Speculative and Informative: Lessons from Market Reactions to Speculation Cues

J. Anthony Cookson S. Katie Moon Joonki Noh*

December 28, 2021

*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, Arizona State University, Brigham Young University, Case Western Reserve University, Drexel 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, Wendy Rotenberg, Noah Stoffman, Shawn Thomas, Ralph Walkling, Brian Waters, 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

Speculation 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 10 weeks 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 use more R&D terms. Together, our results imply that speculative statements in 10-Ks contain new information on positive but yet immature prospects of future cash flows.

1 Introduction

Scientific writing often includes expressions of belief or likelihood as scientific research processes involve interpreting results and inductive reasoning to draw conclusions. When it comes to firm financial disclosures, such expressions of belief or likelihood (in other words, speculation) can be regarded as the lack of clarity and may concern investors and regulators. Speculative language in financial disclosures have understudied, if not, in the academic literature, and the closest possible strand of the literature is on linguistic complexity. Furthermore, the previous literature on complex language 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)).

Relating to the disagreement on complex language, in this paper, we focus on the extent of speculation in financial disclosures and examine whether speculative language contains more or less information. We develop a novel measure for the degree of speculation in financial disclosures — the number of speculation cues extracted from firms' 10-K filings — and use it to understand the information content of speculative statements. We find that greater speculation in financial disclosures predicts permanent and positive abnormal returns, improvements to stock liquidity, the arrival of positive-sentiment news about the firm, and greater intensities of insider and informed buying (but not those of insider and informed selling). Together, these findings support the view that management speculation conveys value-relevant firm activities.

It is empirically challenging to measure the extent of speculation in financial disclosures for at least two reasons. First, the selection of a list of speculation keywords 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 keywords. 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 that are based on unverifiable information — "weasel tags." Specifically, we analyze the text of Wikipedia articles and, more importantly, the weasel tags embedded into these articles. Wikipedia advises its users to attach weasel tags when they encounter sentences or phrases in Wikipedia articles that accompany unverifiable information. By appealing to Wikipedia's crowdsourced solution, we simultaneously provide an externally-reliable basis for identifying cues associated with speculation and "tie our hands" by eliminating researchers' subjective choices typically involved in building a dictionary of keywords.[1]

Using our dictionary of speculation cues, we generate a measure for the extent of speculation at the firm-year level by computing the fraction of speculation cues in each firm's annual 10-K filing. Consistent with the idea that our measure captures the management speculation in 10-Ks, we find that 10-Ks with greater extent of speculation tend to exhibit greater uncertainty and to contain more modal words that convey differing shades of meaning. Yet, we uncover that the information contained in our speculation measure has distinctive and unique aspects beyond existing textual measures proposed in earlier studies. We also find that 10-K disclosures with higher extent of speculation tend to have higher positive sentiment.

For a 10-week window following the 10-K disclosure, our core findings are that the extent of speculation in 10-Ks predicts (i) positive return 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

[1]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.

selling, and (iv) more positive-sentiment news, using news sentiment data available from RavenPack. With the exception of insider purchases, which are likely informed before the 10-K disclosure, these outcomes all exhibit no pre-trends in the 5-week period prior to the 10-K disclosure, and there is no perceptible effect of speculation past 10 weeks. Together, these core empirical results provide compelling evidence that the use of speculative language in 10-K disclosures reflects the value-relevant information on firms' fundamentals, which contrasts with the view that management speculation adds more uncertainty.

As a complement to these core tests, we also evaluate the textual content of the disclosures that drives these results. Based on this textual content, our speculation measure appears to reflect immature information on upcoming positive but yet uncertain prospects. Specifically, we detect two notable and 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 uncertain, 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"). It is important to note that both forward-looking disclosure and disclosure on R&D activities are almost entirely voluntary (Muslu, Radhakrishnan, Subramanyam, and Lim (2015) and Jones (2007)). Therefore, when information is voluntarily disclosed before the disclosure is mandated, managers have incentives to disclose relatively good news (Verrecchia 1983, Dye 1985, Skinner 1994). Building on these textual indicators, we conduct a series of heterogeneity tests and find that the cumulative market reactions to the speculative language in 10-Ks are stronger for firms that use more forward-looking terms and disclose more about R&D terms. These findings corroborate our interpretation that the speculative language in 10-Ks reflects uncertain, yet valuable, positive information.

We also reinforce the interpretation from our empirical evidence with a simple conceptual framework in the following section by illustrating why positive (vs. negative) private information about future events is more likely to be disclosed early using speculative language in a 10-K filing. In the theoretical framework, we analyze a manager's decision to

disclose certain private information at a given moment jointly derived from multiple properties of the information including realization timing, direction, and signal strength. This exercise provides us with a predication that only the private information on potentially positive events in the future will be voluntarily disclosed by management using speculative language.

Several aspects of our test results contrast with notable alternative interpretations. First, we observe no pre-trends in abnormal returns prior to the 10-K disclosure date and find that there is no reversal in the returns after 10 weeks. 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.[2] Second, during the same time frame when there are significant positive abnormal returns, stock liquidity also improves, contrasting further with a risk-based explanation. Third, consistent with the realization of positive information in the 10 weeks following 10-K disclosure date, we find that both the probability of informed buying and news sentiment about the firm with high extent of speculation are greater. Furthermore, during the entire test period around 10-K disclosures with high speculation, we find that insiders purchase significantly more of the firm's shares. All these results collectively are inconsistent with a risk-based explanation.[3]

Our paper makes several contributions to the existing literature. First, our evidence on the use of speculative language in firm disclosures relates to the work on the 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, Ely 2017). Following this line of research, recent empirical applications have focused on how the disclosure of bad news can signal firm's quality (Gormley,

[2]Beyond the evidence on this pattern of returns, our finding is not driven by the difference in systematic risks captured by the Fama-French three-factor model. Our finding 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). See Section 5.1 and Appendix Table A.3 for further discussion.

[3]Complementary to this main set of test results, we also find that the speculative language in 10-Ks is associated with future positive earnings surprises. See Appendix Table A.8.

Kim, and Martin 2012, 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 for the extent of speculation is more related to providing early access to information. Our results suggest that managers act in their decisions to provide more voluntary information on immature but positive opportunities.

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, 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, Garcia 2013, 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 at the master dictionary by Loughran and McDonald (2011). Relative to these other textual measures, our speculation measure provides a useful description of the 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 insight on different textual analysis methodologies. In this respect, we anticipate fruitful applications of our speculation measure to understand better the information environment into which the speculative language can be injected.

The remainder of the paper proceeds as follows. Section 2 presents our conceptual framework for a manager's decision to disclose certain private information using speculative language. Section 3 provides a description of Wikipedia's weasel tags, the construction of our dictionary of speculation cues, and the development of our speculation measure in 10-

Ks. Section 4 describes our sample and presents results relating our speculation measure to other existing variables to validate it and to obtain a better understanding of which firms employ more speculative language in their 10-Ks. Section 5 provides the main empirical results from the tests that investigate the relations between the extent of speculation in 10-Ks and subsequent abnormal returns, liquidity, intensities of informed and insiders' trading, and news sentiment after 10-K filing. Section 6 concludes the paper.

2 Conceptual Framework

In this section, we present a theoretical framework on what our speculation measure can capture. We analyze a manager's decision to disclose certain private information in the 10-K in a given year jointly derived from the following three properties of the information: i) realization timing, ii) direction, and iii) signal strength. First, we assume that the content of information is for either current or future realizations of events. For example, discussing earnings of the company for the current fiscal year is subject to the former case, while discussing investment opportunities for the next fiscal year is subject to the latter case. Second, we assume that the information is on either positive or negative news of the firm. For example, the manager's private information can be about the success or failure of a clinical trial that her firm has been conducting that year. Lastly, we assume that the information is based on a strong or weak signal. We expect the manager to use more speculative language when discussing her private information when it is based on a weak signal versus a strong signal. We illustrate these three properties of the private information, all in binary forms for simplicity, in Figure 1 and present five possible disclosure scenarios. For each scenario, we derive the disclosure choice by the manager based on the SEC's disclosure regulations and the previous theories on voluntary disclosures in the literature.

[Insert Figure 1 Here]

In scenario A, as the information content is subject to current realizations of events, regardless of whether the information is good or bad, it is mandated to be disclosed under

the SEC's Regulation FD ("Fair Disclosure") that regulates the disclosure of nonpublic material information. We note that the signal for this type of information must be strong as the information is about current events. Therefore, the information in scenario A is disclosed by mandates without using speculative language.

Next, in both scenarios B and C, the information is for future realizations of events and positive at the same time. The information is based on a strong signal in scenario B, and it is based on a weak signal in scenario C. As the timing of the information is forward-looking, the disclosure of this type of information is voluntary. Many previous studies in the literature on voluntary disclosures find that managers leak or reveal good news immediately to investors (Verrecchia 1983, Dye 1985, Kothari, Shu, and Wysocki 2009, Acharya, DeMarzo, and Kremer 2011). We expect that such incentives for immediate releases of good news will exist regardless of the strength of the signal. Therefore, based on these theory models, we predict that the information in scenarios B and C will be disclosed voluntarily but only the information in scenario C with a weak signal will entail speculative language.

Analogously, in scenarios D and E, the information is for future realizations of events but negative this time. The information is based on a strong signal in scenario D, and it is based on a weak signal in scenario E. Interestingly in scenario D, although the disclosure of the information on future events is voluntary, the manager is more likely to disclose the negative information voluntarily due to the certain nature of the information. Such a management choice to disclosure negative certain information immediately is related to a manager's incentive to avoid high legal costs as shown in Skinner (1994). Therefore, we predict that negative forward-looking information with a strong signal will be disclosed without using speculative language. On the other hand, when the negative information is based on a weak signal as in scenario E, it is less likely to be disclosed until the signal becomes stronger in the future. A recent paper by Bao, Kim, Mian, and Su (2019) shows empirical evidence on managers' incentive to withhold bad news up to a certain threshold, which is supported by multiple theory models as in Verrecchia (1983), Dye (1985), Kothari, Shu, and Wysocki (2009), and Acharya, DeMarzo, and Kremer (2011).

Although the five scenarios we analyze in this section are an extremely simplified version of modeling the complex nature of private information, they effectively identify the type of private information that will be voluntarily disclosed by managers using speculative language. We find from Figure 1 that only scenario C makes the information to the actual disclosure in the form of speculative statements. Thus, a testable prediction derived from the analysis is that managers' private information about future positive events is more likely to be disclosed by management using speculative language. Other types of private information will be either disclosed with non-speculative language or withheld until it receives a stronger signal or is required to be disclosed.

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." Our approach is to appeal to an external source, Wikipedia, which effectively crowdsources our identification of speculative language. Specifically, we take the entire Wikipedia articles 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) suggest three broad categories of weasel words or phrases used in Wikipedia articles: 1) numerically vague expressions (e.g., many), 2) the passive voice (e.g., it is said), and 3) adverbs that weaken (e.g., probably). Examples of these weasel words directly given 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 and at the same time to detect and mark excessive uses of such words by others using a special weasel tag, {{Weasel-inline—{{subst:DATE}}}}, for improvement. The following examples illustrate how the weasel tag is used in a sentence of each Wikipedia article:

- "The Tic Tok Men"

  Many{{weasel inline—date=March 2009}} consider this album to be the quintessen-

tial 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 completed on April 20, 2017 and comprised of 17,483,910 articles, and extract sentences that contain weasel tags.[4] 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 not abundant 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.

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 their bigrams and trigrams as well to better identify potential weasel words and phrases. The bigrams and trigrams are particularly useful to capture weasel phrases that use passive voice and appeal to anonymous authority. We note that a raw sort of the frequency of word usage will not accurately capture the most distinctively 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

[4]Wikipedia dumps are available for downloading at `https://dumps.wikimedia.org/`. For additional context on the nature of weasel tagged sentences, see Wikipedia's own article about weasel words at `https://en.wikipedia.org/wiki/Weasel_word`.

to show up as most frequent, despite not being weasel words themselves (e.g., words like "the", "and", and "that"). This is a much larger issue with the unigrams than it is 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 from the same Wikipedia article. By inspecting them manually, we see 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 text. 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).[5] This saliency measure captures the degree to which the words are overused relative to common language, and is thus, appropriate for screening a list of common language. Panel A(ii) of Table 1 shows how effective the saliency screen that we use is in filtering out common language from the list of words. Essentially, all of the common words (e.g., "the", "and") are least salient, and are thus, 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 (unigrams, bigrams, 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 language twice. In addition, Wikipedia has published guidelines for weasel words with specific examples to

[5] $Salience(word|weasel\ sentence) = p(word\&weasel\ sentence) \times log(\frac{p(weasel\ sentence|word)}{p(weasel\ sentence)})$ is used as an equation for the saliency filter. We also consider a re-weighted version of the list for speculation cues using term frequency–inverse document frequency (tf-idf) weights that mirrors the intuition of the Goldsmith-Pinkham, Hirtle, and Lucca (2016) saliency filter. The saliency filter and tf-idf weighting are similar methodologies that filter out common language. We prefer to use the saliency filtered list for speculation cues since it is more transparent and involves fewer researcher choices.

help users identify weasel language. Our methodology captures the vast majority of the example 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 in our final list. The complete dictionary for speculation cues is present in Appendix A.

Our dictionary for speculation cues is distinct from notable alternatives. 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.[6] Also, a number of 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 2015 and extract the raw counts of how many times a given firm mentions each of the speculation cues in a given year. This generates a full panel of speculation cue vectors with 219,491 firm-year observations. Our final sample is reduced to 46,996 firm-year observations after merging with the Compustat and CRSP databases (in Table 2). The number of firm-year observations decreases further to approximately 30,000 when the sample is merged with the product market threats and financial constraints data from Hoberg, Phillips, and Prabhala (2014) and Hoberg and Maksimovic (2015), respectively, for

[6]The most frequently used unigram, "other", can be simply mentioned in 10-Ks to refer to an accounting item that contains "other", for example, as in "Other Comprehensive Income", "Assets - Other", "Liabilities - Other - Total". We note that our findings discussed in the subsequent sections are robust to excluding "other" from our dictionary for speculation cues.

some analysis.

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

To provide a contextual understanding of our speculation measure, we examine neighboring unigrams and bigrams that co-exist with our speculation cues in 10-K disclosures. Neighboring unigrams and bigrams are those that occur in the same paragraph that contains any of the speculation cues. 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 for verbs, the most frequently mentioned neighboring words are "will", "require(d)", "expected", and "estimated." These words appear to be 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", and "future" is nearly as common. These two words are also associated with forward-looking disclosures. The most frequently mentioned adjective or adverb was "approximately." It is worth noting that 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 the contents of information. 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. Appendix Table A.1 presents the complete list of the top 100 neighboring bigrams that are overused relative to common language with the saliency screen.

We also examine time-series variations in the degree to which speculation cues are used to describe the same topic in a firm's 10-Ks. As a practical example, we present below excerpts from the MD&A sections of Technical Communications Corp's 10-K filings discussing "Liquidity and Capital Resources - Cash Requirements", where our speculation cues are underlined and forward-looking words are italicized based on Muslu, Radhakrishnan, Subramanyam, and Lim (2015).

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 but increases its usage of speculative language in 2010 and more significantly so in 2015. We find that sentences containing speculation cues seem 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 with speculative language in 2010 and 2015 in contrast with its straightforward numerical description of the company's current cash situation in 2000.

Overall, the picture that emerges from examining words neighboring our speculation cues and examples of time-varying usages of speculation cues in 10-K filings within a firm is that corporate disclosures containing speculative language are more likely to express shades of possibility and convey forward-looking information that is by nature less specific and more immature.

3.3 Discussion on the Speculative Language in 10-K Disclosures

Before describing our empirical tests, it is important to comment on the meaning of speculative statements within the context of 10-K disclosures relative to other potential source texts. 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 audited also — 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 the earnings conference call).[7] 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 disclosures.[8]

Because of this high degree of care in preparing the 10-Ks, the speculative language in 10-Ks is more deliberate than other source texts. With this background in mind, 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.

4 Validation and Relation to Firm Characteristics

In this section, we conduct a series of empirical analyses to obtain an understanding of firms' usage of speculative language in 10-K disclosures: Which firms employ more speculative language in their 10-Ks than others under what situations? These analyses provide useful validation of our textual measure before we examine various market reactions to it in Section 5.

4.1 Summary Statistics

Table 2 presents the basic summary statistics for various textual tonal variables (in Panel A) and non-tonal firm characteristics (in Panel B).[9] Each variable is winsorized at the top

[7] 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.

[8] To mitigate a potential concern that our speculation measure merely picks up boilerplate warnings that are not sufficient as meaningful disclosures of value-relevant information, we conduct a robustness check by excluding paragraphs deemed to be boilerplate warnings in 10-Ks. See Section 5.1 and Appendix Table A.3.

[9] In our stock universe, we include common stocks (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 over the past 12-month periods before 10-K filing dates from the CRSP database. We exclude stock-days with trading volume below 100 shares, with return below -1.0, or in a month with less than 15 days of valid return and volume data. We also delete stock-days with prices less than $3 to mitigate potential concerns that might be caused by the bid-ask bounce. These stock filters follow those in Jegadeesh and Wu

and bottom 1% of its distribution, and their definitions are provided in the Appendix B.

[Insert Table 2 Here]

For the textual tonal variables in Panel A of Table 2, 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 percentage. The mean and median of *Speculation* are 1.387% and 1.471%, respectively. The average of Sentiment that is the difference between the percentages of positive words and negative words (out of total words) is -0.716%, indicating that negative sentiment dominates positive one in our sample of 10-K filings. On average, 30% of words are considered as complex (Fog) words, and uncertain or litigious words are mentioned as many times as our speculstion cues.

As for non-tonal firm characteristics, the average of market value of assets is approximately $1.18 billion (Size in logarithm), and the average of firm age (Age) is roughly 11 years. We include two growth opportunities proxies, Tobin's Q and Sales growth, whose means are 1.94 and approximately 10%, respectively. We also consider two proxies for the economic conditions that firms face. Those are Product market fluidity by Hoberg, Phillips, and Prabhala (2014) and Financial constraints by Hoberg and Maksimovic (2015). For the analysis on how markets react to our speculation measure in Section 5, we include share turnover (Turnover), book-to-market ratio (Book-to-market), percentage of institutional investors' holdings (Institutional ownership), risk-adjusted return before 10-K filing (Fama-French alpha), and Filing-day abnormal return as control variables.

(2013) and Amihud and Noh (2020). Our empirical findings in Section 5 are robust to different choices for the low price cutoff ranging from $1 to $5. Based on this stock-day universe, we construct all lower frequency variables (weekly, monthly, and yearly) for our subsequent analyses.

4.2 Relations to Other Textual Tonal Variables

Although 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) because speculative language is appropriate to use in the presence of uncertainty. We validate this intuition of speculation 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, weak modality, and strong 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 are substantial overlaps in the distributions of the speculation cues for high and low uncertainty, weak modality, and strong 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. Panel A of Table 3 reports 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 Panel A, we include Fog to quantify the complexity of 10-K disclosures and Uncertain and Modal constructed based on the master dictionary by Loughran and McDonald (2011) as independent variables. Column (1) shows that complexity, uncertainty, and modality are all positively associated with our speculation measure, as expected,

even when controlling for unobserved firm characteristics by including firm fixed effect. In Column (2), we additionally examine the relations of speculative language to Sentiment and two other textual tonal variables, Constraining and Litigious, which capture firm's discussion on constraining and litigious situations, respectively. We find that our measure of speculation is positively associated with Sentiment, Constraining, and Litigious, suggesting an interpretation that firms use more speculative language when they discuss their positive prospects that likely have not been realized under those negative situations. In Column (3), we add controls for Superfluous and Interesting and find that the results in Columns (1) and (2) remain intact. Although we do not report the results to conserve space, we also find that these test results are robust to controlling for Size, Age, Tobin's Q, and Sales growth additionally, in terms of the magnitudes and statistical significance.

Taken together, the evidence in this section suggests that speculative language used in 10-Ks captures relatively positive tone with high uncertainty and high modality. Because uncertainty and modality aspects of the text are to a large degree parts of the content of speculative language, we do not control for Uncertain and Modal in our subsequent market-reaction tests in Section 5.

4.3 Relations to Non-tonal Firm Characteristics

As a second set of validation evidence, we relate our speculation measure 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)) which 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 with the regression models in Panel B of Table 3. 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 the speculation measure. In Column (1) of Panel B, we consider the first set of non-tonal firm characteristics employed in Figure 3, i.e., Size, Age, Tobin's Q, and Sales growth, which are lagged by one year. The result in Column (1) of Panel B shows that the strong associations between the speculation measure and Age and Tobin's Q in Figure 3, are also present in the regression analysis.

The next set of non-tonal firm characteristics include proxies for product market threats and financial constraints, which are also lagged by one year. Column (2) of Panel B summarizes the estimates, where the speculation measure additionally exhibits a significant and negative association with Size. We find significant positive relations between our measure of speculation and both Product market fluidity and Financial constraints, which are, respectively, proposed by Hoberg, Phillips, and Prabhala (2014) and Hoberg and Maksimovic (2015). This supports the idea that product market threats or financial constraints place pressure on firms to disclose the information, which is potentially useful to alleviate their difficult situations in product markets or financial markets, with using speculative language in their 10-K disclosures.

Finally, in Column (3), we consider a list of variables that have been known to affect firms' returns on the event days of 10-K releases. Those variables include Turnover, Institutional ownership, and Fama-French alpha, which are calculated over periods before 10-K filing dates, and Filling-day abnormal return.[10] The test results in Column (3) show that these firm characteristics are not significantly related to the use of speculative language in 10-Ks.

[10]We do not include Market value and Book-to-market in Column (3) since their inclusion is redundant due to Size and Tobin's Q.

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 Return 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 weekly windows around its filing and test whether our speculation measure predicts abnormal returns using the following regression specification. For stock i, over the nth week around its 10-K filing in year t,

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

where $BHAR_{itn}$ is defined as the return difference between stock i and the CRSP value-weighted index over the nth-week window,[11] $Speculation_{it}$ 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 nth-week market return to capture common economic shocks that affect all individual stocks' abnormal returns over the nth-week around each 10-K filing. All of the baseline explanatory variables are constructed based on the information available as of the 10-K filing date, and their detailed definitions are provided in the 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) for each week separately over the 3-week period prior to the 10-K release date and the 9-week period after the 10-K release date (thus $n = -3, \ldots, 9$). We

[11] The 1st week window starts from the fourth day (inclusive) and ends at the tenth day (inclusive) after the 10-K release date. The 10-K filing-day window, i.e., the 0th week window, covers the four days between the 10-K filing day (inclusive) and three days later (inclusive). The window of week -1 covers seven days before the 10-K filing day.

cluster standard errors by firm and filing year-month to account for potential serial and cross-sectional correlations of abnormal returns, respectively.[12]

The coefficient of interest in equation (1) is β_n, which captures how each stock's price in the nth week reacts to the speculative language used in its 10-K disclosure. If speculative language in 10-K disclosures contains positive value-relevant information about firms that takes time for investors to digest, we expect positive and delayed price reactions to the speculative language in 10-Ks after the release dates.

[Insert Table 4 Here]

The results from estimating equation (1) are presented in Table 4. For clarity of presentation, we only report coefficient estimates of three important variables, *Speculation*, Sentiment, and nth-week market return. Appendix Table A.2 presents the slope coefficient estimates of all control variables. The results in Panel A of Table 4 support the interpretation that speculative language contains positive and immature value-relevant information. There is not an immediate return reaction, but we find a significant and positive return reaction to speculative language between 2 and 7 weeks after 10-K filing. In the pre-period, there is no detectable return reaction (i.e., no pre-trend or anticipation), and after week 7, the statistical significance wanes and the estimated coefficients are smaller (i.e., no return reversal).

As an alternative presentation of the results, we present the results for cumulative BHARs, relative to Week 0 (including the 10-K filing day) in Panel B of Table 4.[13] The

[12]Consistent with conventional wisdom in the literature that serial correlation is low for returns variables on average, we find that statistical significance without the firm cluster is nearly identical to that of the double clustering which we report in the main tests.

[13]Specifically, we make the Week 0 (i.e., four days between the 10-K filing day and three days later) the reference period using the following definitions of cumulative BHARs. For the post-filing weeks, 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 nth week, where $n = 1, \ldots, 9$. For the pre-filing weeks and the week 0, each "reverse" cumulative BHAR is computed over the period from the end of the week 0 (i.e., the third day after 10-K filing) to the start of the nth week, where $n = -3, \ldots, 0$. For example, the reverse cumulative BHAR for Week[-1,0] is obtained by solving $(1 + BHAR_{[-1,0]}) = \frac{1}{(1+BHAR_{-1})*(1+BHAR_0)}$ for $BHAR_{[-1,0]}$, where $BHAR_{-1}$ and $BHAR_0$ are the weekly BHARs over the week -1 and the week 0, respectively. For each of the cumulative BHAR regressions, we control for the nth-week cumulative market return calculated in the same way based on weekly market returns up to the nth week.

positive and significant return reaction to speculative language in 10-K disclosures cumulatively emerges from the 3rd week, monotonically increases until the 9th week. The economic magnitude of this effect in Column of Week[1,9] translates into a 0.78% increase in BHAR over the nine-week period after the 10-K release with a one standard deviation increase in the use of speculative language in the 10-K disclosure (=0.46%).

To provide evidence of no pre-trends and no return reversal, we consider a wider window and report the results graphically. Figure 4 plots the coefficient estimates of *Speculation* for cumulative BHARs from five weeks before to 15 weeks after the 10-K filing date, together with the 95% confidence intervals to indicate statistical significance.

[Insert Figure 4 Here]

Figure 4 shows explicitly that the positive speculation effect is not preceded by any pre-trend for five weeks prior to 10-K filing, nor followed by a reversal through 15 weeks after a cumulative return of 0.78% in the 9 weeks. The positive and delayed price reaction to speculative language is consistent with our interpretation that speculative language in 10-Ks can provide positive but immature value-relevant information about firms to investors in financial markets.

We conduct multiple robustness checks and show the results in Appendix Table A.3. First, to ensure that the positive and significant price reactions to the speculative language in 10-Ks are not driven by future earnings announcements near 10-K filing dates, we repeat the estimation of equation (1) with a refined subsample that excludes all $BHAR_{itn}$ observations having earnings announcements over the three to seven weeks after 10-K filing dates. The results for this test are presented in Panel A of Appendix Table A.3 and show that the positive speculation effect is not mechanically related to upcoming future earnings announcements near 10-K filing. Second, we reconstruct our measure of speculation by excluding the count of speculation cues in the paragraphs that contain boilerplate warnings related to safe harbor provisions, which lack meaningful information. We then repeat the estimation of equation (1) using this alternative speculation measure. From the

results shown in Panel B, we find that our results are robust to excluding the safe harbor paragraphs.

Third, we demean *Speculation* with the firm average and estimate equation (1) with the firm-demeaned measure of *Speculation*. This test is to examine whether our speculation measure has sufficient time-varying information conveyed by varying language in 10-K disclosures. The results for this test are present in Panel C. We find that the positive and significant return effects of speculation are not merely cross-sectional. 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 $BHAR_{itn}$.[14] In Panel D, we find that the positive and significant price reactions to the speculative language in 10-Ks are robust to using an alternative measure of abnormal returns.

Finally, to ensure that our positive speculation effect is not driven by the difference in risk exposure, we estimate equation (1) after adding the sensitivities to Fama-French three factors based on the past one year before 10-K filing date as in calculating Fama-French alpha. In Panel E, we find that our findings are not explained by the risk exposures to the Fama-French three factors. In addition, we note that, to control for a quickly varying risk exposure in the spirit of Patton and Verardo (2012), we include Filing-day abnormal return as a control variable in equation (1). In sum, our speculation effect is unlikely driven by systematic risks.[15]

5.2 Liquidity Reaction

In this section, we examine how the use of speculative language in 10-Ks relates to illiquidity over weekly windows around 10-K release dates. If the positive price reaction to the spec-

[14]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.

[15]To the extent that security analysts are not involved in actual trading stocks, the predictability of future earnings surprise by our speculation measure presented in Appendix Table A.8 is also inconsistent with a risk-based explanation.

ulative language in 10-Ks is primarily driven by the release of value-relevant information and investors digest it over the subsequent period, we expect a negative relation between our speculation measure and firms' illiquidity or information asymmetry after 10-K filing with no pre-trend. To test this potential link between the extent of speculation and illiquidity level, we employ the following regression specification: For stock i, over the nth week around its 10-K filing in year t,

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

where $Spread_{itn}$ 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), over the nth-week window (in logarithm). We employ the same definition of weekly windows as those 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 the Appendix B. For ease of interpretation, explanatory variables are standardized to have the mean of zero and the standard deviation of one.

Existing studies have shown that 1) illiquidity level and risk are priced in the cross-section of asset returns, thus affecting investors' decisions significantly (Amihud and Mendelson (1986), Pástor and Stambaugh (2003), Acharya and Pedersen (2005)), and that 2) market illiquidity has improved substantially over time, leading to a reduced illiquidity premium (Amihud (2002), Ben-Rephael, Kadan, and Wohl (2015)). Following these earlier studies, 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. In addition, it has been known that illiquidity is persistent over time (Amihud (2002), Bali, Peng, Shen, and Tang (2014)) and has a commonality across assets (Chordia, Roll, and Subrahmanyam (2000)). To reflect these properties in estimating equation (2),

we cluster standard errors by firm and filing year-month to account for the serial and cross-sectional correlations of illiquidity level, respectively.

The coefficient of interest in equation (2) is β_n, which captures how speculative language in 10-Ks relates to illiquidity over the nth-week window around 10K filing. The estimation results of equation (2) are presented in Table 5, in which we only report the coefficient estimates of three important variables, *Speculation*, Sentiment, and Pre-filing spread. Those of all control variables are provided in Appendix Table A.4.

[Insert Table 5 and Figure 5 Here]

We find that the estimated coefficients for *Speculation* are negative and significant from the 1st through 9th weeks after 10-K filing. We also find no evidence of pre-trends as all coefficient estimates of *Speculation* before 10-K filing are not statistically different from zero. These results indicate that illiquidity decreases in the weeks following the filing of a 10-K with high speculation.[16] Interestingly, the coefficient estimates are largest in magnitude in the 3rd and 4th weeks after 10-K filing, which coincides with the time period when the stock price reaction is 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 the information is digested by investors in financial markets, thus leading to an improvement to liquidity after 10-K filing.

The graphical representation in Figure 5 shows that our finding on liquidity improvement holds for an extended window combining the previous 5-week period before the 10-K release date and the subsequent 15-week period after the 10-K release date. Indeed, the reaction of illiquidity to speculative language in 10-Ks is not preceded by any pre-trend and concentrated on the nine-week period following the 10-K release date.

[16] Further enhancing our confidence in the estimates, the control variables in equation (2) exhibit relationships to illiquidity that are consistent with prior work. For example, we find evidence that the estimated coefficients for Sentiment are negative in general and significant in the 7th and 8th weeks. This suggests that the percentages of positive and negative words in 10-Ks can contain value-related information, consistent with the pricing evidence of textual tones in Tetlock, Saar-Tsechansky, and Macskassy (2008), Loughran and McDonald (2011), and Jegadeesh and Wu (2013). It is worth noting that we include a pre-filing level of the quoted relative bid-ask spread in equation (2) and find that the significant negative effect of our speculation measure on illiquidity survives.

5.3 Informed Traders' Reactions

If speculative language contains positive value-relevant information and average investors take some time to digest it, it might be able to predict informed purchases (but not informed sales) 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 nth week around 10-K filing in year t,

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

where $\textit{Probability of informed buying}_{itn}$ proxies the trading activity of informed buyers, which is the average of daily (posterior) probability of informed buying proposed by Brennan, Huh, and Subrahmanyam (2018),[17] over the nth-week window. We use the same definition of weekly windows as those for the BHAR and illiquidity regressions. $\textit{Speculation}_{it}$ 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 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 the Appendix B. For ease of interpretation, explanatory variables are standardized to have the mean of zero and the standard deviation of one. We also control for filing year-month fixed effect and cluster standard errors by firm and filing year-month.

The coefficient of interest in equation (3) is β_n, which captures how the speculative language in 10-Ks affects the informed buying activity of corporate outsiders over the nth-week window after 10-K filing. The results from estimating equation (3) are presented

[17] 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 parameter of the PIN model using a three-month rolling window, and then calculate the daily posterior probability that a given trading day has a good news or bad news by conditioning on the number of buyer-initiated and seller-initiated trades on each day. In those models, 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).

in Table 6 in which we only report the coefficient estimates of three important variables, *Speculation*, Sentiment, and Pre-filing informed buying. Those of all control variables are provided in Appendix Table A.5.

[Insert Table 6 and Figure 6 Here]

We find that the estimated coefficients for our speculation measure are positive and significant for the 0th week (including the 10-K filing day) as well as six weeks afterward. This positive reaction of informed buying activities to speculation is strongest over the 0th and 1st weeks, indicating that the informed buyers' reactions to the speculative language in 10-Ks emerge earlier than the rise of stock price and improvement of liquidity over the 3rd and 4th weeks as shown, respectively, in Tables 4 and 5. We also find no evidence of pre-trends as all coefficient estimates of *Speculation* before 10-K filing are not statistically different from zero. It is worthwhile to emphasize that *Probability of informed buying*$_{itn}$ in equation (3) captures a *posterior* and *conditional* probability of informed buying, while *Spread*$_{itn}$ in equation (2) can be related to an *unconditional* probability of informed trading, which captures information asymmetry. Therefore, the positive coefficient of *Speculation* in equation (3) is not contradictory to the negative coefficient of *Speculation* in equation (2).

Figure 6(a) presents the results for informed buying activities over an extended window combining the previous 5-week period before 10-K filing and the subsequent 15-week period after 10-K filing. In addition, we present the analogous test results for informed selling activities in Figure 6(b). In contrast to informed buying activities, when we use the probability of informed selling also proposed by Brennan, Huh, and Subrahmanyam (2018) in Figure 6(b), we find that the coefficient estimates of *Speculation* are all smaller and statistically insignificant throughout the same extended window. Taken together, these test results support the interpretation that *positive* value-relevant information disseminates into markets with speculative language in 10-Ks and is digested by informed traders more quickly than other investors.

Next, we investigate how the informed buying activity by corporate insiders is 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 over the nth-week window (in logarithm).[18] Different from the tests on the probability of informed buying, we expect insider buying to be greater throughout all weekly windows because insiders have access to information prior to the 10-K filing date. The results are presented in Table 7 in which we only report the coefficient estimates of two important variables, *Speculation* and Sentiment. Those of all control variables are provided in Appendix Table A.6.

[Insert Table 7 Here]

We find that the estimated coefficients on *Speculation* are positive and significant throughout all weekly windows in Table 7. Interestingly, the estimated coefficient on *Speculation* is largest in magnitude in the 3rd and 4th weeks after 10-K filing, which is consistent with the timing of the return reaction to speculation and of the corresponding liquidity improvement. This matched timing of liquidity improvement and insider buying activities suggests that informed corporate insiders time their buying trades when stock liquidity becomes most in favor of them, which is consistent with Collin-Dufresne and Fos (2015) and Collin-Dufresne and Fos (2016). Looking at a wider event window, Figure 6(c) shows that the positive estimated coefficient on *Speculation* attenuates over time, becoming insignificant by the 13th week after 10-K filing. In comparison, Figure 6(d) presents the results for insider selling activities, which show no statistically significant association with speculative statements over all weekly windows. Taken together, these findings on insider trading activities reinforce the conclusion that *positive* value-relevant information accompanies the speculative language used in 10-K disclosures.

[18] Because of the change in the dependent variable, we also drop Pre-filing informed buying from the vector of control variables. For each firm i and year t, we construct *Dollar volume of insider buying*$_{itn}$ by aggregating all records of insider purchases in dollar volume over the nth-week window around 10-K filing date 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 in equation (3).

5.4 News Sentiment in the Subsequent Periods

We now test more directly whether speculative language in 10-K disclosures predicts subsequent news sentiment, which is informative of whether the information embedded in speculative statements is positive versus negative. Specifically, we employ the following specification: For stock i, over the nth week around 10-K filing in year t,

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

where $News\ sentiment_{itn}$ is the average news-related sentiment score over the nth-week window,[19] $Speculation_{it}$ 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 nth-week market news sentiment to proxy for the common economic shocks that potentially affect all individual firms' news sentiment scores over the nth-week around each 10-K filing.[20] 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 the 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 estimation results of equation (4) are presented in Table 8 in which we only report the coefficient estimates of four important variables, *Speculation*, Sentiment, Pre-filing news sentiment, and nth-week market news sentiment. The results for all control variables are provided in Appendix Table A.7.

[Insert Table 8 and Figure 7 Here]

We find that the estimated coefficients on *Speculation* are positive and significant from the 1st through 7th weeks after 10-K filing, indicating the arrival of good news about the

[19] We use the same definition of weekly windows as those for our earlier regression models in Sections 5.1 to 5.3. 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 observations with Relevance=100 and Novelty$\geq$75 are used for our analysis.

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

firm's prospects for weeks after the 10-K filing. The estimated magnitude on *Speculation* is greatest in the 3rd and 4th weeks after the 10-K filing date, which aligns with the timing of the return reactions. We also find no evidence of pre-trends since all coefficient estimates of *Speculation* before 10-K filing are not statistically different from zero. Figure 7 over an extended window also confirms that the news media describes firms with higher extent of speculation in their 10-Ks with more positive sentiment, implying the arrival of good news about those firms after their 10-K filing.[21]

5.5 Firm Heterogeneous Effects

In this section, we conduct a series of heterogeneity tests to shed light into the mechanisms underlying the return reaction to speculative language in 10-Ks. Specifically, we aim to understand whether the return reaction is driven by forward-looking statements, disclosure of proprietary information, and periods of uncertainty or difficulty in processing and trading on new information. To evaluate these questions, we perform a heterogeneity analysis in which we estimate equation (1) using cumulative BHARs and above-median versus below-median sample splits for five firm-specific variables: Forward-looking disclosure, R&D disclosure, pre-filing idiosyncratic volatility of returns, analyst coverage, and quoted relative bid-ask spread. These variables are constructed based on the information available as of 10-K filing dates.[22] Table 9 reports the subgroup estimation results of equation (1), and Figure 8 shows their graphical representations over an extended estimation window.

[Insert Table 9 and Figure 8 Here]

First, we evaluate whether the return reaction is greater for firms with more forward-

[21] In Appendix Table A.8, we also investigate the standardized unexpected earnings (SUE) predictability by our speculation measure. We find that speculative language in 10-Ks can predict future SUE positively and significantly for the SUE indicator and two rank variables. The positive effect attenuates over quarters indicating that security analysts initially under-react to the information accompanying speculative language in 10-Ks possibly due to its embedded immaturity, although they eventually digest and reflect its implication related to future cash flows into their earnings forecasts.

[22] We also ensure that the sample splits do not depend on any forward-looking information. Specifically, the median sample split is determined based on the distribution of each variable over the one-year period before each 10-K filing date.

looking disclosure based on the measure from Muslu, Radhakrishnan, Subramanyam, and Lim (2015).[23] In Panel A of Table 9, we find that the price reactions to the extent of speculation in 10-Ks are greater for firms that use more forward-looking terms. The positive and significant estimated coefficients on *Speculation* are present only in the high forward-looking disclosure group from the 2nd week onward, while those in the low forward-looking disclosure group are insignificant regardless of the horizon. Based on Column Week[1,9], we find that a one standard deviation increase in the use of speculative language in the high forward-looking disclosure group is associated with about 1.06% higher BHAR over the nine-week period after 10-K filing. In terms of its magnitude, this effect is stronger than that for the whole sample at 0.78% in Panel B of Table 4.

Second, we examine the role of proprietary information, using the R&D disclosure measure based on Merkley (2013).[24] Based on Panel B of Table 9, we find that the return reaction to the extent of speculation is larger for firms that disclose more R&D keywords. Based on Column Week[1,9], a one standard deviation increase in the use of speculative language in the high R&D disclosure group is associated with about 1.01% higher BHAR over the nine-week period after 10-K filing. The estimated magnitudes for the high R&D disclosure group throughout columns are similar to those for the high forward-looking disclosure group.

Third, we examine three indicators of firm's information environment to understand whether the return reaction to the extent of speculation is driven by periods of uncertainty or difficulty in processing and trading upon new information. Specifically, we consider sample splits based on pre-filing idiosyncratic volatility of return (Panel C), presence versus absence of analyst coverage (Panel D), and quoted relative bid-ask spread (Panel E).[25] In

[23]For the measure of forward-looking disclosure, we use the percentage of forward-looking disclosure keywords (out of total words) in a firm's 10-K.

[24]For the measure of R&D disclosure, we use the percentage of narrative R&D disclosure keywords (out of total words) in a firm's 10-K.

[25]The pre-filing idiosyncratic 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 the 10-K filing date. The pre-filing presence or absence of analyst coverage is determined based on the most recent quarter before the 10-K filing date.

these tests, we find that firms with greater idiosyncratic volatility, no analyst coverage, and lower liquidity experience relatively larger positive price reactions than their counterparts, starting from the 3rd or 4th week. For example, in Column Week[1,9] of Panel E, firms with high pre-filing information asymmetry (captured by quoted relative bid-ask spread) have about 1.7 times greater return reaction to speculative language than those with low pre-filing information asymmetry. From these test results, we infer that the return reaction to the extent of speculation is greater in the presence of greater uncertainty and in relatively more opaque information environments. These findings corroborate our interpretation of the information content of speculative language in 10-Ks as uncertain, yet positive, value-relevant information that is eventually digested by various market participants.

5.6 Discussion on Forward-looking Disclosure

Finally, we present an analysis that distinguishes the market reaction to the extent of speculative statements in 10-Ks from the market reaction to forward-looking statements. Specifically, we decompose our measure of speculation into two components: speculation in the paragraphs in which forward-looking disclosure keywords coexist and speculation in the other paragraphs. Apart from distinguishing from forward-looking disclosure itself, the non-forward-looking component of speculation captures disclosures that are speculative for other natural reasons (e.g., proprietary disclosures like R&D as in Panel B of Table 9).

For this analysis, we estimate equation (1) by replacing $Speculation_{it}$ with $Speculation_{it}^{FW}$ and $Speculation_{it}^{-FW}$. $Speculation_{it}^{FW}$ is the percentage of our speculation cues in paragraphs that also contain forward-looking disclosure keywords from Muslu, Radhakrishnan, Subramanyam, and Lim (2015). $Speculation_{it}^{-FW}$ is the percentage of speculation cues in other paragraphs that do not have the forward-looking disclosure keywords, calculated as $Speculation_{it}$ minus $Speculation_{it}^{FW}$. Table 10 presents the results from estimating this specification. Panels A and B use weekly BHARs and cumulative BHARs as the dependent variable, respectively.

[Insert Table 10 Here]

From Table 10, we find that the return reaction to the speculation is similar for both components of our speculation measure. For both components, the estimated coefficients are positive and significant from the 2nd or 3rd week with slightly larger return reactions to $Speculation^{-FW}$. Based on this evidence, we conclude that our speculation effect is not purely driven by forward-looking statements. Indeed, these findings corroborate our bigram analysis in Panel B of Table 1, which found many of the words near speculation cues had to do with the disclosure of proprietary information.

6 Conclusions

In this paper, we introduce a textual measure to the finance and accounting literature, which quantifies the degree of speculation in firms' disclosures with a minimum level of researchers' subjectivity. Our speculation measure is distinct from existing textual measures such as sentiment, uncertainty, and modality, and has 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 uncovers that this speculative language is most valuable when it accompanies forward-looking statements, statements about proprietary information, and the information environment surrounding the firm is relatively opaque. Thus, it is natural for firms to use speculative language during uncertain times to deliver new information on positive but yet immature prospects of future cash flows. Market participants initially under-react to the information contained in this speculative language possibly due to its embedded immaturity but they eventually understand and digest it. Collectively, our findings and approach suggest that there is much more to learn from the qualitative content of firm disclosures.

Appendix B. Variable Definitions

Speculation	is the number 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 filings (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 filings (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 filings (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 filings (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 filings (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 filings (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 filings (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 filings (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).
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	is a 10-K based textual measure for financial constraints from Hoberg and Maksimovic (2015) with higher values indicating that firms are more at the risk of delaying their investments due to issues with liquidity.
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 six days (inclusive) before 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 ending at six days before 10-K filing are required to be included in the sample.
Institutional ownership	is the percentage of institutional investors' holdings available from the CDA/Spectrum database for the most recent quarter before 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 ending at six days (inclusive) before 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).
Nasdaq dummy	is one if the stock is listed in the Nasdaq on the day before 10-K filing and zero otherwise.
nth-week market return	is the CRSP value-weighted index over the nth week around each 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 six days (inclusive) before 10-K filing and is based on the CRSP database.
Pre-filing informed buying	is the average of daily probabilities of informed buying based on Brennan, Huh, and Subrahmanyam (2018) over the period from the beginning of the prior month to six days (inclusive) before 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 six days (inclusive) before 10-K filing, where the daily news sentiment score is defined as (ESS-50)/50 based on the ESS variable from the RavenPack News Analytics database.
nth-week market news sentiment	is the average of daily market-wide news sentiment score over the nth week around each 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.
Analyst dispersion	is the standard deviation of analysts' forecasts for the most recent quarter prior to the earnings announcement used to compute SUE, divided by the stock price at the end of the quarter.
Analyst revision	is the change in the consensus (mean) of analysts' forecasts, divided by the stock price in the prior month before the earnings announcement day.
Market earnings-announcement-day return	is a value-weighted average of individual stocks' earnings-announcement-day returns across stocks that have earnings announcements within a given month, where their market capitalizations in the preceding month are used as weights. For each stock and each earnings announcement, the earnings-announcement-day return is defined as the cumulative return over the period between the earnings announcement day (inclusive) and the three days later (inclusive).

References

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

Acharya, V. V., Pedersen, L. H., 2005. Asset pricing with liquidity risk. *Journal of Financial Economics* 77, 375 – 410.

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

Amihud, Y., Mendelson, H., 1986. Asset pricing and the bid-ask spread. *Journal of Financial Economics* 17, 223–249.

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.

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

Ben-Rephael, A., Kadan, O., Wohl, A., 2015. The Diminishing Liquidity Premium. *The Journal of Financial and Quantitative Analysis* 50, 197–229.

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.

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

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

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.

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*.

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.

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.

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.

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.

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.

Merkley, K. J., 2013. 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.

Pástor, L., Stambaugh, R. F., 2003. Liquidity Risk and Expected Stock Returns. *Journal of Political Economy* 111, 642–685.

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.

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

Tetlock, P. C., Saar-Tsechansky, M., Macskassy, S., 2008. More than words: Quantifying language to measure firms' fundamentals. *Journal of Finance* 63, 1437–1467.

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

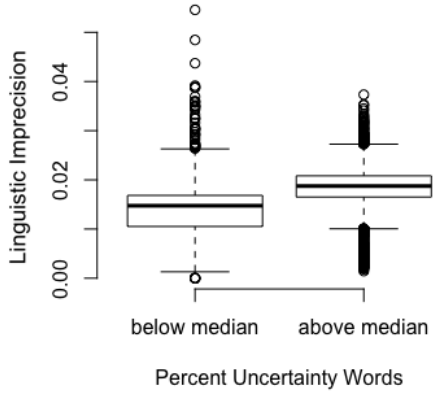
Figure 1: Management Disclosure Choices with Different Information Properties

This figure shows a theoretical framework that analyzes a manager's decision to disclose certain private information in the 10-K in a given year jointly derived from the following three properties of the information: i) realization timing, ii) direction, and iii) signal strength. First, we assume that the content of information is for either current or future realizations of events. Second, we assume that the information is on either positive or negative news of the firm. Third, we assume that the information is based on a strong or weak signal. We then present five possible disclosure scenarios with types of language based on the SEC's disclosure regulations and the previous theories on voluntary disclosures in the literature.

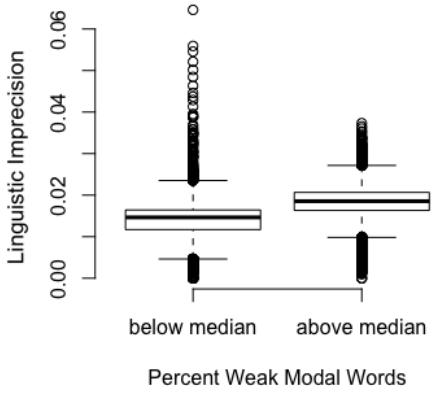
Scenario	**Realization of the event**	**Direction of the event**	**Signal strength**	**Disclosure choice by management**	**Language**
A	Current	Positive/Negative	Strong	Mandatory: Yes	Non-speculative
B	Future	Positive	Strong	Voluntary: Yes Verrecchia (1983) and Dye (1985) Kothari, Shu, Wysocki (2008)	Non-speculative
C	Future	Positive	Weak	Voluntary: Yes Verrecchia (1983) and Dye (1985) Kothari, Shu, Wysocki (2008)	Speculative
D	Future	Negative	Strong	Voluntary: Yes To avoid legal costs as in Skinner (1994)	Non-speculative
E	Future	Negative	Weak	Voluntary: No Verrecchia (1983) and Dye (1985) Kothari, Shu, Wysocki (2008) Bao, Kim, Mian, Su (2019)	Speculative

Figure 2: Speculation versus Uncertainty and Modality

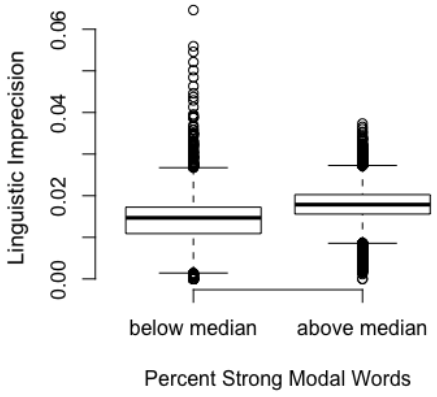
This figure shows the relation between each of notable textual tonal measures, (a) uncertainty words, (b) weak modal words, and (c) strong modal words (from the master dictionary by Loughran and McDonald (2011)), and our speculation measure. Each panel presents two side-by-side box plots for the distribution of the speculation measure by above and below the median of each textual tonal 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



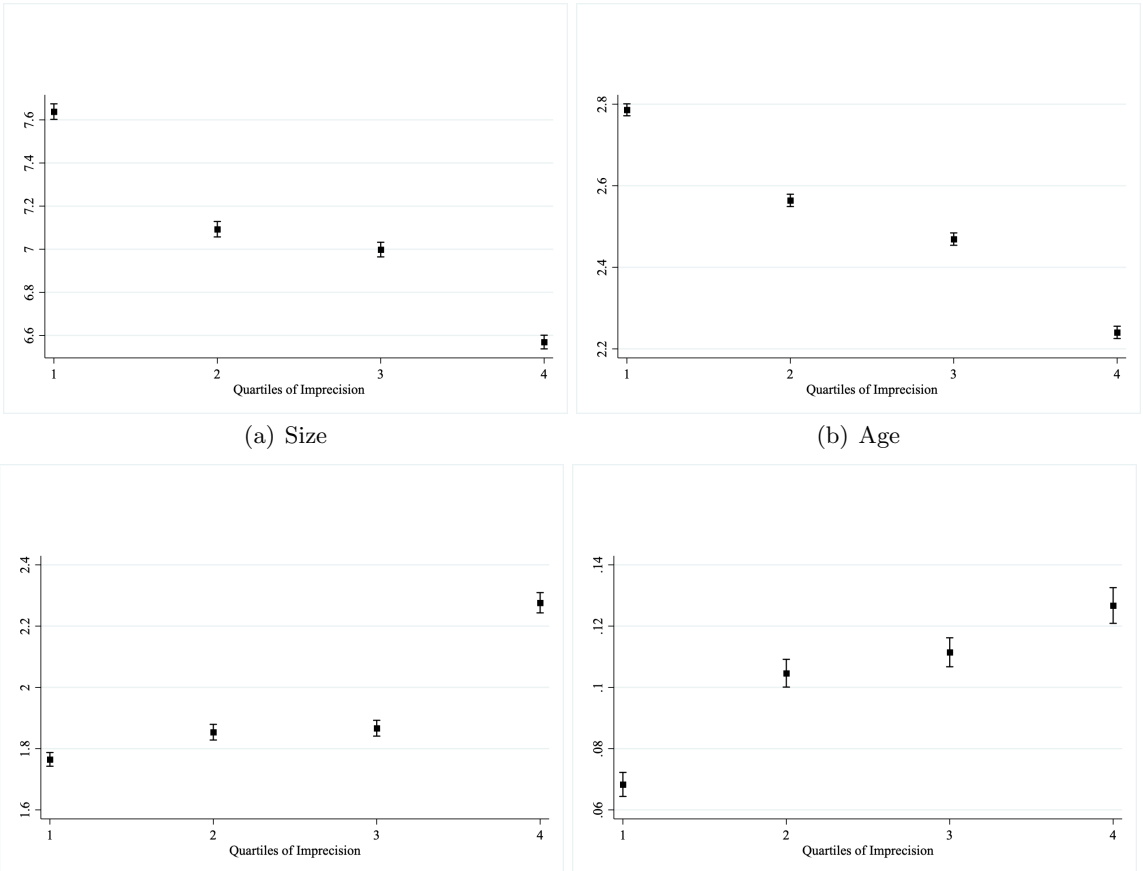
(b) Percentage of Weak Modal Words



(c) Percentage of Strong Modal Words

Figure 3: Speculation and Non-tonal Firm Characteristics

This figure shows the relation between each of notable non-tonal firm characteristics and the propensity to use speculative language in its 10-K disclosure. Each panel presents the 95% confidence interval for the mean of each firm characteristics for the first, second, third, and fourth quartile of the distribution of the 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 coefficient estimates of our speculation measure (*Speculation*) in the regressions of cumulative buy and hold abnormal returns (BHARs) over various multiple-week windows. 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 nth week ($n = 1, \ldots, 15$). For the pre-filing weeks and the week 0, each "reverse" cumulative BHAR is computed over the period from the end of the week 0 to the start of the nth week ($n = -5, \ldots, 0$). The vertical solid line at the week 0 indicates the 10-K filing-day window 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.

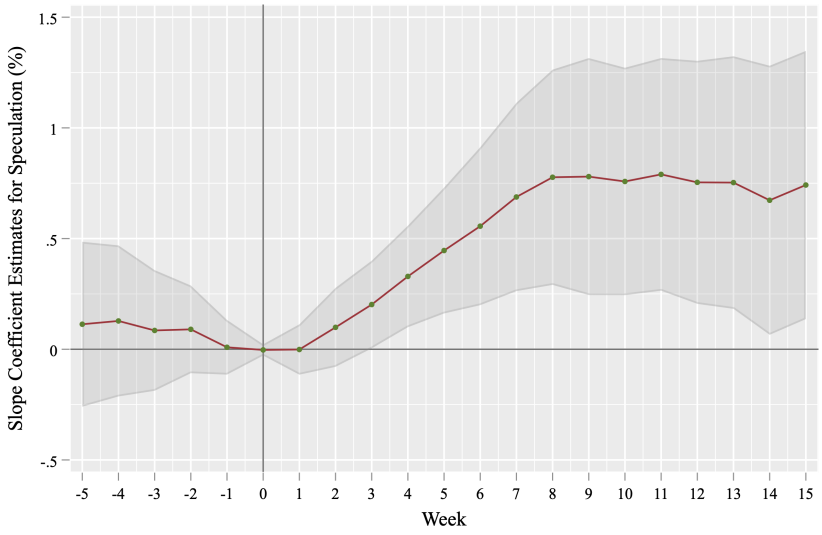


Figure 5: Speculation and Illiquidity

This figure presents coefficient estimates of our speculation measure (*Speculation*) in the regressions of the quoted relative bid-ask spread over an extended window combining the previous 5-week period before the 10-K release date and the subsequent 15-week period after the 10-K release date. The shaded area represents the 95% confidence interval for each slope coefficient estimate (circle marker). The vertical solid line at the week 0 indicates the 10-K filing-day window 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.

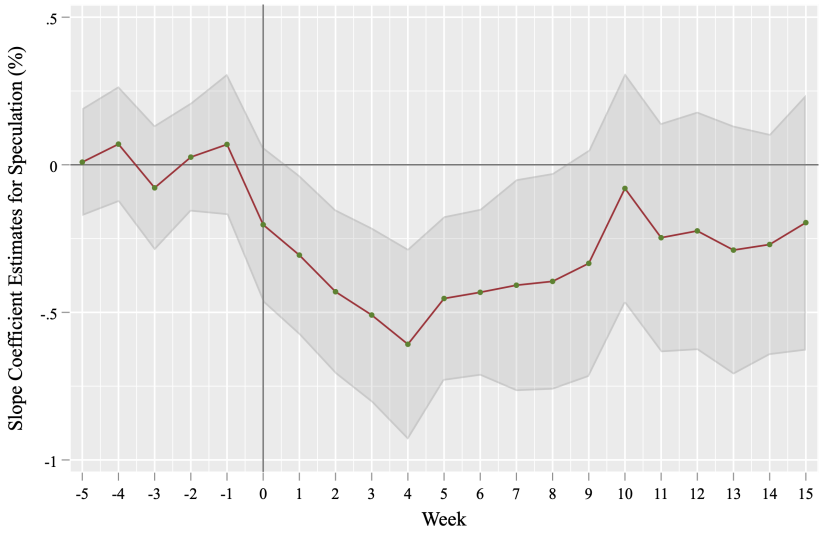
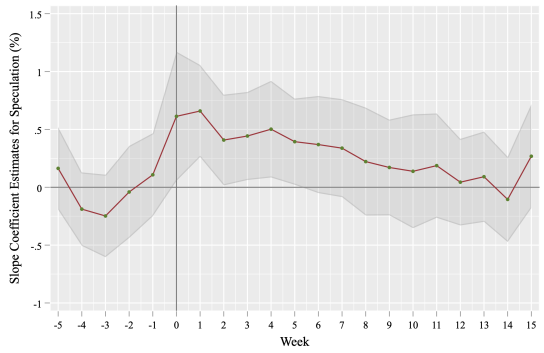
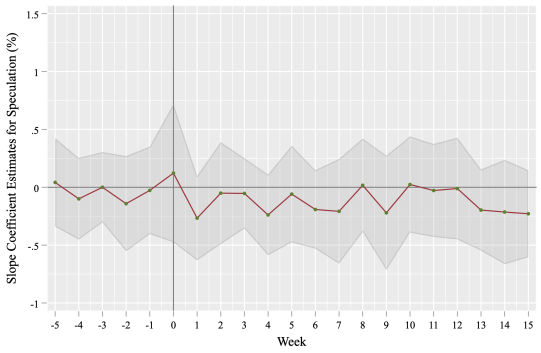


Figure 6: Speculation and Informed Trading Activities

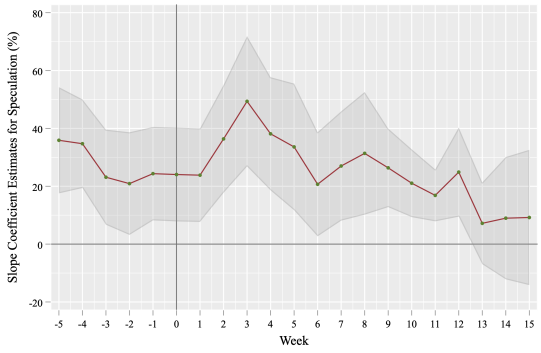
This figure presents coefficient estimates of our speculation measure (*Speculation*) in the regressions of the following weekly variables: Probability of informed buying, Probability of informed selling, Dollar volume of insider buying, and Dollar volume of insider selling over an extended window combining the previous 5-week period before the 10-K release date and the subsequent 15-week period after the 10-K release date. The shaded area represents the 95% confidence interval for each slope coefficient estimate (circle marker). The vertical solid line at the week 0 indicates the 10-K filing-day window 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.



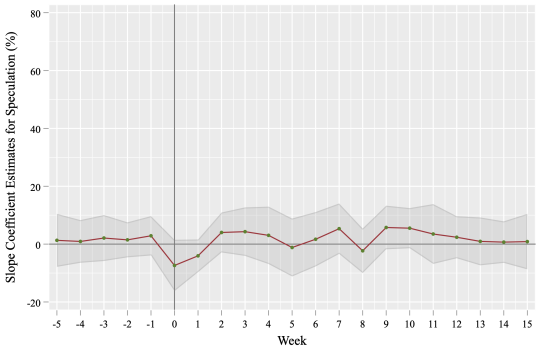
(a) Probability of informed buying



(b) Probability of informed selling



(c) Dollar volume of insider buying



(d) Dollar volume of insider selling

Figure 7: Speculation and News Sentiment

This figure presents coefficient estimates of our speculation measure (*Speculation*) in the regressions of the news sentiment score over an extended window combining the previous 5-week period before the 10-K release date and the subsequent 15-week period after the 10-K release date. The shaded area represents the 95% confidence interval for each slope coefficient estimate (circle marker). The vertical solid line at the week 0 indicates the 10-K filing-day window 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.

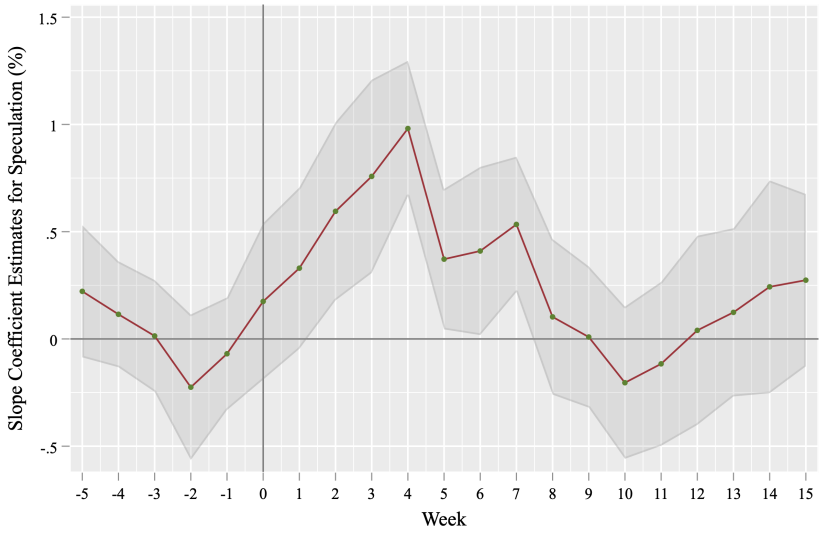
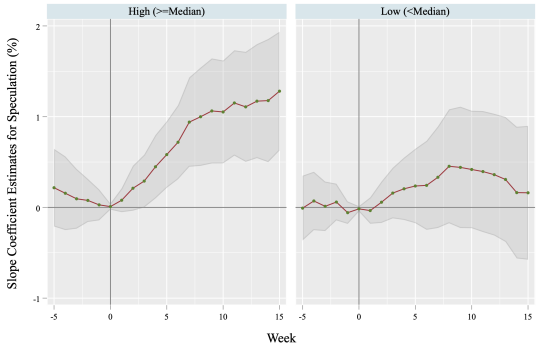
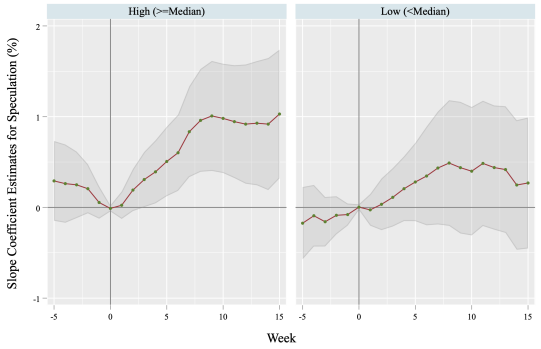


Figure 8: Speculation and Cumulative BHARs by Heterogeneous Variables

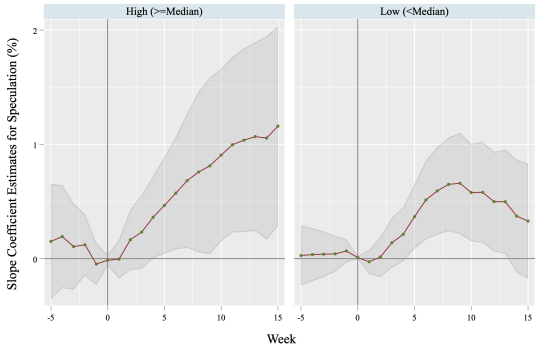
This figure presents coefficient estimates of our speculation measure (*Speculation*) in the regressions of cumulative buy and hold abnormal returns (BHARs) over various multiple-week windows by heterogeneous firm-specific variables. The shaded area represents the 95% confidence interval for each coefficient 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 nth week ($n = 1, \ldots, 15$). For the pre-filing weeks and the week 0, each "reverse" cumulative BHAR is computed over the period from the end of the week 0 to the start of the nth week ($n = -5, \ldots, 0$). The vertical solid line at the week 0 indicates the 10-K filing-day window that covers the four days between the 10-K filing day and three days later. We consider the following five firm-specific variables: forward-looking disclosure, R&D disclosure, pre-filing idiosyncratic volatility of returns, analyst coverage, and quoted relative bid-ask spread, which are constructed based on the information available as of 10-K filing dates. High and Low of each of these variables refer to firms with the variable above and below its median, respectively, which is determined based on its distribution over the one-year period before 10-K filing. For pre-filing analyst coverage, Presence and Absence are firms with at least one analyst following and those with no analyst following, respectively, which are determined based on the most recent quarter before 10-K filing. The coefficient estimates and confidence intervals are presented in percentage.



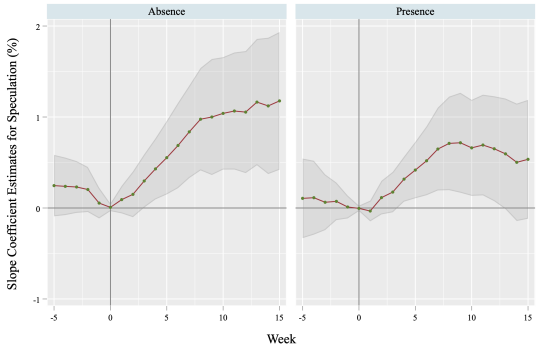
(a) Forward-looking disclosure



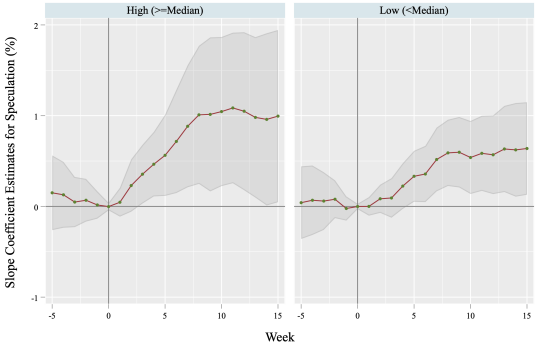
(b) R&D disclosure



(c) Pre-filing idiosyncratic volatility of return



(d) Pre-filing analyst coverage



(e) Pre-filing quoted relative bid-ask spread

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 neighbor 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 A.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	clear	could	could	must
3	may	approximately	possible	best
4	could	risk	might	highest
5	would	intangible	depend	never
6	number of	believe	uncertain	lowest
7	can	assumptions	depending	always
8	well	risks	depends	clearly
9	however	believes	appears	strongly
10	various	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 2015. Panel A presents the summary statistics for our speculation measure (*Speculation*), existing textual tonal variables based on the master dictionary by Loughran and McDonald (2011), and Fog words initially proposed by Robert Gunning in 1952 and used in the literature (e.g., Li (2008)), where all statistics are expressed in percentage. Panel B presents the summary statistics for non-tonal firm characteristics. 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.

Panel A: Textual Tonal Variables

	Mean	Std.Dev	Min	Median	Max	Num. of Obs.
Speculation	1.387	0.458	0.000	1.471	4.901	46996
Sentiment	-0.716	0.443	-4.362	-0.671	1.670	46996
Uncertain	1.011	0.362	0.000	1.042	3.230	46996
Modal	0.790	0.369	0.000	0.832	2.607	46996
Constraining	0.571	0.245	0.000	0.590	2.116	46996
Litigious	1.207	0.874	0.039	0.966	6.819	46996
Superfluous	0.009	0.011	0.000	0.006	0.253	46996
Interesting	0.123	0.078	0.000	0.115	1.666	46996
Fog	30.416	4.504	14.066	30.145	53.947	46996

Panel B: Non-tonal Firm Characteristics

	Mean	Std.Dev	Min	Median	Max	Num. of Obs.
Size	7.075	1.942	0.515	6.981	14.776	46996
Age	2.515	0.852	0.000	2.565	3.970	46996
Tobin's Q	1.940	1.506	0.703	1.407	10.485	46996
Sales Growth	0.103	0.266	-0.778	0.079	1.223	46996
Product market fluidity	6.994	3.500	1.449	6.379	17.336	46248
Financial constraints	-0.017	0.089	-0.192	-0.022	0.231	30630
Turnover	-1.871	1.141	-10.372	-1.752	3.225	46984
Book-to-market	-0.715	0.846	-8.864	-0.635	3.631	45771
Institution ownership	59.419	28.734	1.340	64.360	100.000	39370
Fama-French alpha	0.001	0.002	-0.004	0.000	0.010	46996
Filing-day abnormal return	0.000	0.050	-0.168	-0.001	0.185	46996

Table 3: Relations between Speculation and Other Firm-specific Variables

This table presents results from the regressions of our speculation measure (*Speculation*) on various textual tonal measures (Panel A) and non-tonal firm characteristics (Panel B). In Panel A, 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 based on the same 10-Ks. Panel B considers 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. Standard errors (SEs) reported in parentheses are clustered by firm and year. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% level, respectively.

Panel A: Textual Tonal Variables

	Dependent variable = Speculation		
	(1)	(2)	(3)
Fog(Z)	0.065***	0.066***	0.066***
	(0.008)	(0.009)	(0.009)
Uncertain(Z)	0.102***	0.126***	0.125***
	(0.013)	(0.018)	(0.018)
Modal(Z)	0.298***	0.250***	0.244***
	(0.023)	(0.018)	(0.017)
Sentiment(Z)		0.005*	0.006**
		(0.003)	(0.003)
Constraining(Z)		0.052***	0.052***
		(0.005)	(0.005)
Litigious(Z)		0.067***	0.069***
		(0.006)	(0.006)
Superfluous(Z)			0.009***
			(0.002)
Interesting(Z)			0.014***
			(0.003)
Fixed effect		Firm / Year	
Clustered SE		Firm / Year	
Observations	45672	45672	45672
Adjusted R^2	0.871	0.889	0.889

Panel B: Non-tonal Firm Characteristics

	Dependent variable = Speculation		
	(1)	(2)	(3)
Size(Z)	-0.035	-0.054**	-0.058*
	(0.022)	(0.023)	(0.029)
Age(Z)	-0.066***	-0.063***	-0.072***
	(0.018)	(0.015)	(0.018)
Tobin's Q(Z)	0.013**	0.018***	0.015**
	(0.006)	(0.005)	(0.006)
Sales Growth(Z)	0.000	-0.000	-0.001
	(0.003)	(0.004)	(0.006)
Product market fluidity(Z)		0.023***	0.022**
		(0.007)	(0.009)
Financial constraints(Z)		0.019***	0.020***
		(0.003)	(0.004)
Turnover(Z)			-0.000
			(0.005)
Institution ownership(Z)			0.004
			(0.010)
Fama-French alpha(Z)			0.000
			(0.004)
Filing-day abnormal return(Z)			0.001
			(0.001)
Fixed effect		Firm / Year	
Clustered SE		Firm / Year	
Observations	45672	29229	24128
Adjusted R^2	0.653	0.658	0.660

Table 4: Speculative Language in Disclosure and BHARs

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

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

where $BHAR_{itn}$ is the return difference between stock i and the CRSP value-weighted index over the nth-week window (the 1st week window starts from the fourth day after the 10-K filing day, and the 0th week window covers 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), and $\mathbf{X}_{it}$ is a column vector of our baseline control variables used in prior studies: Sentiment, Market value, Book-to-market, Turnover, Institutional ownership, Fama-French alpha, and Filing-day abnormal return. $RMkt_{tn}$ is the nth-week market return to capture the common economic shocks that affect all individual stocks' abnormal returns over the nth-week around each 10-K filing. The detailed definitions of these explanatory variables are given in the Appendix B. In Panel A, we estimate equation (1) for each week separately over the previous 3-week period before 10-K filing and the subsequent 9-week period after 10-K filing ($n = -3, \ldots, 9$). In Panel B, we repeat to estimate equation (1) with cumulative BHARs over multiple-week windows as the dependent variable. 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 nth week, i.e., Week[1,n] where $n = 1, \ldots, 9$. For the pre-filing weeks and the week 0, each "reverse" cumulative BHAR is computed over the period from the end of the week 0 to the start of the nth week, i.e., Week[n,0] where $n = -3, \ldots, 0$. The nth-week cumulative market return is calculated in the same way based on weekly market returns up to the nth week. (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.

Panel A: Weekly BHARs	Dependent variable = BHARs												
	Week(-3)	Week(-2)	Week(-1)	Week(0)	Week(1)	Week(2)	Week(3)	Week(4)	Week(5)	Week(6)	Week(7)	Week(8)	Week(9)
Speculation(Z)	0.004	-0.084	-0.001	-0.000	-0.001	0.102*	0.128**	0.143***	0.126**	0.115**	0.136**	0.048	0.007
	(0.048)	(0.061)	(0.063)	(0.000)	(0.058)	(0.053)	(0.054)	(0.047)	(0.059)	(0.054)	(0.062)	(0.058)	(0.055)
Sentiment(Z)	0.009	0.029	-0.069	-0.000	-0.009	0.030	-0.002	0.048	-0.070	0.007	0.057	-0.049	-0.045
	(0.048)	(0.041)	(0.058)	(0.000)	(0.048)	(0.044)	(0.044)	(0.036)	(0.050)	(0.055)	(0.043)	(0.039)	(0.036)
nth-week market return(Z)	-0.045	-0.197	0.047	0.000	0.173	-0.141	-0.121	-0.015	-0.013	0.021	0.034	0.084	0.018
	(0.099)	(0.207)	(0.140)	(0.000)	(0.210)	(0.186)	(0.162)	(0.153)	(0.098)	(0.084)	(0.082)	(0.082)	(0.160)
Clustered SE	Firm / Filing year-month												
Observations	42063	42173	42262	42298	41885	41711	41567	41505	41398	41248	41123	41010	40953
Adjusted R^2	0.022	0.013	0.021	1.000	0.003	0.002	0.004	0.002	0.002	0.001	0.001	0.002	0.001

Panel B: Cumulative BHARs	Dependent variable = Cumulative BHARs												
	Week[-3,0]	Week[-2,0]	Week[-1,0]	Week[0,0]	Week[1,1]	Week[1,2]	Week[1,3]	Week[1,4]	Week[1,5]	Week[1,6]	Week[1,7]	Week[1,8]	Week[1,9]
Speculation(Z)	0.085	0.090	0.009	-0.003	-0.001	0.099	0.202**	0.329***	0.446***	0.556***	0.688***	0.777***	0.780***
	(0.139)	(0.101)	(0.063)	(0.013)	(0.058)	(0.091)	(0.101)	(0.117)	(0.145)	(0.182)	(0.217)	(0.248)	(0.273)
Sentiment(Z)	-0.068	-0.038	0.016	-0.030**	-0.009	0.037	0.058	0.108	0.047	0.057	0.110	0.093	0.026
	(0.100)	(0.069)	(0.053)	(0.013)	(0.048)	(0.064)	(0.076)	(0.080)	(0.100)	(0.127)	(0.156)	(0.178)	(0.190)
nth-week cumulative	0.563*	0.366	0.153	0.033	0.173	0.278	0.395	0.474	0.682	0.997**	1.063**	1.246**	1.602***
market return(Z)	(0.321)	(0.257)	(0.114)	(0.020)	(0.210)	(0.269)	(0.306)	(0.356)	(0.416)	(0.445)	(0.468)	(0.534)	(0.525)
Clustered SE	Firm / Filing year-month												
Observations	42010	42156	42262	42298	41885	41681	41498	41376	41116	40909	40723	40570	40397
Adjusted R^2	0.225	0.280	0.395	0.923	0.003	0.003	0.005	0.005	0.007	0.010	0.009	0.010	0.012

Table 5: Speculative Language in Disclosure and Illiquidity

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

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

where $Spread_{itn}$ is the average quoted relative bid-ask spread (in logarithm) over the nth-week window (the 1st week window starts from the fourth day after the 10-K filing day, and the 0th week window covers the four days between the 10-K filing day and three days later), $Speculation_{it}$ is the percentage of speculation cues (out of total words), 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. We estimate equation (2) for each week separately over the previous 3-week period before 10-K filing and the subsequent 9-week period after 10-K filing ($n = -3, \ldots, 9$). (Z) indicates that the variable is standardized to have the mean of zero and the standard deviation of one. Following earlier studies, e.g., Amihud and Mendelson (1986) and Ben-Rephael, Kadan, and Wohl (2015), we also control for 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. Standard errors (SEs) reported in parentheses are clustered by firm and filing year-month to account for the serial and cross-sectional correlations in illiquidity, respectively, which are based on the evidence in Amihud (2002) and Bali, Peng, Shen, and Tang (2014), and Chordia, Roll, and Subrahmanyam (2000). 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												
	Week(-3)	Week(-2)	Week(-1)	Week(0)	Week(1)	Week(2)	Week(3)	Week(4)	Week(5)	Week(6)	Week(7)	Week(8)	Week(9)
Speculation(Z)	-0.078	0.026	0.069	-0.203	-0.306**	-0.430***	-0.509***	-0.608***	-0.453***	-0.432***	-0.408**	-0.395**	-0.334*
	(0.108)	(0.094)	(0.122)	(0.134)	(0.138)	(0.142)	(0.151)	(0.165)	(0.142)	(0.144)	(0.183)	(0.187)	(0.196)
Sentiment(Z)	-0.012	-0.076	0.037	-0.005	0.105	0.090	0.104	-0.179	-0.009	-0.135	-0.320*	-0.397**	-0.175
	(0.104)	(0.123)	(0.110)	(0.175)	(0.134)	(0.136)	(0.155)	(0.171)	(0.159)	(0.186)	(0.181)	(0.179)	(0.189)
Pre-filing spread(Z)	49.655***	48.158***	46.462***	41.002***	39.534***	37.737***	36.937***	36.188***	35.278***	34.365***	32.573***	31.836***	31.554***
	(0.337)	(0.410)	(0.501)	(0.742)	(0.764)	(0.935)	(1.026)	(1.115)	(1.241)	(1.269)	(1.421)	(1.508)	(1.585)
Fixed effect	Firm / Filing year-month												
Clustered SE	Firm / Filing year-month												
Observations	40541	40645	40734	40771	39335	39249	39192	39185	39177	39113	39096	39062	39085
Adjusted R^2	0.939	0.937	0.933	0.87	0.896	0.887	0.883	0.88	0.876	0.871	0.865	0.86	0.853

Table 6: 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,

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

where $\textit{Probability of informed buying}_{itn}$ proxies the trading activity of informed buyers based on Brennan, Huh, and Subrahmanyam (2018) over the nth-week window (the 1st week window starts from the fourth day after the 10-K filing day, and the 0th week window covers the four days between the 10-K filing day and three days later), $\textit{Speculation}_{it}$ is the percentage of speculation cues (out of total words), 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. We estimate equation (3) for each week separately over the previous 3-week period before 10-K filing and the subsequent 9-week period after 10-K filing ($n = -3, \ldots, 9$). (Z) indicates that the variable is standardized to have the mean of zero and the standard deviation of one. We also control for filing year-month fixed effect. 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												
	Week(-3)	Week(-2)	Week(-1)	Week(0)	Week(1)	Week(2)	Week(3)	Week(4)	Week(5)	Week(6)	Week(7)	Week(8)	Week(9)
Speculation(Z)	-0.248	-0.041	0.108	0.613**	0.659***	0.408**	0.443**	0.502**	0.394**	0.369*	0.338	0.222	0.171
	(0.182)	(0.203)	(0.183)	(0.285)	(0.203)	(0.200)	(0.194)	(0.213)	(0.190)	(0.214)	(0.216)	(0.238)	(0.211)
Sentiment(Z)	-0.051	-0.274***	-0.147	-0.425**	-0.309**	-0.059	-0.057	-0.117	-0.081	0.287**	0.123	-0.157	0.139
	(0.096)	(0.102)	(0.125)	(0.199)	(0.149)	(0.144)	(0.140)	(0.171)	(0.165)	(0.134)	(0.149)	(0.181)	(0.146)
Pre-filing informed buying(Z)	16.496***	15.763***	14.564***	8.613***	6.509***	5.351***	4.334***	3.419***	2.275***	1.607***	1.206***	1.008***	0.650***
	(0.152)	(0.188)	(0.188)	(0.215)	(0.182)	(0.294)	(0.209)	(0.176)	(0.180)	(0.184)	(0.174)	(0.161)	(0.189)
Fixed effect	Filing year-month												
Clustered SE	Firm / Filing year-month												
Observations	33218	33339	33334	32766	31747	31553	31355	31320	31390	31171	31151	30956	31135
Adjusted R^2	0.430	0.396	0.345	0.122	0.087	0.065	0.053	0.055	0.050	0.051	0.054	0.048	0.040

Table 7: Speculative Language in Disclosure and Insider Buying Activity

This table presents the estimation results of the following equation: For stock i in year t,

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

where $\textit{Dollar volume of insider buying}_{itn}$ captures the insiders' buying activity in the dollar volume over the nth-week window (the 1st week window starts from the fourth day after the 10-K filing day, and the 0th week window covers the four days between the 10-K filing day and three days later), $\textit{Speculation}_{it}$ is the percentage of speculation cues (out of total words), 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. We estimate the regression model above for each week separately over the previous 3-week period before 10-K filing and the subsequent 9-week period after 10-K filing ($n = -3, \ldots, 9$). (Z) indicates that the variable is standardized to have the mean of zero and the standard deviation of one. We also control for filing year-month fixed effect. 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												
	Week(-3)	Week(-2)	Week(-1)	Week(0)	Week(1)	Week(2)	Week(3)	Week(4)	Week(5)	Week(6)	Week(7)	Week(8)	Week(9)
Speculation(Z)	23.132***	20.916**	24.360***	24.057***	23.836***	36.369***	49.362***	38.129***	33.601***	20.689**	26.999***	31.381***	26.373***
	(8.383)	(9.050)	(8.240)	(8.270)	(8.248)	(9.516)	(11.488)	(9.980)	(11.170)	(9.184)	(9.639)	(10.816)	(6.931)
Sentiment(Z)	-8.951	-8.446	-0.209	-0.880	-2.867	-7.230	4.575	10.509	-3.294	-5.164	-5.823	-0.732	-2.157
	(6.616)	(7.044)	(7.614)	(8.235)	(7.334)	(8.037)	(9.803)	(7.652)	(8.220)	(6.731)	(6.813)	(6.779)	(5.678)
Fixed effect	Filing year-month												
Clustered SE	Firm / Filing year-month												
Observations	1939	1824	1774	1448	1896	1421	1056	1085	1258	1444	1769	2054	2247
Adjusted R^2	0.141	0.167	0.134	0.124	0.125	0.102	0.146	0.135	0.057	0.079	0.146	0.111	0.152

Table 8: Speculative Language in Disclosure and News Sentiment

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

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

where $News\ sentiment_{itn}$ is the average news sentiment score over the nth-week window (the 1st week window starts from the fourth day after the 10-K filing day, and the 0th week window covers the four days between the 10-K filing day and three days later). The average news sentiment score is calculated based on the ESS variable from the RavenPack News Analytics database. $Speculation_{it}$ is the percentage of speculation cues (out of total words) 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 nth-week market news sentiment to proxy for the common economic shocks that potentially affect all individual firms' news sentiment scores over the nth-week around each 10-K filing. The detailed definitions of these explanatory variables are given in the Appendix B. We estimate equation (4) for each week separately over the previous 3-week period before 10-K filing and the subsequent 9-week period after 10-K filing ($n = -3, \ldots, 9$). (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 in *News sentiment*, respectively. 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												
	Week(-3)	Week(-2)	Week(-1)	Week(0)	Week(1)	Week(2)	Week(3)	Week(4)	Week(5)	Week(6)	Week(7)	Week(8)	Week(9)
Speculation(Z)	0.014	-0.225	-0.069	0.175	0.330*	0.595***	0.758***	0.981***	0.372**	0.410**	0.534***	0.103	0.009
	(0.133)	(0.173)	(0.134)	(0.186)	(0.192)	(0.212)	(0.230)	(0.161)	(0.167)	(0.200)	(0.161)	(0.185)	(0.168)
Sentiment(Z)	-0.018	-0.244	-0.069	0.003	-0.534**	-0.213	-0.173	0.258	-0.567***	-0.376	-0.341**	0.105	-0.248
	(0.128)	(0.153)	(0.152)	(0.222)	(0.210)	(0.266)	(0.193)	(0.172)	(0.204)	(0.234)	(0.166)	(0.162)	(0.239)
Pre-filing news sentiment(Z)	13.552***	13.777***	14.141***	2.636***	2.865***	2.840***	2.536***	2.628***	2.234***	2.318***	2.609***	2.421***	2.195***
	(0.185)	(0.172)	(0.232)	(0.269)	(0.203)	(0.247)	(0.268)	(0.261)	(0.242)	(0.205)	(0.263)	(0.203)	(0.203)
nth-week market news sentiment(Z)	0.507***	0.198	0.051	0.619**	-0.139	0.307	0.168	0.372**	-0.033	0.765***	0.391**	0.328	0.493**
	(0.133)	(0.122)	(0.168)	(0.269)	(0.184)	(0.287)	(0.279)	(0.164)	(0.188)	(0.216)	(0.186)	(0.236)	(0.201)
Clustered SE	Firm / Filing year-month												
Observations	21632	21499	22047	18274	19146	18568	19093	19208	20189	21266	21833	21886	21232
Adjusted R^2	0.310	0.324	0.349	0.027	0.024	0.022	0.021	0.021	0.019	0.017	0.018	0.014	0.012

Table 9: Speculative Language in Disclosure and Cumulative BHARs by Heterogeneous Variables

We consider the following five firm-specific variables for heterogeneity tests: forward-looking disclosure, R&D disclosure, pre-filing idiosyncratic volatility of returns, analyst coverage, and quoted relative bid-ask spread. High and Low of each of these variables refer to firms with the variable above versus below its median, respectively, which is determined based on its distribution over the one-year period before 10-K filing. For pre-filing analyst coverage, Presence and Absence are firms with at least one analyst following and those with no analyst following, respectively, which are determined based on the most recent quarter before 10-K filing. For each subsample based on these variables, this table presents the estimation results of equation (1) as follows: For stock i in year t,

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

where $Cumulative\ BHAR_{itn}$ is the return difference between stock i and the CRSP value-weighted index over multiple-week window (as in Panel B of Table 4). $Speculation_{it}$ is the percentage of speculation cues (out of the total words), and $\mathbf{X}_{it}$ is a column vector of our baseline control variables used in prior studies: Sentiment, Market value, Book-to-market, Turnover, Institutional ownership, Fama-French alpha, and Filing-day abnormal return. $RMkt_{tn}$ is the nth-week market return to capture the common economic shocks that affect all individual stocks' abnormal returns over the nth-week around each 10-K filing. The detailed definitions of these explanatory variables are given in the Appendix B. 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 nth week, i.e., Week[1,n] where $n = 1, \ldots, 9$. For the pre-filing weeks and the week 0, each "reverse" cumulative BHAR is computed over the period from the end of the week 0 to the start of the nth week, i.e., Week[n,0] where $n = -3, \ldots, 0$. The nth-week cumulative market return is calculated in the same way based on weekly market returns up to the nth week. (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 serial and cross-sectional correlations of cumulative BHARs, respectively. Estimated coefficients and SEs are reported in percentage. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% level, respectively. For brevity, only the slope coefficient of $Speculation_{it}$ is reported although all control variables are included in the estimation.

		Dependent variable = Cumulative BHARs											
Subsamples by		Week[-3,0]	Week[-2,0]	Week[-1,0]	Week[0,0]	Week[1,1]	Week[1,2]	Week[1,3]	Week[1,4]	Week[1,5]	Week[1,6]	Week[1,7]	Week[1,8]	Week[1,9]
Panel A: Forward-looking disclosure														
High	Speculation(Z)	0.095	0.077	0.028	0.009	0.079	0.211*	0.290**	0.448**	0.582***	0.718***	0.940***	0.999***	1.063***
		(0.168)	(0.121)	(0.087)	(0.017)	(0.067)	(0.126)	(0.148)	(0.178)	(0.187)	(0.208)	(0.251)	(0.276)	(0.295)
Low	Speculation(Z)	0.012	0.059	-0.058	-0.016	-0.035	0.057	0.158	0.205	0.236	0.243	0.332	0.453	0.441
		(0.139)	(0.103)	(0.063)	(0.015)	(0.074)	(0.116)	(0.142)	(0.175)	(0.209)	(0.250)	(0.285)	(0.320)	(0.341)
Panel B: R&D disclosure														
High	Speculation(Z)	0.249	0.206	0.054	-0.011	0.024	0.191	0.307**	0.392**	0.506***	0.602***	0.834***	0.959***	1.008***
		(0.187)	(0.138)	(0.091)	(0.017)	(0.077)	(0.118)	(0.155)	(0.177)	(0.195)	(0.214)	(0.255)	(0.289)	(0.309)
Low	Speculation(Z)	-0.158	-0.087	-0.079	0.003	-0.027	0.034	0.111	0.206	0.280	0.346	0.434	0.489	0.438
		(0.139)	(0.107)	(0.062)	(0.016)	(0.089)	(0.145)	(0.164)	(0.182)	(0.220)	(0.277)	(0.317)	(0.353)	(0.370)
Panel C: Pre-filing idiosyncratic volatility of returns														
High	Speculation(Z)	0.106	0.121	-0.047	-0.014	-0.004	0.166	0.233	0.362**	0.466**	0.573**	0.682**	0.757**	0.812**
		(0.193)	(0.137)	(0.094)	(0.024)	(0.087)	(0.135)	(0.165)	(0.183)	(0.216)	(0.250)	(0.300)	(0.357)	(0.394)
Low	Speculation(Z)	0.039	0.042	0.067	0.009	-0.027	0.015	0.139	0.213*	0.368**	0.513***	0.593***	0.650***	0.660***
		(0.102)	(0.080)	(0.051)	(0.007)	(0.053)	(0.090)	(0.112)	(0.120)	(0.143)	(0.176)	(0.196)	(0.209)	(0.226)
Panel D: Pre-filing analyst coverage														
Absence	Speculation(Z)	0.231	0.203	0.054	0.007	0.092	0.150	0.297**	0.430**	0.553***	0.687***	0.837***	0.976***	1.000***
		(0.145)	(0.126)	(0.086)	(0.021)	(0.077)	(0.128)	(0.149)	(0.171)	(0.206)	(0.239)	(0.259)	(0.287)	(0.325)
Presence	Speculation(Z)	0.063	0.073	0.011	-0.004	-0.033	0.114	0.175	0.317**	0.417***	0.519***	0.648***	0.710***	0.717**
		(0.156)	(0.105)	(0.064)	(0.015)	(0.058)	(0.094)	(0.113)	(0.126)	(0.158)	(0.193)	(0.232)	(0.262)	(0.280)
Panel E: Pre-filing quoted relative bid-ask spread														
Hgih	Speculation(Z)	0.048	0.068	0.016	-0.002	0.046	0.230	0.355**	0.465**	0.563**	0.716**	0.883**	1.009***	1.014**
		(0.141)	(0.120)	(0.077)	(0.021)	(0.081)	(0.147)	(0.165)	(0.181)	(0.228)	(0.290)	(0.343)	(0.388)	(0.433)
Low	Speculation(Z)	0.059	0.079	-0.023	-0.001	-0.000	0.084	0.093	0.224*	0.332**	0.357**	0.517***	0.590***	0.597***
		(0.162)	(0.106)	(0.068)	(0.015)	(0.052)	(0.079)	(0.111)	(0.128)	(0.143)	(0.158)	(0.179)	(0.186)	(0.198)

Table 10: Speculative Language in Disclosure and BHARs: Decomposition with Forward-looking Disclosure

This table presents the estimation results of the following regression specification: For stock i in year t,

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

where $BHAR_{itn}$ is the return difference between stock i and the CRSP value-weighted index over the nth-week window (the 1st week window starts from the fourth day after the 10-K filing day, and the 0th week window covers the four days between the 10-K filing day and three days later), $Speculation_{it}^{FW}$ is the percentage of speculation cues only in paragraphs containing forward-looking disclosure keywords from Muslu, Radhakrishnan, Subramanyam, and Lim (2015), $Speculation_{it}^{-FW}$ is the percentage of speculation cues in the other paragraphs, calculated as $Speculation_{it}$ minus $Speculation_{it}^{FW}$, and $\mathbf{X}_{it}$ is a vector of our baseline control variables used in prior studies: Sentiment, Market value, Book-to-market, Turnover, Institutional ownership, Fama-French alpha, and Filing-day abnormal return. $RMkt_{tn}$ is the nth-week market return to capture the common economic shocks that affect all individual stocks' abnormal returns over the nth-week around each 10-K filing. The detailed definitions of these explanatory variables are given in the Appendix B. In Panel A, we estimate the above model for each week separately over the previous 3-week period before 10-K filing and the subsequent 9-week period after 10-K filing. In Panel B, we repeat the estimation of the above model with cumulative BHARs over multiple-week windows as the dependent variable. 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 nth week. For the pre-filing weeks and the week 0, each "reverse" cumulative BHAR is computed over the period from the end of the week 0 to the start of the nth week and used as the dependent variable. The nth-week cumulative market return is calculated in the same way based on weekly market returns up to the nth week. (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. For brevity, only the estimated coefficients on $Speculation_{it}^{FW}$, $Speculation_{it}^{-FW}$, and two other important variables are reported although all control variables are included in the estimation.

Panel A: Weekly BHARs	Dependent variable = BHARs												
	Week(-3)	Week(-2)	Week(-1)	Week(0)	Week(1)	Week(2)	Week(3)	Week(4)	Week(5)	Week(6)	Week(7)	Week(8)	Week(9)
Speculation^{FW}(Z)	0.006	-0.106	0.001	-0.000	-0.003	0.097	0.097**	0.146**	0.120**	0.110*	0.121*	0.067	0.014
	(0.054)	(0.071)	(0.063)	(0.000)	(0.071)	(0.062)	(0.047)	(0.056)	(0.058)	(0.062)	(0.063)	(0.068)	(0.069)
Speculation^{-FW}(Z)	0.004	-0.045	-0.006	-0.000	-0.015	0.146***	0.121***	0.127***	0.126**	0.120**	0.115**	0.034	0.022
	(0.049)	(0.056)	(0.052)	(0.000)	(0.048)	(0.049)	(0.045)	(0.037)	(0.051)	(0.050)	(0.047)	(0.047)	(0.044)
Sentiment(Z)	0.011	0.012	-0.069	-0.000	-0.012	0.035	0.007	0.059	-0.071	0.012	0.065	-0.039	-0.040
	(0.048)	(0.039)	(0.057)	(0.000)	(0.045)	(0.043)	(0.043)	(0.036)	(0.051)	(0.055)	(0.042)	(0.037)	(0.036)
nth-week market return(Z)	-0.045	-0.199	0.047	0.000	0.173	-0.137	-0.122	-0.015	-0.013	0.021	0.033	0.085	0.017
	(0.100)	(0.205)	(0.140)	(0.000)	(0.210)	(0.185)	(0.161)	(0.153)	(0.097)	(0.084)	(0.082)	(0.082)	(0.159)
Clustered SE	Firm / Filing year-month												
Observations	42063	42173	42262	42298	41885	41711	41567	41505	41398	41248	41123	41010	40953
Adjusted R^2	0.022	0.014	0.021	1.000	0.003	0.002	0.005	0.002	0.002	0.001	0.001	0.002	0.001

Panel B: Cumulative BHARs	Dependent variable = Cumulative BHARs												
	Week[-3,0]	Week[-2,0]	Week[-1,0]	Week[0,0]	Week[1,1]	Week[1,2]	Week[1,3]	Week[1,4]	Week[1,5]	Week[1,6]	Week[1,7]	Week[1,8]	Week[1,9]
Speculation^{FW}(Z)	0.043	0.070	-0.017	-0.011	-0.003	0.082	0.155	0.280**	0.380**	0.478**	0.582**	0.683**	0.689**
	(0.125)	(0.098)	(0.061)	(0.015)	(0.071)	(0.105)	(0.118)	(0.127)	(0.154)	(0.189)	(0.229)	(0.268)	(0.298)
Speculation^{-FW}(Z)	0.151	0.131	0.025	0.004	-0.015	0.147*	0.244**	0.364***	0.493***	0.627***	0.759***	0.828***	0.845***
	(0.164)	(0.111)	(0.056)	(0.010)	(0.048)	(0.078)	(0.096)	(0.108)	(0.144)	(0.170)	(0.199)	(0.225)	(0.234)
Sentiment(Z)	-0.069	-0.033	0.008	-0.033**	-0.012	0.038	0.066	0.124	0.058	0.072	0.131	0.123	0.060
	(0.098)	(0.068)	(0.053)	(0.013)	(0.045)	(0.061)	(0.073)	(0.077)	(0.094)	(0.121)	(0.147)	(0.166)	(0.177)
nth-week cumulative	0.566*	0.367	0.155	0.033*	0.173	0.281	0.398	0.478	0.688*	1.005**	1.069**	1.257**	1.604***
market return(Z)	(0.319)	(0.255)	(0.114)	(0.020)	(0.210)	(0.266)	(0.303)	(0.353)	(0.412)	(0.439)	(0.462)	(0.525)	(0.521)
Clustered SE	Firm / Filing year-month												
Observations	42010	42156	42262	42298	41885	41681	41498	41376	41116	40909	40723	40570	40397
Adjusted R^2	0.226	0.280	0.395	0.923	0.003	0.003	0.005	0.006	0.008	0.010	0.010	0.011	0.013

Appendix Tables to:

Speculative and Informative: Lessons from Market Reactions to Speculative Disclosure

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

This table presents the top 100 meaningful (i.e., 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 pairs of neighbor words in 10-K paragraphs and compute the saliency score of each bigram in paragraphs with the speculation cues relative to paragraphs without such keywords.

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 A.2: Speculative Language in Disclosure and BHARs (All Slope Coefficients)

This table present the coefficient estimates for all variables in Table 4. Estimated coefficients and standard errors (SEs) are reported in percentage.

Panel A: Weekly BHARs	Dependent variable = BHARs												
	Week(-3)	Week(-2)	Week(-1)	Week(0)	Week(1)	Week(2)	Week(3)	Week(4)	Week(5)	Week(6)	Week(7)	Week(8)	Week(9)
Speculation(Z)	0.004	-0.084	-0.001	-0.000	-0.001	0.102*	0.128**	0.143***	0.126**	0.115**	0.136**	0.048	0.007
	(0.048)	(0.061)	(0.063)	(0.000)	(0.058)	(0.053)	(0.054)	(0.047)	(0.059)	(0.054)	(0.062)	(0.058)	(0.055)
Sentiment(Z)	0.009	0.029	-0.069	-0.000	-0.009	0.030	-0.002	0.048	-0.070	0.007	0.057	-0.049	-0.045
	(0.048)	(0.041)	(0.058)	(0.000)	(0.048)	(0.044)	(0.044)	(0.036)	(0.050)	(0.055)	(0.043)	(0.039)	(0.036)
Market value(Z)	0.206***	0.294***	0.246***	-0.000	-0.112	-0.148**	-0.202***	-0.237***	-0.165***	-0.166**	-0.057	-0.089	0.021
	(0.063)	(0.101)	(0.055)	(0.000)	(0.072)	(0.066)	(0.054)	(0.073)	(0.059)	(0.067)	(0.070)	(0.063)	(0.073)
Book-to-market(Z)	0.153***	0.105	0.047	0.000	0.118*	0.067	0.043	-0.014	0.004	0.022	-0.002	0.096**	0.028
	(0.058)	(0.071)	(0.048)	(0.000)	(0.067)	(0.081)	(0.065)	(0.064)	(0.075)	(0.070)	(0.048)	(0.048)	(0.063)
Turnover(Z)	0.048	-0.128	-0.251**	0.000	-0.074	-0.079	0.212**	-0.017	0.079	0.052	-0.039	-0.054	-0.014
	(0.102)	(0.146)	(0.103)	(0.000)	(0.137)	(0.225)	(0.086)	(0.086)	(0.088)	(0.109)	(0.101)	(0.112)	(0.108)
Institutional ownership(Z)	-0.017	0.179**	0.031	-0.000	0.212***	0.032	-0.011	0.083	-0.017	-0.017	0.158**	0.038	0.023
	(0.076)	(0.084)	(0.084)	(0.000)	(0.078)	(0.124)	(0.043)	(0.073)	(0.063)	(0.071)	(0.070)	(0.066)	(0.067)
Fama-French alpha(Z)	1.045***	0.739***	1.015***	0.000	-0.011	-0.173	0.241	0.053	0.057	0.021	-0.022	-0.237**	0.106
	(0.114)	(0.188)	(0.165)	(0.000)	(0.142)	(0.262)	(0.169)	(0.067)	(0.066)	(0.087)	(0.077)	(0.094)	(0.150)
Filing-day abnormal return(Z)	-0.066	0.128**	-0.216***	6.045***	-0.249***	0.024	-0.186***	0.097**	-0.090	-0.000	-0.008	0.035	-0.221***
	(0.066)	(0.062)	(0.079)	(0.000)	(0.090)	(0.104)	(0.069)	(0.043)	(0.059)	(0.058)	(0.055)	(0.082)	(0.068)
nth-week market return(Z)	-0.045	-0.197	0.047	0.000	0.173	-0.141	-0.121	-0.015	-0.013	0.021	0.034	0.084	0.018
	(0.099)	(0.207)	(0.140)	(0.000)	(0.210)	(0.186)	(0.162)	(0.153)	(0.098)	(0.084)	(0.082)	(0.082)	(0.160)
Clustered SE	Firm / Filing year-month												
Observations	42063	42173	42262	42298	41885	41711	41567	41505	41398	41248	41123	41010	40953
Adjusted R^2	0.022	0.013	0.021	1.000	0.003	0.002	0.004	0.002	0.002	0.001	0.001	0.002	0.001

Panel B: Cumulative BHARs	Dependent variable = Cumulative BHARs												
	Week[-3,0]	Week[-2,0]	Week[-1,0]	Week[0,0]	Week[1,1]	Week[1,2]	Week[1,3]	Week[1,4]	Week[1,5]	Week[1,6]	Week[1,7]	Week[1,8]	Week[1,9]
Speculation(Z)	0.085	0.090	0.009	-0.003	-0.001	0.099	0.202**	0.329***	0.446***	0.556***	0.688***	0.777***	0.780***
	(0.139)	(0.101)	(0.063)	(0.013)	(0.058)	(0.091)	(0.101)	(0.117)	(0.145)	(0.182)	(0.217)	(0.248)	(0.273)
Sentiment(Z)	-0.068	-0.038	0.016	-0.030**	-0.009	0.037	0.058	0.108	0.047	0.057	0.110	0.093	0.026
	(0.100)	(0.069)	(0.053)	(0.013)	(0.048)	(0.064)	(0.076)	(0.080)	(0.100)	(0.127)	(0.156)	(0.178)	(0.190)
Market value(Z)	-1.660***	-1.236***	-0.690***	-0.190***	-0.112	-0.260***	-0.443***	-0.662***	-0.838***	-0.985***	-1.058***	-1.172***	-1.160***
	(0.181)	(0.143)	(0.087)	(0.017)	(0.072)	(0.080)	(0.117)	(0.161)	(0.206)	(0.252)	(0.306)	(0.360)	(0.399)
Book-to-market(Z)	-0.483***	-0.302***	-0.155***	-0.049***	0.118*	0.169	0.228*	0.196	0.191	0.235	0.228	0.342	0.401
	(0.157)	(0.110)	(0.058)	(0.013)	(0.067)	(0.120)	(0.125)	(0.165)	(0.193)	(0.231)	(0.259)	(0.278)	(0.289)
Turnover(Z)	1.732***	1.330***	0.751***	0.173***	-0.074	-0.129	0.011	-0.015	0.023	0.155	0.155	0.172	0.136
	(0.409)	(0.303)	(0.133)	(0.019)	(0.137)	(0.296)	(0.344)	(0.411)	(0.422)	(0.492)	(0.552)	(0.601)	(0.646)
Institutional ownership(Z)	-0.698**	-0.537***	-0.168	-0.041**	0.212***	0.190	0.221	0.286	0.276	0.214	0.349	0.351	0.386
	(0.286)	(0.206)	(0.108)	(0.016)	(0.078)	(0.146)	(0.172)	(0.209)	(0.202)	(0.230)	(0.265)	(0.282)	(0.300)
Fama-French alpha(Z)	-2.406***	-1.578***	-0.901***	0.037	-0.011	-0.167	-0.048	0.014	0.074	0.073	0.023	-0.183	-0.233
	(0.465)	(0.369)	(0.182)	(0.022)	(0.142)	(0.354)	(0.334)	(0.347)	(0.391)	(0.452)	(0.500)	(0.550)	(0.525)
Filing-day abnormal return(Z)	-6.102***	-6.076***	-5.892***	-5.946***	-0.249***	-0.215*	-0.392***	-0.284**	-0.356**	-0.349**	-0.335*	-0.336	-0.620**
	(0.213)	(0.173)	(0.127)	(0.098)	(0.090)	(0.127)	(0.126)	(0.133)	(0.144)	(0.167)	(0.183)	(0.218)	(0.259)
nth-week cumulative	0.563*	0.366	0.153	0.033	0.173	0.278	0.395	0.474	0.682	0.997**	1.063**	1.246**	1.602***
market return(Z)	(0.321)	(0.257)	(0.114)	(0.020)	(0.210)	(0.269)	(0.306)	(0.356)	(0.416)	(0.445)	(0.468)	(0.534)	(0.525)
Clustered SE	Firm / Filing year-month												
Observations	42010	42156	42262	42298	41885	41681	41498	41376	41116	40909	40723	40570	40397
Adjusted R^2	0.225	0.280	0.395	0.923	0.003	0.003	0.005	0.005	0.007	0.010	0.009	0.010	0.012

Table A.3: Speculative Language in Disclosure and Return Reactions: Robustness Tests

This table presents results from the multiple robustness tests estimating equation (1) similar to Table 4. In Panel A, we use a refined subsample that excludes all $BHAR_{itn}$ observations having earnings announcements over the three to seven weeks after 10-K filing dates. In Panel B, we use *Speculation* adjusted for safe-harbor boilerplate paragraphs. Specifically, we subtract the number of speculation cues used in safe-harbor boilerplate paragraphs of a 10-K filing from the original total number of speculation cues in the 10-K filing. In Panel C, for each firm, we demean *Speculation* with the firm average of *Speculation*. In Panel D, we employ cumulative abnormal returns (CARs) as the dependent variable instead of BHARs for equation (1). For the post-filing weeks, each cumulative CAR is computed over the period from the start of the 1st week to the end of the nth week, i.e., Week[1,n] where $n = 1, \ldots, 9$. For the pre-filing weeks and the week 0, each "reverse" cumulative CAR is computed over the period from the end of the week 0 to the start of the nth week, i.e., Week[n,0] where $n = -3, \ldots, 0$. In Panel E, we additionally control for the exposures to the Fama-French three factors. Standard errors (SEs) reported in parentheses are clustered by firm and filing year-month to account for the serial and cross-sectional correlations of abnormal returns, respectively. Estimated coefficients and SEs are reported in percentage. For brevity, only the slope coefficient of $Speculation_{it}$ is reported although all control variables in equation (1) are included in the estimation.

Dependent Variable		Week(-3)	Week(-2)	Week(-1)	Week(0)	Week(1)	Week(2)	Week(3)	Week(4)	Week(5)	Week(6)	Week(7)	Week(8)	Week(9)
Panel A: Excluding Observations with Future Earnings Announcements														
Weekly BHARs	Speculation(Z)	-0.024	-0.067	-0.005	0.000	0.009	0.105*	0.119**	0.138***	0.122**	0.123**	0.115**	0.023	0.008
		(0.048)	(0.065)	(0.061)	(0.000)	(0.061)	(0.057)	(0.052)	(0.048)	(0.055)	(0.057)	(0.054)	(0.062)	(0.056)
Cumulative BHARs	Speculation(Z)	0.122	0.073	0.070	0.014	0.009	0.105	0.192*	0.367***	0.381**	0.524***	0.635***	0.678**	0.768**
		(0.153)	(0.101)	(0.063)	(0.011)	(0.061)	(0.093)	(0.101)	(0.124)	(0.148)	(0.187)	(0.221)	(0.267)	(0.305)
Panel B: Removing Safe-harbor Boilerplate Paragraphs														
Weekly BHARs	Speculation(Z)	0.001	-0.084	-0.011	-0.000	0.012	0.107*	0.131***	0.136***	0.124**	0.117**	0.123**	0.041	0.009
		(0.049)	(0.061)	(0.065)	(0.000)	(0.056)	(0.055)	(0.049)	(0.048)	(0.059)	(0.058)	(0.061)	(0.058)	(0.053)
Cumulative BHARs	Speculation(Z)	0.117	0.111	0.026	0.002	0.012	0.117	0.225**	0.345***	0.461***	0.572***	0.697***	0.780***	0.786***
		(0.145)	(0.103)	(0.066)	(0.013)	(0.056)	(0.090)	(0.105)	(0.116)	(0.144)	(0.178)	(0.212)	(0.242)	(0.267)
Panel C: Using Demeaned Speculation														
Weekly BHARs	Speculation(Z)	0.034	-0.088	0.048	-0.000	-0.006	0.121**	0.141***	0.148***	0.138**	0.121**	0.127**	0.078	0.029
		(0.043)	(0.063)	(0.057)	(0.000)	(0.072)	(0.053)	(0.051)	(0.043)	(0.062)	(0.060)	(0.057)	(0.054)	(0.052)
Cumulative BHARs	Speculation(Z)	-0.041	0.014	-0.046	-0.009	-0.006	0.107	0.219**	0.354**	0.467***	0.573***	0.708***	0.810***	0.817***
		(0.116)	(0.087)	(0.051)	(0.013)	(0.072)	(0.104)	(0.105)	(0.139)	(0.167)	(0.205)	(0.241)	(0.265)	(0.285)
Panel D: Using CARs														
Weekly CARs	Speculation(Z)	-0.000	-0.072	-0.002	-0.000	-0.001	0.107**	0.118**	0.142***	0.128**	0.122**	0.115*	0.048	0.008
		(0.047)	(0.060)	(0.061)	(0.000)	(0.057)	(0.054)	(0.047)	(0.047)	(0.059)	(0.058)	(0.060)	(0.059)	(0.054)
Cumulative CARs	Speculation(Z)	0.025	0.042	-0.006	0.000	-0.001	0.104	0.204**	0.345***	0.458***	0.569***	0.670***	0.738***	0.751***
		(0.107)	(0.084)	(0.061)	(0.000)	(0.057)	(0.087)	(0.102)	(0.118)	(0.146)	(0.181)	(0.214)	(0.239)	(0.262)
Panel E: Controlling for Fama-French Three Factors' Loadings														
Weekly BHARs	Speculation(Z)	0.007	-0.088	0.006	-0.000	-0.003	0.106**	0.125**	0.139***	0.126**	0.111**	0.136**	0.043	0.005
		(0.048)	(0.062)	(0.063)	(0.000)	(0.058)	(0.053)	(0.053)	(0.047)	(0.059)	(0.054)	(0.065)	(0.057)	(0.053)
Cumulative BHARs	Speculation(Z)	0.057	0.071	-0.010	-0.007	-0.003	0.101	0.203**	0.326***	0.442***	0.550***	0.682***	0.766***	0.767***
		(0.143)	(0.104)	(0.065)	(0.013)	(0.058)	(0.091)	(0.101)	(0.119)	(0.149)	(0.183)	(0.218)	(0.247)	(0.270)

Table A.4: Speculative Language in Disclosure and Illiquidity (All Slope Coefficients)

This table present coefficient estimates for all control variables in Table 5. Estimated coefficients and standard errors (SEs) are reported in percentage.

	Dependent variable = Quoted relative bid-ask spread												
	Week(-3)	Week(-2)	Week(-1)	Week(0)	Week(1)	Week(2)	Week(3)	Week(4)	Week(5)	Week(6)	Week(7)	Week(8)	Week(9)
Speculation(Z)	-0.078	0.026	0.069	-0.203	-0.306**	-0.430***	-0.509***	-0.608***	-0.453***	-0.432***	-0.408**	-0.395**	-0.334*
	(0.108)	(0.094)	(0.122)	(0.134)	(0.138)	(0.142)	(0.151)	(0.165)	(0.142)	(0.144)	(0.183)	(0.187)	(0.196)
Sentiment(Z)	-0.012	-0.076	0.037	-0.005	0.105	0.090	0.104	-0.179	-0.009	-0.135	-0.320*	-0.397**	-0.175
	(0.104)	(0.123)	(0.110)	(0.175)	(0.134)	(0.136)	(0.155)	(0.171)	(0.159)	(0.186)	(0.181)	(0.179)	(0.189)
Market value(Z)	0.573*	-0.663*	-1.494***	-3.963***	-4.242***	-4.720***	-4.682***	-5.145***	-4.525***	-5.443***	-5.855***	-6.302***	-6.311***
	(0.292)	(0.344)	(0.342)	(0.560)	(0.530)	(0.543)	(0.585)	(0.836)	(0.878)	(1.005)	(0.828)	(0.717)	(0.828)
Book-to-market(Z)	-0.240*	-0.227*	-0.045	0.052	0.391**	0.256	0.258	0.590**	0.410**	0.427**	0.496**	0.580**	0.385
	(0.124)	(0.132)	(0.153)	(0.205)	(0.191)	(0.180)	(0.207)	(0.240)	(0.195)	(0.196)	(0.206)	(0.262)	(0.245)
Turnover(Z)	-0.274	-0.719***	-1.194***	-1.851***	-1.684***	-1.660***	-1.768***	-1.880***	-1.847***	-2.098***	-2.506***	-2.056***	-2.079***
	(0.170)	(0.173)	(0.228)	(0.299)	(0.306)	(0.339)	(0.368)	(0.362)	(0.374)	(0.372)	(0.358)	(0.423)	(0.438)
Institutional ownership(Z)	0.051	0.377**	0.539**	-0.102	-0.213	-0.798**	-0.712**	-0.600**	-0.615**	-0.641*	-0.792**	-1.268***	-1.194***
	(0.190)	(0.187)	(0.217)	(0.301)	(0.285)	(0.334)	(0.322)	(0.301)	(0.285)	(0.336)	(0.361)	(0.358)	(0.371)
Fama-French alpha(Z)	-0.258*	-0.283**	-0.686***	-0.907***	-1.118***	-0.984***	-1.196***	-1.257***	-1.047***	-1.163***	-1.216***	-1.021***	-0.920***
	(0.134)	(0.133)	(0.125)	(0.176)	(0.207)	(0.198)	(0.203)	(0.218)	(0.225)	(0.219)	(0.220)	(0.236)	(0.246)
Filing-day abnormal return(Z)	0.010	-0.051	0.030	-0.623***	-0.957***	-1.034***	-0.898***	-0.859***	-0.823***	-0.781***	-0.787***	-0.855***	-0.874***
	(0.078)	(0.090)	(0.097)	(0.165)	(0.116)	(0.121)	(0.138)	(0.119)	(0.119)	(0.119)	(0.130)	(0.133)	(0.122)
Nasdaq dummy	1.664***	0.155	-1.446**	-1.928*	-2.056**	-1.920*	-1.549	-1.567*	-2.307**	-1.471	-2.473**	-2.448**	-2.418*
	(0.574)	(0.612)	(0.652)	(1.073)	(0.815)	(1.041)	(0.989)	(0.934)	(1.067)	(0.951)	(1.197)	(1.151)	(1.277)
Pre-filing spread(Z)	49.655***	48.158***	46.462***	41.002***	39.534***	37.737***	36.937***	36.188***	35.278***	34.365***	32.573***	31.836***	31.554***
	(0.337)	(0.410)	(0.501)	(0.742)	(0.764)	(0.935)	(1.026)	(1.115)	(1.241)	(1.269)	(1.421)	(1.508)	(1.585)
Fixed effect	Firm / Filing year-month												
Clustered SE	Firm / Filing year-month												
Observations	40541	40645	40734	40771	39335	39249	39192	39185	39177	39113	39096	39062	39085
Adjusted R^2	0.939	0.937	0.933	0.87	0.896	0.887	0.883	0.88	0.876	0.871	0.865	0.86	0.853

Table A.5: Speculative Language in Disclosure and Informed Buying Activity (All Slope Coefficients)

This table presents coefficient estimates for all control variables in Table 6. Estimated coefficients and standard errors (SEs) are reported in percentage.

	Dependent variable = Probability of informed buying												
	Week(-3)	Week(-2)	Week(-1)	Week(0)	Week(1)	Week(2)	Week(3)	Week(4)	Week(5)	Week(6)	Week(7)	Week(8)	Week(9)
Speculation(Z)	-0.248	-0.041	0.108	0.613**	0.659***	0.408**	0.443**	0.502**	0.394**	0.369*	0.338	0.222	0.171
	(0.182)	(0.203)	(0.183)	(0.285)	(0.203)	(0.200)	(0.194)	(0.213)	(0.190)	(0.214)	(0.216)	(0.238)	(0.211)
Sentiment(Z)	-0.051	-0.274***	-0.147	-0.425**	-0.309**	-0.059	-0.057	-0.117	-0.081	0.287**	0.123	-0.157	0.139
	(0.096)	(0.102)	(0.125)	(0.199)	(0.149)	(0.144)	(0.140)	(0.171)	(0.165)	(0.134)	(0.149)	(0.181)	(0.146)
Market value(Z)	-0.164	-0.233	0.074	-0.217	0.322	0.504	0.489	0.241	0.070	0.551	0.395	0.505	0.652*
	(0.223)	(0.226)	(0.273)	(0.446)	(0.348)	(0.316)	(0.370)	(0.357)	(0.389)	(0.399)	(0.373)	(0.334)	(0.374)
Book-to-market(Z)	-0.095	-0.164	-0.042	-0.189	-0.254	-0.267	-0.023	-0.293	-0.206	-0.173	-0.462**	-0.104	-0.239
	(0.128)	(0.160)	(0.155)	(0.203)	(0.176)	(0.197)	(0.183)	(0.235)	(0.204)	(0.204)	(0.215)	(0.197)	(0.170)
Turnover(Z)	0.240*	0.091	-0.713***	-0.185	-1.143***	-1.558***	-1.125***	-1.637***	-1.542***	-1.654***	-1.519***	-1.472***	-1.288***
	(0.132)	(0.214)	(0.208)	(0.294)	(0.212)	(0.335)	(0.234)	(0.240)	(0.215)	(0.268)	(0.262)	(0.207)	(0.221)
Institutional ownership(Z)	0.019	0.378**	0.548***	1.024***	1.314***	1.575***	1.557***	1.539***	1.480***	1.713***	2.060***	2.143***	2.060***
	(0.162)	(0.178)	(0.189)	(0.306)	(0.237)	(0.324)	(0.228)	(0.226)	(0.279)	(0.266)	(0.302)	(0.279)	(0.247)
Fama-French alpha(Z)	0.205	0.050	0.610***	0.445*	0.718***	0.753***	0.646***	0.885***	0.633***	0.636***	0.401**	0.309	0.391**
	(0.151)	(0.194)	(0.198)	(0.226)	(0.193)	(0.213)	(0.227)	(0.228)	(0.214)	(0.169)	(0.196)	(0.222)	(0.160)
Filing-day abnormal return(Z)	-0.144	0.259*	0.050	5.003***	1.182***	0.580***	0.815***	0.804***	0.704***	0.493***	0.434***	0.379**	0.305*
	(0.099)	(0.148)	(0.125)	(0.404)	(0.231)	(0.189)	(0.166)	(0.182)	(0.164)	(0.141)	(0.153)	(0.157)	(0.163)
Nasdaq dummy	-0.296	-0.033	0.069	0.532	0.665	1.698**	1.767**	1.303*	1.682**	1.764**	1.766***	1.104*	1.564**
	(0.325)	(0.319)	(0.383)	(0.652)	(0.616)	(0.702)	(0.682)	(0.682)	(0.671)	(0.754)	(0.640)	(0.597)	(0.643)
Pre-filing informed buying(Z)	16.496***	15.763***	14.564***	8.613***	6.509***	5.351***	4.334***	3.419***	2.275***	1.607***	1.206***	1.008***	0.650***
	(0.152)	(0.188)	(0.188)	(0.215)	(0.182)	(0.294)	(0.209)	(0.176)	(0.180)	(0.184)	(0.174)	(0.161)	(0.189)
Fixed effect	Filing year-month												
Clustered SE	Firm / Filing year-month												
Observations	33218	33339	33334	32766	31747	31553	31355	31320	31390	31171	31151	30956	31135
Adjusted R^2	0.430	0.396	0.345	0.122	0.087	0.065	0.053	0.055	0.050	0.051	0.054	0.048	0.040

Table A.6: Speculative Language in Disclosure and Insider Buying Activity (All Slope Coefficients)

This table presents coefficient estimates for all control variables in Table 7. Estimated coefficients and standard errors (SEs) are reported in percentage.

	Dependent variable = Dollar volume of insider buying												
	Week(-3)	Week(-2)	Week(-1)	Week(0)	Week(1)	Week(2)	Week(3)	Week(4)	Week(5)	Week(6)	Week(7)	Week(8)	Week(9)
Speculation(Z)	23.132***	20.916**	24.360***	24.057***	23.836***	36.369***	49.362***	38.129***	33.601***	20.689**	26.999***	31.381***	26.373***
	(8.383)	(9.050)	(8.240)	(8.270)	(8.248)	(9.516)	(11.488)	(9.980)	(11.170)	(9.184)	(9.639)	(10.816)	(6.931)
Sentiment(Z)	-8.951	-8.446	-0.209	-0.880	-2.867	-7.230	4.575	10.509	-3.294	-5.164	-5.823	-0.732	-2.157
	(6.616)	(7.044)	(7.614)	(8.235)	(7.334)	(8.037)	(9.803)	(7.652)	(8.220)	(6.731)	(6.813)	(6.779)	(5.678)
Market value(Z)	24.452***	27.172***	21.717***	29.732***	25.939***	30.885**	37.308***	11.052	5.615	27.702***	46.466***	29.984***	37.298***
	(6.795)	(8.796)	(8.290)	(6.435)	(7.966)	(12.694)	(10.846)	(11.562)	(11.709)	(6.931)	(6.649)	(8.412)	(7.874)
Book-to-market(Z)	-23.806***	-28.253***	-22.064***	-6.859	-16.524**	-4.350	-4.602	-22.475***	-7.481	-10.599	-8.260	-12.069	-12.031**
	(7.209)	(7.589)	(7.307)	(6.735)	(7.473)	(8.397)	(9.607)	(8.454)	(7.688)	(7.379)	(6.114)	(9.659)	(5.294)
Turnover(Z)	38.932***	59.125***	43.910***	41.598***	35.660***	43.496***	38.853***	37.991***	39.127***	24.189***	43.479***	27.418***	37.282***
	(6.712)	(7.485)	(7.196)	(7.238)	(7.865)	(10.204)	(11.401)	(8.927)	(7.687)	(8.034)	(6.767)	(8.310)	(5.634)
Institutional ownership(Z)	15.992*	-13.538	5.212	-2.827	4.715	0.680	-7.290	17.916	-5.094	12.458	0.530	16.609*	1.951
	(8.272)	(9.093)	(8.439)	(8.126)	(7.919)	(9.220)	(11.886)	(12.511)	(12.269)	(9.083)	(8.509)	(9.649)	(7.861)
Fama-French alpha(Z)	8.595	-1.948	7.756	2.163	1.759	2.369	-9.619	-12.123	12.728*	6.745	-1.990	6.800	4.343
	(6.478)	(5.751)	(6.033)	(6.437)	(5.263)	(8.769)	(10.057)	(9.855)	(6.988)	(8.870)	(7.183)	(6.156)	(6.694)
Filing-day abnormal return(Z)	-5.989	3.049	10.768**	4.763	-12.052***	1.304	-3.102	8.407	-7.171	1.100	-4.119	8.676	5.158
	(5.457)	(4.065)	(4.667)	(4.272)	(4.451)	(3.959)	(8.237)	(6.035)	(5.926)	(8.287)	(5.966)	(5.473)	(4.202)
Nasdaq dummy	-21.650	-41.511***	-40.229***	-37.666***	-30.573**	-8.721	-58.495***	-20.111	-49.249**	-29.172	-27.152**	-16.799	-45.731***
	(15.461)	(10.754)	(11.609)	(13.275)	(13.562)	(16.864)	(21.103)	(22.538)	(18.875)	(18.077)	(12.873)	(14.437)	(10.424)
Fixed effect	Filing year-month												
Clustered SE	Firm / Filing year-month												
Observations	1939	1824	1774	1448	1896	1421	1056	1085	1258	1444	1769	2054	2247
Adjusted R^2	0.141	0.167	0.134	0.124	0.125	0.102	0.146	0.135	0.057	0.079	0.146	0.111	0.152

Table A.7: Speculative Language in Disclosure and News Sentiment (All Slope Coefficients)

This table presents coefficient estimates for all control variables in Table 8. Estimated coefficients and standard errors (SEs) are reported in percentage.

	Dependent variable = Average news sentiment score												
	Week(-3)	Week(-2)	Week(-1)	Week(0)	Week(1)	Week(2)	Week(3)	Week(4)	Week(5)	Week(6)	Week(7)	Week(8)	Week(9)
Speculation(Z)	0.014	-0.225	-0.069	0.175	0.330*	0.595***	0.758***	0.981***	0.372**	0.410**	0.534***	0.103	0.009
	(0.133)	(0.173)	(0.134)	(0.186)	(0.192)	(0.212)	(0.230)	(0.161)	(0.167)	(0.200)	(0.161)	(0.185)	(0.168)
Sentiment(Z)	-0.018	-0.244	-0.069	0.003	-0.534**	-0.213	-0.173	0.258	-0.567***	-0.376	-0.341**	0.105	-0.248
	(0.128)	(0.153)	(0.152)	(0.222)	(0.210)	(0.266)	(0.193)	(0.172)	(0.204)	(0.234)	(0.166)	(0.162)	(0.239)
Market value(Z)	0.151	0.059	-0.250	0.352	0.348	0.401*	0.497**	-0.127	0.671***	0.382*	0.950***	1.025***	0.733***
	(0.153)	(0.145)	(0.167)	(0.236)	(0.255)	(0.234)	(0.193)	(0.205)	(0.191)	(0.209)	(0.217)	(0.194)	(0.208)
Book-to-market(Z)	0.111	0.031	-0.262**	-0.210	-0.224	-0.062	-0.389***	-0.206	-0.242	-0.368*	0.046	0.313*	-0.239
	(0.115)	(0.143)	(0.126)	(0.203)	(0.185)	(0.183)	(0.117)	(0.154)	(0.160)	(0.200)	(0.160)	(0.181)	(0.166)
Turnover(Z)	-0.045	0.023	-0.133	-0.984***	-0.092	0.422*	0.283	0.000	-0.417*	-0.011	-0.560**	-0.366**	0.006
	(0.184)	(0.164)	(0.139)	(0.269)	(0.187)	(0.250)	(0.238)	(0.234)	(0.236)	(0.249)	(0.257)	(0.184)	(0.274)
Institutional ownership(Z)	-0.095	-0.313*	-0.174	-0.577*	-1.995***	-1.724***	-1.905***	-1.203***	-1.581***	-1.088***	-0.848***	-0.648***	-1.006***
	(0.197)	(0.184)	(0.186)	(0.296)	(0.251)	(0.295)	(0.332)	(0.346)	(0.325)	(0.227)	(0.251)	(0.212)	(0.277)
Fama-French alpha(Z)	-0.092	-0.193	-0.577***	0.145	-0.629**	-0.196	-0.111	-0.022	0.406*	0.833***	0.247	0.658***	-0.078
	(0.181)	(0.217)	(0.152)	(0.304)	(0.283)	(0.335)	(0.295)	(0.263)	(0.245)	(0.276)	(0.249)	(0.233)	(0.261)
Filing-day abnormal return(Z)	-0.269*	0.030	0.283***	2.822***	-0.052	-0.454*	0.002	0.094	0.139	0.282	-0.021	0.105	-0.241
	(0.143)	(0.102)	(0.105)	(0.222)	(0.192)	(0.238)	(0.240)	(0.165)	(0.183)	(0.189)	(0.135)	(0.166)	(0.184)
Pre-filing news sentiment(Z)	13.552***	13.777***	14.141***	2.636***	2.865***	2.840***	2.536***	2.628***	2.234***	2.318***	2.609***	2.421***	2.195***
	(0.185)	(0.172)	(0.232)	(0.269)	(0.203)	(0.247)	(0.268)	(0.261)	(0.242)	(0.205)	(0.263)	(0.203)	(0.203)
nth-week market news sentiment(Z)	0.507***	0.198	0.051	0.619**	-0.139	0.307	0.168	0.372**	-0.033	0.765***	0.391**	0.328	0.493**
	(0.133)	(0.122)	(0.168)	(0.269)	(0.184)	(0.287)	(0.279)	(0.164)	(0.188)	(0.216)	(0.186)	(0.236)	(0.201)
Clustered SE	Firm / Filing year-month												
Observations	21632	21499	22047	18274	19146	18568	19093	19208	20189	21266	21833	21886	21232
Adjusted R^2	0.310	0.324	0.349	0.027	0.024	0.022	0.021	0.021	0.019	0.017	0.018	0.014	0.012

Table A.8: Speculative Language in Disclosure and Subsequent Earnings Surprise

This table presents results from the regression of an indicator or two rank functions for future standardized unexpected earnings (SUE) on our speculation measure as follows:

$$Earnings\ surprise_{itq} = \alpha_q + \beta_q Speculation_{it} + \eta_q^{'}\mathbf{X}_{it} + \gamma_q EARMkt_{tq} + \epsilon_{itq},$$

where $Earnings\ surprise_{itq}$ is one of three functions of the nearest future SUE in the qth quarter to stock i's 10-K filing-day window in year t. For the SUE indicator, $Earnings\ surprise_{itq}$ takes 1 or -1 when the nearest future SUE in the qth quarter is above or below zero, respectively. For the SUE quintile rank, $Earnings\ surprise_{itq}$ is +2, +1, 0, -1, or -2 when the nearest future SUE in the qth quarter is above 80%, between 80% and 60%, between 60% and 40%, between 40% and 20%, or below 20%, respectively, where the percentiles are computed based on all available SUEs of the other firms (excluding stock i) within the three-week period before the nearest future SUE. For the SUE decile rank, $Earnings\ surprise_{itq}$ is +5, +4, +3, +2, +1, -1, -2, -3, -4, or -5 based on the corresponding decile cutoffs similarly. Each of these three SUE functions is scaled to have the standard deviation of one. Quarter 0 means that the nearest future earnings announcement and 10-K filing are made in the same quarter, and Quarter 1 means that the nearest future earnings is announced in the next quarter to 10-K filing. $Speculation_{it}$ is the percentage of speculation cues (out of total words), and $\mathbf{X}_{it}$ is a column vector of our baseline control variables as in Table 4 and two additional analyst-related variables: Analyst dispersion and Analyst revision. $EARMkt_{tq}$ is the market earnings-announcement-day return to capture the common economic shocks that potentially affect all individual firms' SUEs. The detailed definitions of these explanatory variables are provided 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-quarter to account for the serial and cross-sectional correlations in $Earnings\ surprise_{itq}$, respectively. Estimated coefficients and SEs are reported in percentage. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% level, respectively.

	Dependent variable = Earnings surprise					
	Quarter 0			Quarter 1		
	Indicator	Quintile	Decile	Indicator	Quintile	Decile
	(1)	(2)	(3)	(4)	(5)	(6)
Speculation(Z)	6.058**	6.054***	6.054***	1.854**	1.687**	1.429
	(2.709)	(2.185)	(2.271)	(0.922)	(0.839)	(0.880)
Sentiment(Z)	1.639	-0.831	-0.908	-0.117	-2.428**	-2.634***
	(1.784)	(1.798)	(1.830)	(0.988)	(0.926)	(0.914)
Market value(Z)	12.205***	4.199	3.986	11.219***	3.254**	3.214**
	(2.352)	(3.005)	(3.252)	(0.917)	(1.259)	(1.426)
Book-to-market(Z)	1.233	7.270***	8.199***	0.869	5.917***	6.514***
	(1.914)	(2.565)	(2.498)	(0.900)	(0.918)	(0.980)
Turnover(Z)	6.988**	8.498***	9.200***	1.781	5.019***	5.469***
	(2.683)	(2.975)	(3.259)	(1.214)	(1.078)	(1.169)
Institutional ownership(Z)	1.339	-2.457	-2.169	3.846***	2.018	1.809
	(2.462)	(2.613)	(2.626)	(1.215)	(1.460)	(1.467)
Fama-French alpha(Z)	8.222***	1.683	2.543	9.902***	5.362***	5.784***
	(2.434)	(2.541)	(2.828)	(1.635)	(1.413)	(1.554)
Filing-day abnormal return(Z)	6.861***	5.952***	6.151***	2.054***	1.393*	1.448*
	(1.899)	(2.085)	(2.141)	(0.675)	(0.767)	(0.788)
Analyst dispersion(Z)	-7.710***	-5.178***	-5.354***	-7.748***	-4.759***	-4.993***
	(1.950)	(1.661)	(1.845)	(1.608)	(0.758)	(0.749)
Analyst revision(Z)	0.704	1.406	1.472	-0.270	0.278	0.386
	(2.351)	(2.700)	(2.773)	(0.623)	(0.872)	(0.900)
Market earnings-announcement	0.056	1.013	0.998	0.595	2.738***	2.793***
-day return(Z)	(1.582)	(2.258)	(2.301)	(0.901)	(0.709)	(0.748)
Clustered SE	Firm / Filing year-quarter					
Observations	2626	2626	2626	26345	26345	26345
Adjusted R^2	0.047	0.016	0.017	0.034	0.012	0.012

A.1 Discussion on the SUE Predictability

In this section, to understand what value-relevant information comes with the speculative language in 10-K disclosures on a deeper level, we additionally test whether our speculation measure can predict subsequent earnings surprise, a proxy for news on future cash flows, after 10-K release dates.

For each firm and each quarterly earnings announcement, we first compute standardized unexpected earnings (SUE) as actual earnings minus the mean of analysts' forecasts divided by price which is available as of one day before the earnings announcement day. We then estimate the following regression specification with an indicator or two rank functions of future SUE as the dependent variable:

$$\textit{Earnings surprise}_{itq} = \alpha_q + \beta_q \textit{Speculation}_{it} + \eta_q^{'}\mathbf{X}_{it} + \gamma_q \textit{EARMkt}_{tq} + \epsilon_{itq}, \tag{A.1}$$

where $\textit{Earnings surprise}_{itq}$ is one of three functions of the nearest future SUE in the qth quarter to stock i's 10-K filing-day window (i.e., four days between the 10-K filing day and three days later) in year t. For the SUE indicator, $\textit{Earnings surprise}_{itq}$ takes 1 or -1 when the nearest future SUE in the qth quarter is above or below zero, respectively. For the SUE quintile rank, $\textit{Earnings surprise}_{itq}$ is +2, +1, 0, -1, or -2 when the nearest future SUE in the qth quarter is above 80%, between 80% and 60%, between 60% and 40%, between 40% and 20%, or below 20%, respectively, where the percentiles are computed based on all available SUEs of the other firms (excluding stock i) within the three-week period before the nearest future SUE. For the SUE decile rank, $\textit{Earnings surprise}_{itq}$ is +5, +4, +3, +2, +1, -1, -2, -3, -4, or -5 based on the corresponding decile cutoffs similarly. Each of these three SUE functions is scaled to have the standard deviation of one. Quarter 0 means that the nearest future earnings announcement and 10-K filing are made in the same quarter, and Quarter 1 means that the nearest future earnings is announced in the next quarter to 10-K filing. $\textit{Speculation}_{it}$ is the percentage of speculation cues (out of total words), and $\mathbf{X}_{it}$ is a column vector of our baseline control variables as in Table 4 and two additional analyst-related variables: Analyst dispersion and Analyst revision. $\textit{EARMkt}_{tq}$ is the market earnings-announcement-day return to capture the common economic shocks that potentially affect all individual firms' SUEs.[26] For the ease of interpretation, all explanatory variables are standardized to have the mean of zero and the standard deviation of one, and their detailed definitions are provided in the Appendix B. To account for the serial and cross-sectional correlations of SUEs, standard errors are clustered by firm and filing year-quarter, respectively. Estimated coefficients and standard errors are reported in percentage.

The test results of SUE predictability in equation (A.1) are presented in Table A.8. In Columns (1) to (3), when the nearest future SUE and 10-K filing are in the same quarter, the number of observations is significantly smaller (than the other columns), which can potentially lead to lower test power. Despite of this limitation, we find that our speculation measure can predict future SUE positively and significantly at the 5% level for all three

[26] Our market-wide earnings-announcement-day return is value-weighted across firms that have earnings announcements in each month. Thus the value of $\textit{EARMkt}_{tq}$ depends on the actual month when each future SUE is calculated in equation (A.1). Our findings and conclusion from Appendix Table A.8 are robust to equal-weighting.

SUE functions. In Columns (4) to (6), when the nearest future SUE is in the next quarter to the corresponding 10-K filing, the number of observations substantially increases relative to the first three columns, and we continue to find that the speculative language in 10-Ks predicts positive future SUE (except Column (6)). Furthermore, the coefficient estimate for *Speculation* is much larger for Quarter 0 than for Quarter 1, which indicates that the magnitude of SUE predictability by speculation attenuates quickly as a longer time horizon is allowed for security analysts to digest the informational implication that comes with speculative language in 10-K disclosures.[27]

In sum, the test results in Table A.8 support the idea that the speculative language used in 10-K disclosures contains the new information on firms' cash flows in the near future and that security analysts initially under-react to it possibly due to its embedded immaturity although they eventually digest and reflect its implication related to future cash flows into their earnings forecasts.

[27] We also estimate equation (A.1) for Quarter 2 when the nearest future SUE is required to be in the second next quarter to each 10-K filing, and find positive but insignificant coefficient estimates of *Speculation*.