (adverse, effect) | 51 | (sole, discretion) |
| 2 | (adversely, affect) | 52 | (accounting, standard) |
| 3 | (material, adverse) | 53 | (impairment, test) |
| 4 | (internal, control) | 54 | (material, effect) |
| 5 | (third, party) | 55 | (significant, estimate) |
| 6 | (forward-looking, statement) | 56 | (business, day) |
| 7 | (market, value) | 57 | (property, right) |
| 8 | (company, belief) | 58 | (reasonable, basis) |
| 9 | (comprehensive, income) | 59 | (new, product) |
| 10 | (financial, reporting) | 60 | (trade, secret) |
| 11 | (actual, result) | 61 | (public, offering) |
| 12 | (market, price) | 62 | (regulatory, approval) |
| 13 | (financial, condition) | 63 | (holding, company) |
| 14 | (operating, result) | 64 | (closing, price) |
| 15 | (intellectual, property) | 65 | (stock, outstanding) |
| 16 | (management, belief) | 66 | (period, presented) |
| 17 | (adversely, affected) | 67 | (good, faith) |
| 18 | (fair, market) | 68 | (significant, role) |
| 19 | (loan, document) | 69 | (material, information) |
| 20 | (certifying, officer) | 70 | (made, known) |
| 21 | (economic, condition) | 71 | (obtain, reasonable) |
| 22 | (written, notice) | 72 | (involves, management) |
| 23 | (clinical, trial) | 73 | (voting, power) |
| 24 | (applicable, law) | 74 | (standard, require) |
| 25 | (market, condition) | 75 | (company, issued) |
| 26 | (share, outstanding) | 76 | (material, impact) |
| 27 | (materially, affect) | 77 | (report, financial) |
| 28 | (financial, institution) | 78 | (information, included) |
| 29 | (average, number) | 79 | (adverse, impact) |
| 30 | (future, cash) | 80 | (person, performing) |
| 31 | (product, candidate) | 81 | (materially, adversely) |
| 32 | (administrative, agent) | 82 | (material, misstatement) |
| 33 | (outstanding, share) | 83 | (security, act) |
| 34 | (reported, amount) | 84 | (maintaining, disclosure) |
| 35 | (material, weakness) | 85 | (requires, management) |
| 36 | (taxable, income) | 86 | (equivalent, function) |
| 37 | (materially, affected) | 87 | (reasonable, assurance) |
| 38 | (par, value) | 88 | (pay, dividend) |
| 39 | (differ, materially) | 89 | (company, also) |
| 40 | (carrying, value) | 90 | (either, party) |
| 41 | (future, period) | 91 | (overall, financial) |
| 42 | (act, rule) | 92 | (prior, written) |
| 43 | (ordinary, course) | 93 | (bank, holding) |
| 44 | (stock, price) | 94 | (estimate, made) |
| 45 | (make, estimate) | 95 | (certain, circumstance) |
| 46 | (fiscal, quarter) | 96 | (exclude, empty) |
| 47 | (equity, instrument) | 97 | (market, participant) |
| 48 | (circumstance, indicate) | 98 | (also, includes) |
| 49 | (reporting, period) | 99 | (financial, result) |
| 50 | (significant, deficiency) | 100 | (trading, day) |

# Table IA.2: Speculative Language in Disclosure and Weekly BHARs

This table presents the estimation results of Equation (1) as follows: For stock $i$ in year $t$,

$$BHAR_{itn} \quad = \quad \alpha_n + \beta_n Speculation_{it} + \eta_n^{'} \mathbf{X}_{it} + \gamma_n RMkt_{tn} + \epsilon_{itn}.$$

$BHAR_{itn}$ is defined as the cumulative return difference between stock $i$ and the CRSP value-weighted index relative to Week 0 up to the $n$th-week, where $n = -3, \ldots, 16$ and Week 0 has the four days between the 10-K filing day and three days later. $Speculation_{it}$ is the percentage of our speculation cues (out of the total words) in the 10-K disclosure, and $\mathbf{X}_{it}$ is a column vector of our baseline control variables based on prior studies including Sentiment, Market value, Book-to-market, Turnover, Institutional ownership, Fama-French alpha, and Filing-day abnormal return. $RMkt_{tn}$ is the cumulaitve market return to capture common economic shocks that affect all individual stocks over the same $n$-week period as $BHAR_{itn}$. The detailed definitions of these explanatory variables are given in the Appendix B. We estimate Equation (1) with cumulative BHARs over various estimation windows based on weekly intervals as the dependent variables. For the post-filing period, each cumulative BHAR is computed over the period from the start of the 1st week (i.e., the fourth day after 10-K filing) to the end of the $n$th week, where $n = 1, \ldots, 16$. For the pre-filing period, each "reverse" cumulative BHAR is computed over the period from the end of the week -1 to the start of the $n$th week, where $n = -3, \ldots, -1$. For example, the reverse cumulative BHAR for Week[-3,-1] is obtained by solving $(1 + BHAR_{[-3,-1]}) = \frac{1}{(1+BHAR_{-3})*(1+BHAR_{-2})*(1+BHAR_{-1})}$ for $BHAR_{[-3,-1]}$, where $BHAR_{-k}$ is the weekly BHAR for Week $-k$. (Z) indicates that the variable is standardized to have the mean of zero and the standard deviation of one. Standard errors (SEs) reported in parentheses are clustered by firm and filing year-month to account for the serial and cross-sectional correlations of BHARs, respectively. Estimated coefficients and SEs are reported in percentage. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% level, respectively.

| | Dependent variable = Cumulative BHARs | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | (1) Week[-3,-1] | (2) Week[-2,-1] | (3) Week -1 | (4) Week 0 | (5) Week 1 | (6) Week[1,2] | (7) Week[1,3] | (8) Week[1,4] | (9) Week[1,5] | (10) Week[1,6] | (11) Week[1,7] | (12) Week[1,8] | (13) Week[1,9] | (14) Week[1,10] | (15) Week[1,11] | (16) Week[1,12] | (17) Week[1,13] | (18) Week[1,14] | (19) Week[1,15] | (20) Week[1,16] |
| Speculation (Z) | 0.090 | 0.088 | -0.027 | -0.071 | 0.022 | 0.187** | 0.314*** | 0.479*** | 0.615*** | 0.696*** | 0.832*** | 0.927*** | 0.893*** | 0.905*** | 0.911*** | 0.919*** | 0.943*** | 0.934*** | 1.030*** | 1.113*** |
| | (0.136) | (0.109) | (0.072) | (0.047) | (0.056) | (0.094) | (0.116) | (0.129) | (0.160) | (0.183) | (0.206) | (0.226) | (0.245) | (0.241) | (0.250) | (0.263) | (0.281) | (0.301) | (0.306) | (0.340) |
| Sentiment (Z) | 0.151 | 0.119 | 0.108** | 0.021 | -0.049 | 0.003 | 0.065 | 0.127 | 0.119 | 0.165 | 0.235 | 0.185 | 0.072 | -0.007 | -0.051 | -0.089 | -0.157 | -0.208 | -0.077 | 0.010 |
| | (0.111) | (0.077) | (0.052) | (0.031) | (0.057) | (0.089) | (0.106) | (0.115) | (0.144) | (0.159) | (0.184) | (0.193) | (0.197) | (0.199) | (0.204) | (0.207) | (0.221) | (0.235) | (0.251) | (0.275) |
| Market value (Z) | -1.481*** | -1.090*** | -0.588*** | -0.098** | -0.061 | -0.138 | -0.294** | -0.541*** | -0.719*** | -0.895*** | -1.039*** | -1.183*** | -1.257*** | -1.343*** | -1.345*** | -1.511*** | -1.640*** | -1.803*** | -1.888*** | -2.060*** |
| | (0.164) | (0.137) | (0.089) | (0.050) | (0.076) | (0.097) | (0.120) | (0.147) | (0.183) | (0.212) | (0.253) | (0.301) | (0.340) | (0.357) | (0.371) | (0.391) | (0.416) | (0.444) | (0.457) | (0.447) |
| Book-to-market (Z) | -0.370*** | -0.256** | -0.122** | -0.028 | 0.089 | 0.130 | 0.147 | 0.139 | 0.133 | 0.157 | 0.122 | 0.163 | 0.213 | 0.161 | 0.045 | -0.072 | -0.087 | -0.125 | -0.130 | -0.220 |
| | (0.136) | (0.101) | (0.054) | (0.037) | (0.083) | (0.131) | (0.142) | (0.166) | (0.186) | (0.208) | (0.233) | (0.254) | (0.279) | (0.311) | (0.304) | (0.305) | (0.318) | (0.311) | (0.300) | (0.309) |
| Turnover (Z) | 1.460*** | 1.115*** | 0.601*** | -0.144** | -0.062 | -0.176 | -0.074 | -0.050 | 0.012 | 0.167 | 0.155 | 0.198 | 0.269 | 0.320 | 0.164 | 0.144 | 0.230 | 0.267 | 0.126 | 0.055 |
| | (0.315) | (0.234) | (0.106) | (0.065) | (0.120) | (0.234) | (0.285) | (0.345) | (0.367) | (0.440) | (0.475) | (0.520) | (0.568) | (0.564) | (0.566) | (0.571) | (0.582) | (0.608) | (0.627) | (0.625) |
| Institutional ownership (Z) | -0.575** | -0.441*** | -0.115 | 0.203*** | 0.113 | 0.051 | 0.054 | 0.082 | 0.013 | -0.102 | 0.025 | 0.049 | 0.076 | 0.014 | 0.149 | 0.125 | 0.125 | 0.121 | 0.215 | 0.195 |
| | (0.239) | (0.167) | (0.080) | (0.054) | (0.078) | (0.123) | (0.147) | (0.194) | (0.203) | (0.233) | (0.263) | (0.283) | (0.286) | (0.298) | (0.325) | (0.354) | (0.368) | (0.394) | (0.431) | (0.443) |
| Fama-French alpha (Z) | -2.450*** | -1.670*** | -0.902*** | -0.116** | 0.041 | -0.119 | -0.064 | -0.068 | -0.022 | 0.009 | -0.051 | -0.258 | -0.247 | -0.035 | 0.072 | 0.264 | 0.391 | 0.305 | 0.423 | 0.523 |
| | (0.349) | (0.284) | (0.141) | (0.051) | (0.119) | (0.256) | (0.240) | (0.250) | (0.301) | (0.341) | (0.367) | (0.394) | (0.394) | (0.381) | (0.379) | (0.385) | (0.407) | (0.422) | (0.430) | (0.432) |
| Contemporaneous market return (Z) | 0.788** | 0.558* | 0.357 | 0.065 | 0.329** | 0.610** | 0.812** | 0.701** | 0.908** | 1.026*** | 1.093** | 1.207** | 1.627*** | 1.728*** | 1.684*** | 1.700*** | 2.014*** | 2.079*** | 1.932*** | 1.654*** |
| | (0.342) | (0.305) | (0.227) | (0.081) | (0.155) | (0.291) | (0.337) | (0.315) | (0.417) | (0.395) | (0.425) | (0.488) | (0.474) | (0.433) | (0.435) | (0.462) | (0.529) | (0.524) | (0.545) | (0.603) |
| Filing-day abnormal return (Z) | 0.159 | 0.098 | 0.186*** | - | -0.194** | -0.180 | -0.211* | -0.224** | -0.264** | -0.252* | -0.244 | -0.265 | -0.488** | -0.454* | -0.392* | -0.375 | -0.345 | -0.385 | -0.410 | -0.397* |
| | (0.103) | (0.090) | (0.059) | - | (0.079) | (0.112) | (0.109) | (0.111) | (0.128) | (0.146) | (0.149) | (0.177) | (0.224) | (0.248) | (0.235) | (0.232) | (0.256) | (0.255) | (0.249) | (0.234) |
| Observations | 63186 | 63172 | 63141 | 63199 | 62261 | 61837 | 61632 | 61621 | 60603 | 60707 | 60798 | 60842 | 59848 | 59983 | 60023 | 60061 | 59156 | 59267 | 59297 | 59297 |
| Adjusted $R^2$ | 0.051 | 0.037 | 0.022 | 0.002 | 0.004 | 0.007 | 0.008 | 0.007 | 0.010 | 0.011 | 0.010 | 0.011 | 0.013 | 0.013 | 0.012 | 0.012 | 0.014 | 0.014 | 0.013 | 0.012 |