# Information or Spin? Evidence from Language Differences Between 8-Ks and Press Releases

Paul Ma*

JOB MARKET PAPER

January 25th, 2013

**Abstract**

This paper examines whether investors correctly distinguish qualitative information from promotional language in press releases related to material events of US public firms. For a variety of material events, firms are required to issue a Form 8-K, but 37% of the time also voluntarily issue a press release concerning the same event. Using textual analysis, I find that firms are more likely to issue a press release if the underlying 8-K tone is positive, and that tonal differences between the 8-K and the press release are driven in part by quotes from firm officers. I also find economically significant responses in firms' stock returns to tonal language in the 8-K, as well as to tonal differences between the two disclosures. To verify whether my strategy of comparing the press release against the 8-K is isolating the effects of promotional language or additional information, I test and find evidence of an initial positive reaction but a subsequent negative drift from positively toned press releases. This evidence implies that investors may have initially responded to both information and spin. Nominating investor inattention as a possible mechanism for overreaction, I use novel search traffic micro-data from the SEC EDGAR website and detect lower 8-K search intensity in the presence of a press release. Together, my results are consistent with some investors overestimating the degree of substitutability between the two disclosures and thus failing to readjust expectations accordingly.

**JEL:** D03, D83, G02, G14, M41
**Keywords:** voluntary disclosure, mandatory disclosure, persuasion, EDGAR search traffic

*Department of Economics, Stanford University. I am extremely grateful for valuable insights and support from my dissertation committee: David Larcker, Charles Lee, Ivan Marinovic, and John Shoven. I owe particular thanks to Scott Bauguess at the SEC who assisted me with the EDGAR traffic data and questions concerning the 8-K, and to Paul Tetlock for his generous feedback. I am also very thankful to Mary Barth, Elizabeth Blankespoor, Joseph Grundfest, Dirk Jenter, Allan McCall, Alejandro Molnar, Kester Tong, Charles Wang, as well as seminar participants from Carnegie Mellon and Stanford for their comments and suggestions. All views expressed are my own. Email: paulma@stanford.edu

# 1 Introduction

For public firms, the press release is perhaps the most common channel of communication to inform investors of potentially relevant information. While firms are bound by the Securities and Exchange Commission (SEC) to report truthfully[1], language, unlike numbers, is inherently flexible, and there is often more than one way to convey the same underlying set of facts. This latitude in language is especially relevant when describing unexpected events concerning qualitative information, which can be more challenging for the market to process into expectations for future cash flows or discount rates than for quantitative information such as earnings. Not surprisingly, firms devote considerable resources towards "communication expertise", such as public relations, which is an obvious input factor into the production function of press releases.[2] Consequently, one viewpoint is that press releases contain a mixture of factual information and non-informative persuasive content with the intent to bias investors' beliefs. The main question this study asks is whether investors correctly distinguish qualitative information from promotional language in press releases related to material events of US public firms.

Existing theories of voluntary disclosure and cheap talk argue that investors should respond to the portion of the information signal which they find credible and ignore non-credible talk (Benabou and Laroque, 1992; Stocken, 2000). These concepts have also been carried over into the litigation arena, where the doctrine of puffery is also predicated on the idea that a reasonable investor would be able to ignore puffery or sales talk (Padfield, 2007).[3] Testing these theories remains an empirical challenge, however, because researchers generally do not have access to a counterfactual of the press release without the "spin". Thus, despite firms' revealed preferences toward communication expertise, the literature is largely silent regarding the frequency and degree to which firms are injecting non-informative but persuasive language into press releases and the consequences of such actions, if any.

To overcome this challenge, I focus on a broad set of non-earnings related material events that the SEC requires firms to disclose publicly through a Form 8-K.[4] In contrast to pre-scheduled disclosures such as earnings, 8-Ks are largely unanticipated and irregularly spaced throughout the year. They often also concern events which are not easily comparable either across firms or over time, and hence are more difficult to associate with hard information

[1] SEC Rule 10b-5 prohibits firms from making fraudulent statements.

[2] According to Veronis Suhler Stevenson, a private equity firm specializing in the communications industry, US firms spent $3.7 billion on public relations activities in 2009 and are forecasted to increase spending to $8 billion in 2013.

[3] An application of the puffery doctrine may be used to undermine the ability of the plaintiff to satisfy the element of materiality. However, legal courts have been inconsistent in its application as a valid legal defense due to the inherent subjectivity in identifying puffery (Rogers *et al.*, 2011).

[4] Firms also provide earnings press releases as an 8-K filing, which is not the focus of this paper; unless otherwise stated, my reference to 8-Ks excludes earnings based filings.

about future cash flows. Given the 8-K's unanticipated and qualitative nature, both the effectiveness and incentive for the firm to engage to spin may be ex-ante greater if time constraints catch both analysts and journalists off guard such that they cannot respond in a timely manner (Gandy, 1982; Turk, 1986) in order to provide a competing viewpoint.

Common events which trigger an 8-K include "entry into a material agreement" and "appointment or resignation of directors and company officers".[5] In addition to filing the 8-K, firms voluntarily issue a press release concerning the same event 37% of the time, half of which occur prior to the 8-K filing date. This disclosure environment creates the opportunity to employ textual analysis to contrast language differences between the two disclosures under the assumption that, due to SEC enforcement, firms are more constrained in the usage of extraneus language in the 8-K relative to the press release. Given that firms already satisfy their reporting requirements through the 8-K, a natural suspicion is that the press release contains information that is contained in the 8-K but repackaged using more persuasive and positive language. The existence and timing of the press release relative to the 8-K also raises the concern that investors may be inattentive to the subsequent 8-K because they believe the two disclosures to be adequate substitutes, a hypothesis I test directly and confirm later through an analysis of novel search traffic data from the SEC EDGAR website. An alternative interpretation, however, is that firms view the press release as a vehicle to release additional qualitative information beyond the requirements of the 8-K.

I use standard textual analysis techniques to generate measurements of positive and negative tone in the qualitative disclosures and find economically large abnormal returns in response to both the underlying tone of 8-K, and in response to tonal differences between the 8-K and the associated press release around the short term window beginning with the press release and ending with the 8-K filing. However, over the post 8-K announcement window, I find evidence of negative drift as a function of the degree of positive-toned words in the press release after controlling for tone in the 8-K. Together, my results suggest that investors react to both qualitative information and to promotional language, which is subsequently revealed over time.

As an example of possible language differences, consider the following series of events. On November 9, 2005, Mr. Williams, the co-founder of Manakoa Services, a securities technology company, submitted a letter of resignation to the board of directors. Two days later, the firm issued the following press release:

**Press Release**

*It became apparent to Manakoa's board of directors that the board and Mr. Williams had philosophical differences regarding the focus, financing and direction of the*

[5]Table A.I provides the entire set of general categories.

*company. The board will take the opportunity to refocus management on short-term, critical objectives such as improving sales, customer relations and bolstering earnings. We wish Mr. Williams all the success and happiness with his future endeavors.*

On November 29, 2005, the firm filed (late) the required 8-K, which included Mr. William's original resignation letter dated November 7:

**8-K**

*The Company strongly disagrees with the contents of Dr. Williams's resignation letter and regrets that he felt the need to write such a letter. It is our understanding that Dr. Williams has recently suffered serious head injuries and trauma. Hopefully, as he recovers he may regain his perspective. However, should he persist in his behavior and his assertions, the Company will vigorously defend itself in any resulting litigation.*

The admittedly colorful example provides intuition for the idea that the firm selectively framed the resignation in the press release despite full knowledge of the resignation letter but was eventually forced to reveal the acrimonious details through the 8-K platform.[6]

My study uses the entire sample of 8-Ks and associated press releases filed with EDGAR from 1996 to 2011.[7] To quantify the information in the press release and 8-K, I use word lists constructed from Loughran and McDonald (2011) to generate positive- and negative-toned word frequencies for each 8-K and press release. I chose this dictionary relative to alternatives because it is public and non-proprietary and, more importantly it is specifically constructed from the language patterns of the annual 10-K reports.

Recognizing that press releases are voluntary,[8] I find that firms are more likely to issue them if the underlying 8-K tone is more positive, where tone is defined as the number of positive minus negative words, scaled by the word count. Firms also implicitly choose the actual tonal difference between the 8-K and the press release, which I dub the "sentiment bias". The sentiment bias is on average positive, and investigating its determinants, one source is the inclusion of quotes from company officers in the press release, which are generally effusive in nature, but absent from 8-Ks. Press releases written with the assistance

[6]The 8-K's instructions compel firms to disclose director resignation letters. The entire filing is available on EDGAR at
http://sec.gov/Archives/edgar/data/1091967/000119983505000675/0001199835-05-000675-index.htm.

[7]The acronym EDGAR stands for Electronic Data-Gathering Analysis and Retrieval System, which is hosted on the SEC website.

[8]There are specific caveats to this in the face of recent SEC interpretive guidance with respect to Regulation Fair Disclosure FD and disclosure through corporate websites. I discuss this in more detail in the robustness section.

of an external investor or public relations person are also 20% more optimistic than their in-house counterparts and driven by their usage of both more positive words and fewer negative words.[9]

In tests of market reaction, I find abnormal firm returns of 88 to 109 basis points surrounding the event window which encapsulates both the press release and the 8-K for each standard deviation increase in the sentiment bias after controlling for the underlying 8-K tone. The response to the 8-K itself is also positive, with each standard deviation increase in tone generating 112 to 126 basis points in abnormal returns. Magnitudes are large economically and correspond to more than the effect of a decile change in earnings surprise on returns. The response to the sentiment bias indicates that investors perceive the voluntary information released through the press release above and beyond the 8-K as credible and not simply noise.

To investigate whether investors also respond to promotional language in press releases, I test for post 8-K filing date price drift as a function of the degree of optimism in the press release after controlling for the 8-K. I find evidence of a modest price drift of approximately -13 to -20 basis points spread over the month following the 8-K in response to one standard deviation increase in the press release's positive tone. Exploring further, I find the response to be exclusively tied to the frequency of positive words for press releases that precede the 8-K. In contrast, I find evidence of positive drift to tone for 8-Ks which do not have an associated press release, suggesting investors in general under react to 8-Ks, unless firms issue a related press release.

Dynamic disclosure models argue that investors should de-sensitize themselves if a firm acquires a reputation for being bombastic in their language (Stocken, 2000). Therefore for many specifications, I also use within firm variation and as expected obtain even stronger results, which indicate that investors are more responsive to the degree of over-optimism relative to the firm's own norms.

A candidate market friction rationalizing my results is that investors may be inattentive to the 8-K when they also have access to the press release. I test directly for the possibility that investors' demand for the 8-K depends on the availability of a press release using search traffic data between 2008 and 2011, representing all activity on the SEC's EDGAR website. Using search traffic measured at the level of the individual 8-K, I detect a 30 to 50% decrease in the number of unique searchers for the 8-K in the presence of a press release, and an even larger, 80% decrease for the sub-sample of 8-Ks that are released simultaneously with a press release. These findings in combination with the negative price drift are consistent with inattention and search costs as a channel through which investors overreact to a press

[9]These are often external investor relations or public relations contacts. I discuss this more in the data section.

release, but subsequently fail to re-adjust their expectations, even after the 8-K is eventually released.[10]

This paper connects with several lines of research in the accounting and finance literature. It makes a direct contribution to the sparse literature on 8-Ks and directly answers the call for research by Beyer *et al.* (2010) who recognize the need to "apply natural language techniques ... to better understand how capital markets react to and process the detailed information provided on Form 8-K." It also closely relates to the voluntary disclosure literature and specifically Lang and Lundholm (2000), who find that firms hype themselves through more positively toned soft disclosures prior to seasoned equity offerings in order to reduce their cost of capital. In a consistent fashion to their findings but applied to a broader set of disclosures, the evidence here suggests that the market is unable to completely discount spin or hype at the level of the individual disclosure which resulted in negative post announcement drift.

More broadly, this paper contributes to the growing empirical and experimental (Cain *et al.*, 2005) literature which finds evidence that investors have difficulty in de-biasing information from ex-ante known non-neutral sources such as analyst recommendations (Malmendier and Shanthikumar, 2007a), firms' accounting information (Hirshleifer *et al.*, 2004), and media reporting (Dyck and Zingales, 2003). Complementing these studies, this paper provides new evidence of investors' subsequent information acquisition behavior in response to a biased information signal (DellaVigna and Gentzkow, 2009). The finding that EDGAR users, who are ex-ante more likely to be retail investors, search less for the 8-K when there is a press release is new and surprising, but also rationalizeable if investors face search costs.

The paper also adds to the vast literature that applies textual analysis to qualitative information embedded within regulatory filings, press releases, and media content. The overwhelming majority of these studies however, focus on the contents of a single information source[11], whereas this study analyzes the textual interaction between 8-Ks and associated press releases.[12]

As an additional contribution beyond previous textual analysis papers, I find market reactions not only to the net positive word measure, but also to the individual components of positive and negative tone measures. In contrast, existing literature uses either 1) the negative word list only, or 2) the difference between the number of positive and negative words, because researchers have had difficulty showing any market response to positive words

[10] Another implication of the spin hypothesis is that over-reaction will be greater if investors do not read the 8-K, which I do not currently explore.

[11] The pitfall of relying on a single information source is also echoed in Karpoff *et al.* (2012).

[12] The only paper I am aware of which specifically examines textual interactions across information sources is Tetlock (2011), more details are discussed in the literature section.

(Engelberg, 2008). While the latter measure is intuitive to explain, it imposes an arbitrary equal weighting scheme on the effects of positive and negative words, an assumption that is not justified since investors respond asymmetrically to bad news (Kothari *et al.*, 2009b). These results indicate that investors also respond directly to positive tone, consistent with the intention of spin, and not only to the lack of a negative tone.

My study also complements existing studies that ask if investors respond to tone in earnings press releases above and beyond the hard earnings information (Davis *et al.*, 2012; Demers and Vega, 2011). These studies find positive drift following positively toned earnings press releases, whereas I find negative drift following 8-K filings with relatively optimistic press releases, which emphasizes the point that disclosure environments matter and therefore results from one setting may not carry over to another.[13]

The rest of the paper is organized as follows. Section 2 provides background on 8-Ks and discusses related literature. Section 3 describes the data, hypotheses, and empirical evidence of investors' reaction. Section 4 provides search traffic evidence of substitution between the 8-K and the press release. Section 5 provides a discussion of robustness concerns. Finally, Section 6 concludes.

# 2 Background and Literature

## 2.1 8-K Background

The SEC requires firms to submit 8-K filings via EDGAR in response to a wide variety of corporate events. Examples of events include "entry into a material agreement", or "changes to the board of director or principal officers". Appendix Table A.I provides the full list of required disclosures as well as their filing frequencies. According to the SEC Final Rule on August 23rd, 2004, the disclosures represent "events that are unquestionably or presumptively have such significance that current disclosure should be required." The instructions for the 8-K are specific: firms must disclose and classify the event which triggered the 8-K within four business days of the actual event date, and must succinctly describe the circumstances surrounding the event. Facing regulatory scrutiny, for example SEC comment letters, the language within the 8-K is generally to the point, and is often written with legal assistance and signed by the firm's general counsel.[14]

On June 17th, 2002, as part of the Sarbanes-Oxley initiative the SEC proposed to in-

[13] In un-tabulated results I am also able to replicate their findings through the sub-sample of earnings related 8-Ks.

[14] This is from conversations with a senior law partner who specializes in 8-Ks. As an independent measure, I extract the signature portion of the 8-K and find that 18% of 8-Ks are signed by the general counsel and 31% for firms in the S&P 500 index, which are likely to be lower bounds for the true propensity.

crease the number of corporate events reportable under the form 8-K. Prior to the proposal, domestic firms were required to file an 8-K in response to nine separate corporate events. The proposal, with modifications, came into effect on August 23, 2004 and included 12 to 14 new events which became reportable.[15] The new 8-K also shortened the filing deadline from the event date, from up to fifteen calendar days previously down to four business days uniformly across all 8-Ks.[16] The expansion of the 8-K in scope and timeliness corresponds with a massive spike in the time series of 8-Ks filings, which is evident from Figure 1. In the two year calendar window around the switch to the new 8-K, the number of filings increased 50% from approximately 20,000 in 2003 Q4 to over 30,000 in 2005 Q4.

Existing literature concerning the cross section of 8-Ks is limited.[17] Carter and Soo (1999) analyzed the determinants of reporting delay for firms in filing the old 8-K form. They also investigated absolute abnormal returns around the filing dates and generally found small reactions, mainly because the 8-Ks were filed so late relative to the reporting date, which led to pre-announcement leakages.[18] Their finding is consistent with my evidence that a fifth of 8-Ks are preceded by a press release. More recently, Lerman and Livnat (2010) undertook an analysis of the new 8-Ks which imposed the shorter filing deadline. They explored whether there is a market reaction to the 8-K filing and tabulated their results across different types of 8-K categories, for different event windows (the reported event date and the actual filing date), and for different measures of market reaction such as volatility, volume, and absolute returns. They found abnormal volume, volatility, and absolute returns around both the filing and the reporting dates. Both of these studies partition and report results by general 8-K classification headings, but do not attempt to measure the signed informational content at the level of the individual 8-K. This research gap serves as motivation for this paper's research design of applying textual analysis techniques for the purpose of quantifying the informational contents of 8-Ks.

## 2.2 Textual Analysis

In selecting the appropriate method for textual analysis, various options exist with the majority of them falling into two broad categories. The first is often referred as the "bag of words" method whereby a document is tokenized into individual words and matched

[15] 2 items in the older 8-K were substantially modified and can be construed as almost new Lerman and Livnat (2010).

[16] Firms are allowed to report multiple triggering events within the same 8-K as long as the earliest event is within the reporting deadline

[17] Many studies implicitly use the 8-K information within a narrow topical context. Examples include research on restatements, auditor changes, bankruptcies, and corporate governance, all of which are related to 8-K filings.

[18] This finding obviously does not suggest that 8-Ks are useless for reducing information asymmetry since firms might not have even released the press release in the counterfactual world of no 8-Ks.

to a specific dictionary containing relevant keywords of interest to create a count of the keywords as a measure of tone. The second category relies on more advanced techniques from computational linguistics. Examples include naive Bayesian classification (Li, 2010) and text regression (Kogan *et al.*, 2009). Much of the empirical literature has relied on the former approach which is also what I adopt in this paper, mainly because of tractability.

Common dictionaries of keywords include the Harvard Psychosociological IV-4 Dictionary (Tetlock *et al.*, 2008), General Inquirer (a superset of the Harvard dictionary) (Tetlock, 2007; Kothari *et al.*, 2009a), as well as proprietary options such as DICTION (Davis *et al.*, 2012). More recently, Loughran and McDonald (2011) introduced a new public and non-proprietary dictionary based on word usage specifically extracted from the language of 10-Ks. They highlight the advantage of their approach, by comparing their word list to the Harvard dictionary, considered one of the de-facto general usage dictionaries, and report large scale misclassifications due to word meanings which are domain specific. For example, the word "liability" is neutral in financial accounting, but negative in general usage dictionaries. Because of its domain specificity, recent research applying textual analysis to language in accounting and finance settings have largely chosen to use their dictionary in lieu of alternatives [19], a strategy which I follow.

Using these broad set of techniques, researchers in accounting, economics, and finance have analyzed the sentiment of news articles to predict stock returns (Tetlock *et al.*, 2008), the readability of 10-Ks and incentives to obfuscate firm performance (Li, 2008), whether CEOs and CFOs use deceptive language during earnings conference calls and its consequences (Larcker and Zakolyukina, 2012), the origins of media bias (Gentzkow and Shapiro, 2006), the effects of media coverage on firms' stock returns (Fang and Peress, 2009), favorable local media coverage for firms with advertising ties to the local newspapers (Gurun and Butler, 2012) and even the contents of internet chat boards (Wysocki, 1998; Antweiler and Frank, 2004).

## 2.3 Disclosure Theory and Empirical Evidence

My empirical questions are also motivated by a rich theoretical literature which seeks to understand the conditions under which managers disclose credible qualitative information and the corresponding investor response. Studies such as Benabou and Laroque (1992) and Stocken (2000) argue that repeated interactions and ex-post verifiability are key to

[19] Henry (2008) is an another dictionary which also targets financial reporting language. I do not know of any literature which compares it to Loughran and McDonald (2011) except that the former word list (as described in the text of the paper) appears to contains approximately 100 positive and 100 negative words, while the former contains over 300 positive, and 2300 negative words. I choose to use the latter dictionary, because its original data source (10-Ks) is closest to the text category I analyze.

truthful disclosure from the manager.[20] Consistent with these theories, Hutton *et al.* (2003) found that management forecasts for positive news only have market impact when they are accompanied by supplementary hard information, which provides a form of ex-post verifiability and also increases the cost to firms of managers from misrepresentation through exposure to higher litigation risk. My paper also overlaps with Rogers *et al.* (2011), who studied whether firms who issued optimistic earnings press releases were more likely to be targets of litigation. In contrast to my empirical strategy, they use a set of firms matched on observable characteristics to measure "abnormal" optimism with presumably the same goal of differentiating spin from actual information. Whereas they focus on the costs to firms from overly optimistic press releases through litigation risk, I expose potential benefits through short term biasing of investors' beliefs.

Existing empirical literature analyzing the impact of soft information has exclusively centered around the contents of earnings press releases Hoskin *et al.* (1986). Both Davis *et al.* (2012) and Demers and Vega (2011) find that investors respond to tonal language in the press releases above and beyond hard information. They also find evidence of drift in the direction of the tone of the press release which suggests that investors initially under-reacted to tonal information, which is the opposite of my result in the case of optimistic 8-K press releases. Earnings announcements are also different from 8-Ks in several respects, therefore results from one setting may not transfer to another. For example, they provide hard information concerning future cash flows, which is more absent in 8-Ks. Earnings announcements are also highly anticipated and visible events relative to 8-Ks, with high coverage from information intermediaries such as the media and analysts, thus both the effectiveness of spin, and hence the incentives to spin is likely lower than the setting of 8-Ks, which are largely unanticipated and less visible.

This research is also tightly related to Lang and Lundholm (2000) who argue that firms hype themselves through more positively toned soft disclosures prior to seasoned equity offerings in order to reduce their cost of capital. In a similar argument, they use the market's subsequent reaction as a test for whether the original sequence of disclosures reflected hype or the management's attempt to reduce information asymmetry. Whereas they measure hype on the extensive margin (the number of disclosures), I propose to measure and explain the effects of potential spin on the intensive margin (the degree within each disclosure) for a broader set of events in the context of 8-Ks. The connection to spin in the context of media coverage is explored in Solomon (2012), who found that firms who use an external public relations firm were able to obtain asymmetrically greater media coverage for good news relative to bad news. He finds firms which use external public relations subsequently

[20] Crawford and Sobel (1982) demonstrate that as long as the manager and investor's preferences are "sufficiently aligned", then cheap talk can also be credible.

fail to meet expectations during earnings announcements in part because of the excess media coverage which created unrealistic expectations.

In testing whether investors can "descramble" information from noise, my work also relates to Hand (1990), who found that investors were unable to fully distinguish between true cash flow information in the setting of debt-equity swaps. Frederickson and Miller (2004) find a similar result in an experimental setting where MBA students (but not professional investors) failed to appropriately distinguish GAAP from non-GAAP earnings, a result also echoed earlier in Bradshaw and Sloan (2002).[21] In another context, Tetlock (2011) argues that investors fail to distinguish between old and new information in media contents. He compares the textual content between new and old news articles to determine how much new information is presented and finds investor overreaction and subsequent reversal to the stale components of news.

In using EDGAR search traffic to understand whether investors substitute between press releases and 8-Ks, my work also connects to the scant literature which addresses the complementarity and substitutability of different informational sources. Lerman and Livnat (2010) found evidence that the introduction of the new 8-K, which included items previously disclosed in the 10-Q, did not diminish the subsequent information content of 10-Qs. They suggest that investors instead use the information provided in the 10-Q to re-interpret the earlier disclosures made in the 8-K, a hypothesis which is also explored in Drake *et al.* (2012a). In contrast, McNichols and Manegold (1983) found that the introduction of mandatory 10-Qs reduced the information content of 10-Ks.

# 3 Data and Empirical Strategy

## 3.1 Sample Construction

This section describes the data sample while Section 3.3 describes the main empirical tests in the paper. I use the master index file provided on the EDGAR website which provides each 8-K's unique accession code as an identifier. Using the accession code, I retrieve every 8-K filed until the end of 2011, which results in 1,046,487 unique 8-Ks (excluding amended filings). Each 8-K contains the report date, which is when the 8-K triggering event took place, the filing date, and the filing timestamp, which is the date and time EDGAR accepted and began disseminating the filing. In addition, I extract various characteristics of the filing such as which categories ("item numbers") the 8-K is listed under as reported by the firm,

[21] However in a similar strategy to my study, Dambra *et al.* (2012) contrast GAAP and non-GAAP based management cash flow forecasts(MCFF) and find that investors do not react to non-credible non-GAAP based MCFF.

the total number of items, and also the types of exhibits that are attached to the 8-K.

In order to identify the associated press releases of the 8-K, I take advantage of the fact that firms bundle previously issued related press releases within the 8-K filing system as an Exhibit 99 attachment according to SEC Regulation S-K Section 6.01. However firms do not always follow instructions, and therefore I use a large set of additional rules which are described in more detail in Appendix A to determine if a press release (or multiple press releases) exists within the 8-K. Once I detect a press release, I extract the text, date, and information of any listed contact persons in the press release.

It is also certainly possible that some 8-Ks refer to the press release but do not provide any other information. A large fraction of these types of 8-Ks relate to earnings announcements or Regulation FD disclosures. These referential 8-Ks by definition do not provide differential informational content through the 8-K, and therefore I may mis-attribute the information entirely to the press release when in a sense they contain equal information. There are several strategies to address this issue- one is mechanical via a minimum word count, which is what many other studies do as well[22]. An alternative method is to code the press release's sentiment as equal to the 8-K sentiment, since by the firm's choice, they reflect equal information. Applying these and other methods do not qualitatively change the findings of this paper and further discussion is available in the robustness section.

My research design relies on firms to provide related press releases as an attachment in the 8-K filing, which provides the necessary observation by observation linkage between the 8-K and the press release. By construction, this choice precludes me from examining potential press releases related to the 8-K which occur after the filing date. The main research concern however is that firms may issue related press releases, but then selectively choose not to bundle them, resulting in an unknown form of selection bias which is possible and a limitation of the research design. The direction and magnitude of the bias would however, depend on the model which describes the firm's bundling decision. The most benign scenario would be if firms randomly omitted bundling press releases because of miscommunication between the legal and public relations department. To investigate this important issue, for a random set of firms, I retrieve the entire universe of press releases, which is described in detail in Appendix B. This allows me to learn whether the set of 8-Ks which contain attached press releases are systematically different than those which do not contain press releases along the dimensions of overall press release disclosure behavior around the timing of the 8-K filing date. Using the random sub-sample, I find that the overall frequency of press releases before and after the 8-K filing date do not differ systematically between the sample of 8-Ks which have and do not have an attached press release. This is certainly not definitive evidence, but does assuage some basic concerns with respect to selection bias; the

[22]For example, papers which examine the tonal contents of 10-Ks typically impose a word count filter.

robustness section also provides further discussion.

Once the text of the 8-Ks and the press releases have been collected- I further process them to remove non text data such as HTML and tables, a common procedure in the literature. I also remove standard boilerplate language at the beginning and end of each 8-K and also any additional exhibits. The text within the press release is likewise processed to remove standard forward looking statements, or safe harbor statements, which are often embedded towards the end of the document. After processing, the collective documents represent a corpus of over 350 million words in the 8-Ks, and 500 million in the press releases.

In order to measure the tone or sentiment of each document, I first tokenize each document into a list of words ("unigrams"). This list is matched against the positive (number of unique words, N=354) and negative (N=2349) word lists which Loughran and McDonald (2011) assembled and made available online.[23]

In providing a measurement of tone, prior literature has largely focused on either the frequency of 1) negative words, or 2) the net difference of positive and negative words. There is existing evidence however that investors weigh good and bad news asymmetrically (Kothari *et al.*, 2009a). This makes the latter definition in-appropriate because it enforces equal weighting between the effects of positive and negative words. For comparability, I follow the literature and define the sentiment as the difference between the number of positive words and negative words, scaled by the total word count. This results in the 8-K and press release with its own sentiment score, denoted 8-K Sentiment and Press Sentiment respectively in the tables. However, I also construct the disaggregated measures of positive and negative tone as well to test if investors respond asymmetrically to positive and negative tone. All tonal measures are scaled by 100. To provide basic descriptive statistics, restricting to the sample of 8-Ks which also contain a press release, the average 8-K has .28% positive words and .51% negative words for a net 8-K tone of -.23%. In contrast, the average press release contains .87% positive, and .62% negative words for a net press tone of .25%. The average difference between each document pair, denoted the "sentiment bias", is .48%. In reference to the example offered in the introduction, it was identified through the sentiment bias which was measured at the 99.975% percentile.

To identify whether the firm is using an external public relations (PR) or investor relations (IR) firm[24] in writing the press release, I examine the contact information commonly provided in the press release. In particular, I look for the contact information of the press

[23] $http://www.nd.edu/\ mcdonald/Word_Lists.html$

[24] I use the term external contact person to refer to the collective. Bushee and Miller (2012) find that external investor relations can help bring visibility to firms through greater analyst coverage, and media attention. Solomon (2012) provides evidence that firms which hire external public relation firms are able to generate more media coverage when firms issue good news press releases.

release and extract the phone number and email domain name of each listed contact person and compare them against the firm's listed phone number and website domain. It is common for there to be both an internal contact person, and also an outside contact person with an obvious external IR or PR domain name. I classify the external contact person dummy variable as 1 if there is at least one person whose email domain does not match the email domain of the firm. In addition, I require that the first four digits of the phone number of the contact person match those of the firm.[25] If there is a discrepancy between the rules, I code the variable as missing. The resulting variable is a filing level indicator proxying the presence of an external IR or PR firm.

For additional context, I provide some general descriptive statistics beyond what is available in the existing 8-K literature. Figure 1 plots the number of 8-Ks filed, the number of filing firms, the average number of item numbers per 8-K, and the total number of items filed per year quarter from the entire universe of EDGAR filings. The spike in filings after 2004 is a direct consequence of the expansion of corporate events which became mandatory disclosures after the new 8-K form. Figure 2 plots the median length of the 8-K and press release texts over time. The length of the 8-K text was in decline from 200 words to 150 words until the new 8-K form in 2004 when it began to increase in response to the increased disclosure requirements. In comparison, the length of the press release has also increased over time from 400 words in 1995 to 500 in 2011, consistent with Francis *et al.* (2002)'s findings in the case of earnings press releases.

To understand the timing of 8-Ks in relation to other events, Figure 1b displays the frequency of earnings announcements as well as other SEC filings such as the 10-K, 10-Q, Form 4 (insider trading), and DEF 14A (annual proxy statement) as a function of the timing of the 8-K filing date which is centered at time=0. As expected, a large fraction (above 15%) of 8-Ks coincides with earnings announcements which is a mechanical relationship since firms are required to file an 8-K item number 2.02 "Results of Operations" in response to Regulation G. There are also smaller incidences of the 8-K coinciding with the 10-Q, and the 10-K. Insider trading filings are increasing up to 5 days prior to the 8-K filing date and continue to increase up to 5 days after the filing. The timing of insider trading relative to 8-Ks raises the possibility of strategic insider trading and disclosure (Cheng and Lo, 2006; Jagolinzer, 2009).

In terms of the timing of press releases relative to the 8-K, Figure 4a plots the frequency of 8-Ks which have a preceding press release as a function of the number of days prior to the eventual 8-K filing date. Approximately half of 8-Ks with press releases actually precede the 8-K, with the majority preceding just the day before. Because press releases do not contain timestamps, I do not know if the press release was released after hours, which would

[25] Only US-based digit numbers are extracted. See Appendix A for complete details.

allow me to classify them into the next trading day. Figure 4b also plots the average 8-K sentiment, average press release sentiment, and the difference (the sentiment bias), as a function of the number of trading days the press release was issued prior to the 8-K filing date, for the sample of 8-Ks with attached press releases. Despite the 8-K sentiment staying largely flat over time, the degree of sentiment in the press release is greater the earlier it precedes the 8-K date.

Figure 5 plots the frequency of positive and negative words for the press release and the 8-K. The level of negative tone is largely similar between 8-K and press releases over time. However we observe a large gap in the degree of positive tone between the press release and 8-K, with the former much more positive. While it is speculative to make conclusions based on raw time series, the increase(decrease) of negative(positive) tone around the financial crisis in 2008 seems to indicate that the tonal measures are able to proxy to a degree actual firm performance. Figure 6a plots the 8-K sentiment against the associated press release and finds a positive relationship if we expect the press release to equal the 8-K plus promotional language. Figure 6b plots the 8-K sentiment against the sentiment bias (the difference between the 8-K and the press release sentiment) and finds a negative relationship. This finding implies that firms seem to either release more positive information in the press release when the 8-K is negative, or alteratively spin more.[26]

Panel A of Table II reports summary statistics of the textual variables for the universe of 8-Ks. First, 37% of 8-Ks in the processed sample have an attached press release. Panel B of the same table reports the sentiment measures for the subset of 8-Ks which also have press releases. Finally, Panel C of Table II tabulates the correlation matrix amongst the various measures of textual sentiment within the press release and also the 8-K. The correlation between positive and negative words is negatively sloped for the press release, but positively sloped for the 8-K, which may reflect a more cautious tonal hedging strategy in the 8-K which is more absent in the press release. The other main stylized fact is that the press release is on average more positively toned than the corresponding the 8-Ks, which is expected given the regulatory constraints firms face.

## 3.2 8-Ks Are Heterogenous

Table III provides richer characterization of sentiment across three different broad categories of 8-Ks. Panel A provides the breakdown of positive and negative word frequencies by whether a press release was also available with the 8-K for all 8-Ks associated with "Entry Into a Material Definitive Agreement". This category represents events where the

[26] A caveat is that there is also the possibility of a mechanically induced relationship through measurement error in the press release sentiment.

firm entered into an agreement outside the scope of its regular business activities and is not considered to have a bright line threshold.[27] The press release is noticeably more positive with greater positive words and fewer negative words than the corresponding 8-K. There also does not seem to be an obvious selection bias in the issuance of press releases with the observable 8-K sentiment similar between 8-Ks which do not have a press release and those that do. Panel B displays a similar breakdown for filings related to change of directors or principal officers. Here there does appear to be selection based on the underlying 8-K tone. 8-Ks without press releases are on net more negative than their corresponding counterparts with press releases. There is also a large positive difference between the press release sentiment relative to the 8-K, which is consistent with statistical evidence of quotes (15% of press releases contain at least one quote) of effusive praise lavished on the incoming or exiting director and or officer. Finally in Panel C, I take advantage of the fact that some 8-K category headings represent on average ex-ante bad news. These event categories include off balance sheet agreements, delisting from exchanges, change of auditors, restatements, and amendments to bylaws or change in fiscal year. In correspondence, Lerman and Livnat (2010) also identified these events as having on average negative abnormal filing returns. For this particular sub-sample, I observe small differences in the 8-K language with respect to the choice of a press release, but again large positive differences between the press release and the 8-K. Determining whether these differences reflect credible information or hype motivates the empirical test for market reaction.

## 3.3 Empirical Strategy and Results

To measure market reactions, the 8-K and press release data is merged with CRSP and Compustat. The original 8-K sample has a sample size of 1,046,487 and using both the WRDS CIK to GVKEY and Compustat CRSP link files, I was able to match 701,061 observations to CRSP and Compustat. Firm fundamentals such as book to market and industry classification are pulled from Compustat.[28] Return characteristics are taken from CRSP, and analyst coverage from IBES. Finally institutional ownership is calculated through Thomson Reuters's database of SEC 13-F filings. A complete set of variable descriptions and construction is available through Table I.

I make a number of restrictions to the sample used in tests of market reaction to sentiment. This study's focus on non-earnings related press releases means that I do not consider any 8-K which contains Item 2.02(Item 12) in the new (old) 8-K form which both refer to

[27]For example, unsigned letters of intent do not need to be reported because it can affect live negotiations. This issue was highlighted in the SEC Final Rule dated August 23, 2004.

[28]I use the Global Industry Classification System which has been shown to generate superior groupings relative to other classfication schemes (Bhojraj and Lee, 2002).

earnings results. To avoid confounding effects, 8-Ks filed on the same day as an earnings announcement, 10-Q, 10-K, or DEF14A (annual proxy statement) are also excluded. Multiple 8-Ks filed by the same firm on the same day are extremely rare, but also excluded. I also require firms to have available stock prices one month before and after the 8-K filing date. Finally the sample of filings starts on June 1996 and ends in December 2011. The choice of the starting date coincides with when the SEC forced all firms to use EDGAR to electronically submit their filings. In the prior period, large firms were selectively biased in their propensity to file via EDGAR. I also make a number of a small number of deletions to observations which are obvious errors e.g. the filing date is before the report date, the details of which are available in Appendix A.

### 3.3.1 Decision to Issue Press Release

I am unaware of previous work which analyzes the firm's voluntary decision to issue a related press release conditional on a mandatory SEC disclosure. The disclosure literature however, has produced a large amount of work showing that managers prefer to delay or obfuscate bad news if possible (Kothari *et al.*, 2009b). Therefore, I expect the press issuance decision to be related to the underlying 8-K tone, since the press release precedes the 8-K typically. Table A.I provides evidence of differential voluntary press releases rates. For example firms issue press releases concerning change of auditors (item 4.01) only 6.2% percent of the time. This finding is not surprising since spin is less likely to work for obviously negative categories where there is relatively more hard information, and firms may prefer the relative obscurity of the 8-K rather than draw attention through a press release. Pooling the ex-ante negative events described in Table III, the press release rate is only 16%, which is lower than the sample average of 37%, suggesting that firms in general prefer to not publicize bad news which is consistent with prior literature.

In order to more formally investigate the determinants of the voluntary press release for the full sample, I specify the following equation and estimate via logistic regression:

$$\text{Press Release}_{i,j} = \beta_1 \text{8-K}_{i,j} + \Pi X_{i,j} + \epsilon_{i,j} \tag{1}$$

where $\text{Press Release}_{i,j}$ is a dummy which equals one if firm $i$ in 8-K filing $j$ and also issued a press release. $X_{i,j}$ is a vector of covariates which include 8-K category codes, the sentiment of the 8-K text language, common firm fundamentals, as well as variables describing the firm's information environment. I also standardize all tonal measures such that reported coefficients represent the effects of a one standard deviation change in the underlying covariate.

Table IV reports results of Equation 1 with un-exponentiated coefficients. As expected in Column 1, a one standard deviation increase in the net 8-K sentiment increases the probability of a press release by 15.6%. I also find that more complicated 8-Ks (which contain more item numbers) increases the incidence of press releases but filings on Fridays are less likely to have a press release. The last point is also consistent with previous research of managers' propensity to dump negative earnings surprise on Fridays. Column 2 decomposes the effects of tone into its positive and negative components and finds that it is the degree of negative tone in the 8-K which drives the firm's issuance decision. Columns 3 and 4 report the same specifications but using a linear probability model with the inclusion of firm fixed effects and find qualitatively similar results.

### 3.3.2 What Explains The Sentiment Bias?

Conditional on the press release, I next investigate what determines the degree of tonal contrast between the press release and the 8-K. I specify the equation and estimate via least squares:

$$\text{Sentiment Bias}_{i,j} = \Pi X_{i,j} + \epsilon_{i,j} \tag{2}$$

where $\text{Sentiment Bias}_{i,j}$ is defined as the difference in sentiment between the press release and the 8-K for firm $i$ and filing $j$. $X_{i,j}$ is a vector of covariates which include 8-K category codes, characteristics of the press release such as whether there are quotes, and whether it lists an external contact person, as well the regular set of controls for the firm's fundamentals and information environment.

From Column 1 of Table V, I find that both the presence of an external contact person and the existence of quotes increases the tonal difference between the 8-K and press release. The latter is consistent with 8-Ks rarely having quotes due to SEC instructions to be succinct. The results are preserved if we restrict to the set of press releases which are released on the same day as the 8-K. Finally, Columns 3 and 4 repeat the specifications but include firm fixed effects with similar results. Both Bushee and Miller (2012) and Solomon (2012) provide evidence that public relations and investor relations may help bring additional coverage of firm news. While the decision to use an external contact person is certainly a choice variable, nevertheless, it is an interesting finding that they also appear to write more optimistic press releases relative to in-house counterparts, both cross sectionally and within firm. In connection with the managerial disclosure incentives literature, I also find that the sentiment bias is increasing in the propensity for insider trading, proxied by a Form 4 filing, in both the prior week and the subsequent week relative to the 8-K disclosure date and provides a

possible strategic motive for spinning.

While opportunistic insider trading is a natural explanation for why managers may engage in spin, other explanations are also possible. For example, overconfident managers may incorrectly estimate the effectiveness of spin because they do not correctly identify the marginal investor. Alternatively, rational expectation equilibria may exist where managers spin, but investors are not fooled, but managers do not deviate because it may signal something worse. Finally firms may rely on the doctrine of puffery to lower their expected litigation costs. Rogers and Van Buskirk (2011) show that firms with abnormally high optimism in earnings press releases are more likely to be sued than a control sample, but the puffery defense may change the litigation payoff distribution conditional on being sued.

### 3.3.3 Short Term Market Reaction

To motivate the empirical analysis, Dye and Sridhar (2004) model the existence of an unique price equilibrium where prices are a linear function of both soft and hard information, a concept which can be mapped to the press release and 8-K respectively. Treating the 8-K as hard information, I expect the market to react positively if the 8-K is positively toned, hence $\beta_1 > 0$. With respect to $\beta_2$, if the sentiment bias only included promotional cheap talk then efficient markets would predict $\beta_2 = 0$ in both the short and long term. Alternatively, if the sentiment bias contain credible soft information (Benabou and Laroque, 1992), then we would expect $\beta_2 > 0$ in the short term, and $\beta_2 = 0$ in the long term. If however the press release contains a mixture of soft information and spin, but investors have difficulty in distinguishing between the two, then we would also expect $\beta_2 > 0$ initially, but $\beta_2 < 0$ because investors overreacted and reformed expectations from subsequent new information or alternatively the lack of new information (Giglio and Shue, 2012), which discredits the firm's earlier spin.

Because press releases can precede the 8-K, I test for investors' initial response around the combined window surrounding the press release and the 8-K filing date. The event window is defined as 15 trading days prior to the 8-K, which is when I begin to observe press releases, and ends the day after the 8-K filing date. The dependent variable is therefore CAR[-15,1] centered around the 8-K filing date.[29] Cumulative abnormal returns are generated based on raw CRSP returns minus a value weighted portfolio matched on quintiles of both size and book to market[30]. The 25 portfolios are formed as described in Fama and French (1992) at the end of June every calendar year. To determine how investors respond to the qualitative

[29]My results are not sensitive to shortening the window to CAR[-4,1] which encompasses 90% of the press release sample.

[30]Both Barber and Lyon (1997) and Daniel and Titman (1997) provide evidence that an event study based on characteristics adjusted returns generates more powerful test statistics than a factor based approach, an approach which I adopt here.

information embedded in the press release and the 8-K, I specify and estimate the following via least squares:[31]

$$\text{CAR[-15,1]}_{i,j} = \beta_1 \text{8-K}_{i,j} + \beta_2 \text{Sentiment Bias}_{i,j} + \Pi X_{i,j} + \epsilon_{i,j} \tag{3}$$

$X_{i,j}$ is a vector of controls and contains: momentum[-12 months,-2 months], 5 quintiles of log market cap, log book to market, log number of analysts, institutional ownership, log share turnover, and a Nasdaq dummy. Furthermore all specifications include dummies for each of the 8-K categories, as well as filing year and industry dummies. Standard errors are clustered by firm and filing date unless otherwise reported (Petersen, 2009; Gow *et al.*, 2010).

Table VI produces regression results from estimation of Equation 3. Columns 1 estimates $\beta_1$ as 33 basis points for each standard deviation in the net 8-K sentiment. Column 2 reports the same regression but restricted to the sub-sample of 8-Ks with press releases and finds a higher $\beta_1$ of 59 basis points. Column 3 introduces the effect of the sentiment bias alongside with the 8-K sentiment which implicitly restricts the sample to 8-Ks which have a press release. The coefficient on the 8-K sentiment is 112 basis points, while the sentiment bias's effect is 88 basis points. In Column 4, I restrict the sample to press releases which were released concurrently on the same day as the 8-K which results in slightly larger coefficients relative to Column 3. In Columns 5 and 6, I include firm fixed effects in addition to the specification in Column 2. This specification removes concerns of correlated unobserved firm heterogeneity fixed over time. Column 5 reports results for the entire sample, while Column 6 again restricts to the sub-sample where the 8-K and press release were concurrent. The magnitudes are larger under fixed effects, which we expect given that firms will acquire reputation for being non-credible (Stocken, 2000) if they always include promotional language[32]. In every specification, the effect of the sentiment bias

[31]Note that estimates are numerically equivalent mappings between using the press release sentiment instead of the sentiment bias since they are linear combinations. For example the two regressions below are numerically equivalent:

$$\begin{aligned} y &= \beta_1 \text{8K Sentiment} + \beta_2 \text{Press Sentiment} \\ y &= \alpha_1 \text{8K Sentiment} + \alpha_2 (\text{Press Sentiment - 8K Sentiment}) \\ \beta_2 &= \alpha_2 \\ \beta_1 + \beta_2 &= \alpha_1 \end{aligned}$$

[32]Note that investors may still respond if they are not Bayesian and thus fail to update sufficiently. (Eyster and Rabin, 2009)

($\beta_2$) is positive. The magnitudes are economically large, indicating that firms are actually providing credible information through the 8-K, but also incrementally through the press release.

### 3.3.4 Asymmetry Between Positive and Negative Words

Much of the prior textual analysis literature uses either 1) negative words only or 2) net positive minus negative words as a measure of tone. Prior research justified the usage of the former case by the fact that they were unable to find any market reaction to positive words Tetlock (2007); Engelberg (2008). While the latter measure is intuitive to explain, it imposes an arbitrary equal weighting[33] scheme on the effects of positive and negative words, an assumption which is not justified if investors respond asymmetrically to bad news (Kothari *et al.*, 2009b). To provide clarity on this issue, I decompose both the net 8-K and net press release tone into its positive and negative components and re-estimate Equation 3, which is reported in Table VII.

Columns 1 estimates $\beta_1$ as 33 basis points for each standard deviation in the net 8-K sentiment. Column 2 decomposes the net 8-K sentiment into its positive and negative components and finds that markets react positively to positive tone and negatively to negative tone, but asymmetrically in magnitude. The effect of negative words is -33 basis points while positive words increase returns by 14 basis points. Column 3 introduces the effect of the press release sentiment alongside with the 8-K sentiment which implicitly restricts the sample to 8-Ks which have a press release. While the coefficient on the 8-K remains un-changed, the effect from the press sentiment is large, at 79 basis points per standard deviation. Analogous to Column 2, I separate the net measures of 8-K and press release sentiment into their positive and negative components in Column 4. The 8-K positive and negative word coefficients are of similar magnitude at 22 and -24 basis points respectively. The press release positive toned coefficient is 40 basis points while the negative tone coefficient is -68 basis points.

Starting in Column 5, I restrict the sample to press releases which were released concurrently on the same day as the 8-K which results in larger magnitudes for the coefficients relative to Column 4. In Column 6, I again show the decomposed results based on the sub-sample where the press release and 8-K were released concurrently. The coefficients are larger in magnitude relative to those in Column 4, which can be partially driven through lower information pre-leakage. Finally, in Columns 7 and 8, I include firm fixed effects in

[33] Loughran and McDonald (2011) proposed weighing words (term weighing) based on their overall frequency of usage within a document corpus. In a recent working paper, Jegadeesh and Wu (2011) also advocate using word eights optimized on past stock market response. The benefit of their approach is that they do not need to rely on ex-ante classifications of positive and negative words, but instead rely on the market response to infer both the direction and magnitude of each word.

addition to the specification in Column 6. Column 7 reports these results using the net sentiment measures, while Column 8 again decomposes the sentiment into their constituent parts with signs and magnitudes similar to those reported in the larger sample and without firm fixed effects.

Li (2008) raises the valid concern that firms may attempt to obfuscate bad news through longer documents. This concern means that the effects of tone may have greater power if they are unscaled by the word count. In the robustness section, I discuss my results which become stronger when I use the raw unscaled word counts of positive and negative words.

### 3.3.5 Long Term Market Reaction

In order to determine if markets overreacted to the press release but subsequently failed to de-bias themselves through the 8-K, I test for evidence of a negative price drift as a function of the original press release sentiment, after controlling for the 8-K sentiment as well. If overreaction occurred, then I expect $\beta_2 < 0$ when examining post 8-K announcement returns. If the market correctly distinguished information from spin, then I expect $\beta_2 = 0$.

I re-estimate Equation 3 but replace the dependent variable with the cumulative abnormal returns for the event window[2,20] trading days after the filing date. I also decompose the net tonal measurements into their positive and negative components. Table IX produces the regression results and Column 1 finds that a one standard deviation increase in positive words in the press release is associated with a -13 basis point drift. Column 2 restricts to the set of press releases which were filed earlier than the 8-K and finds the effect of positive words in the press release much greater at 20 basis points. Firms are unlikely to deviate strongly in language between the 8-K and the press release if they are timed concurrently as it would be obvious. Column 3 restricts the sample to press releases concurrently with 8-Ks in Column 3 and does not find evidence of negative drift. Columns 4-6 add firm fixed effects and find similar results, but with much larger magnitudes of 20 to 24 basis points.

In order to understand the time horizon over which most of the drift occurs, I re-estimate Equation 3 with different event windows spanning from 2 to 30 trading days after the 8-K filing date and plot the press release positive word coefficients from Column 2 of Table IX in Figure 7a for the set of press releases which preceded the 8-K. The figure indicates that a disproportional amount of the drift occurs between 5 to 10 trading days after the 8-K filing date which also coincides with the spike in the timing of insider trading (Form 4) in Figure 3.[34]

As a robustness check, in un-tabulated results restricting my analysis to the sub-sample of earnings related 8-Ks, identified by item 2.02(12) in the new(old) 8-K, I instead find

[34]While I do not currently investigate insider trading in detail, it may be interesting to investigate if insiders sell prior to the reversal, or if it is the insider trading(selling) itself which triggers the reversal.

positive price drift following positive press releases. This result is consistent with the findings of Demers and Vega (2011) and Davis *et al.* (2012) and reduces concerns about model mis-specification. However these results are still conditional on the firm's issuance of a press release[35] and thus I also separately analyze the set of 8-Ks which do not have press releases.

### 3.3.6 8-Ks Without Press Releases

To complete the analysis of the effects of tone on 8-Ks, I also produce results for the sub-samples of 8-Ks which do not have an associated press release. Panels A and B of Table X presents short term(-1 to 1 window around the filing date) and long term results (2 to 20 window after the filing date) respectively. An one standard deviation increase in the 8-K net sentiment generates a 10 basis point increase in returns in Column 1 of Panel A. Column 2 decomposes the sentiment into positive and negative components and finds the market reacts equally to both positive and negative tone, which is in contrast to the press release tone where negative tone dominated. Including firm firmed effects in Columns 3 and 4 slightly enlarges the size of the coefficients.

Panel B reports results of the same regression but with the cumulative abnormal returns over the next 2 to 20 trading days instead. For 8-Ks without press releases, I find positive drift in response to the 8-K sentiment, with a magnitude of 15 basis points per standard deviation. Again, decomposing the sentiment into separate components, Column 2 reports reaction to both positive and negative tone. Inclusion of firm fixed effects in Columns 3 and 4 again marginally improve the magnitude of the tonal effects as expected.

The magnitude of the drift, 15 basis points per standard deviation in tone, relative to the initial reaction, 10 basis points, suggests that investors significantly under-reacted to information in the class of 8-Ks without press releases. In order to better understand the time horizon over which most of the drift occurs, I re-estimate Equation 3 with different event windows spanning from 2 to 30 trading days after the 8-K filing date, and plot the press release positive word coefficients from Column 2 of Table IX in Figure 7b. Again, the graph indicates that a disproportional amount of the drift occurs between 5 to 10 trading days after the 8-K filing date.

The evidence of over-reaction to press releases and under-reaction to 8-Ks without press releases respectively indicates the possibility of investor inattention, which I investigate further in the next section through an analysis of EDGAR search traffic data.

[35] In my sample nearly 95% of earnings related 8-Ks have a press release so sample selection is less of a concern for prior literature's and also my results related to earnings related press releases.

# 4 Market Frictions

## 4.1 Investor Inattention

A large literature on investor inattention has generated evidence that investors do not react to or acquire all information in a timely manner[36]. Examples include underreaction to customer-supplier links (Cohen and Frazzini, 2008), underreaction during multiple earnings announcements (Hirshleifer *et al.*, 2009), and underreaction to bond rating downgrades (Dichev and Piotroski, 2001). In the specific context of SEC filings such as annual 10-Ks, researchers have also found evidence of investor under-reaction (You and Zhang, 2009). In general investor inattention relies on the simple premise that rational investors face non-trivial information acquisition costs, which also implies that the mode of dissemination may be relevant.

In the context of this study, one explanation for the market's overreaction to press releases is that some participants may not be fully aware of the 8-K and only read the press release, since otherwise they could potentially de-bias themselves through the 8-K.[37] Huberman and Regev (2001) document an extreme example where a biotech firm released important information through the journal Nature, and several months later the essentially same information appeared in the New York Times. Specialists who read the Nature article reacted promptly, but generalists reacted as well with a much larger magnitude when it became more broadly disseminated. While I acknowledge that other market frictions are certainly present, for example transaction costs or short sale constraints, inattention is not entirely unreasonable to consider since press releases rarely mention the availability of a contemporaneous or forthcoming 8-K filing.[38] To test whether investors substitute between the press release and the 8-K, I use novel click by click search traffic data of activities on the SEC EDGAR website itself for the years 2008-2011.

To briefly provide background of the underlying data, the SEC EDGAR website is located at *http://www.sec.gov/edgar.shtml*. From there an user can search for company filings through several ways to ultimately access the record of interest. Each time a particular filing on the EDGAR website is accessed, a log of that is recorded with the following attributes: the timestamp of the request, an unique ID which corresponds to the the internet protocol (IP) address of the visitor, the unique document id("accession number") that was requested,

[36] A large field in economics has found similar evidence that consumers are inattentive to non-salient information such as non-transparent taxes, shipping costs, and complex information. See DellaVigna (2009) for a review of the persuasion literature.

[37] Malmendier and Shanthikumar (2007b) find evidence that retail investors, in contrast to institutions, focus on recommendations relative to earnings forecasts, even when they are in conflict with each other.

[38] This was based on a search of 20 random press releases in my sample, I found only one press release which actually referred the reader to more details in the 8-K. The rest contained boilerplate language about general filings available at the SEC website.

and various user agent characteristics such as the web browser and operating system of the user if available. As an example, a hypothetical observation could be that at 7:45 AM EST on April 15th 2008, an user running Windows XP using Internet Explorer accessed Google's 2008 first quarter 10-Q filed on April 12th, 2008. Drake *et al.* (2012b) provide additional background and descriptive analysis using the data for an earlier subsample.

Both investors as well as automated scripts or web crawlers can access EDGAR. For example it would be possible that data vendors such as 10K Wizard, Factset or Bloomberg have robots which automatically patrol EDGAR through RSS feeds and download any new raw filing it encounters for delivery to its subscribers- these activities would also show up in the web traffic logs. Beyond data vendors, institutions and and other sophisticated investors are also more likely to be using automated crawlers to download filings relative to retail users simply because of the technical expertise required.

## 4.2 Empirical Strategy

Because retail and non-retail investors- who are more likely to appear machine-like are systematically different, I follow heuristics rules adopted in Lee *et al.* (2012) which separates the "humans" from the "robots". The filtering process essentially measures how fast and how many filings a particular unique visitor is downloading on a daily basis and uses a cutoff rule. Further refinements include determining the type of document being accessed, for example, certain filings are encoded in a machine-readable format, which a retail investor would unlikely access via a web browser.

Following these procedures, I am able to generate for each unique 8-K filing, the number of unique "human" visitors per calendar day[39], $\text{HumanTraffic}_{j,t}$ where $j$ references the filing and $t$ represents the number of trading days since the filing, where $t = 0$ is the day of the filing. I focus on unique users rather than total downloads because it is less likely to be skewed by extreme users and also a natural measure of the breadth of interest. Finally, I interpret the flow of traffic from these identified users as closer to retail because institutional investors have available to them other commercial products which provide the same information.

To understand how a competing press release influences the degree of search on EDGAR, I estimate the following specification:

$$\text{Log}[\text{HumanTraffic}_{j,0}] = \beta_1 \text{Press}_{j,0} + \beta_2 \text{8-K Sentiment}_{j,0} + + \Pi X_{j,0} + \epsilon_{j,0} \tag{4}$$

[39]There is a small concern that I measure traffic via calendar day when the the 8-K filing is intraday. These concerns are alleviated with the inclusion of 8-K filing time hourly fixed effects which will remove the mechanical effects of truncation for 8-Ks which are released later in the day. My results also remain if I enlarge the traffic window to include the following calendar day.

where HumanTraffic$_{j,0}$ is the number of unique visitors for filing $j$ on the actual date the 8-K was first available. $X_{j,0}$ is a vector of covariates which include 8-K category codes, hourly fixed effects based on the actual filing time of the 8-K, as well as the regular set of controls for the firm's fundamentals and information environment. If investors treat the press release and 8-K as complementary signals then we should expect $\beta_1 > 0$. On the other hand, search costs may dominate and investors may perceive them as substitutes, which would lead to $\beta_1 < 0$.

Before reporting regression results, Figure 8a provides graphical evidence of lower search traffic for both 8-K specific traffic and also for any SEC filing traffic when a competing press release exists. The lower search traffic is evident on the day of the filing and persists over time. Figure 8b plots 8-K traffic according to whether a press release exists, and also by the actual tone of the underlying 8-K. The indicator for good news is simply defined as sentiment equal to or above the median 8-K sentiment in the sample. Interestingly, when news is negative, we see little graphical difference between having or not having a press release. In contrast, the difference in search seems largely driven by lack of search when the news good is and the existence of the a press release.

Table XI reports regression results based on Equation 4. Column 1 reports the baseline specification with full control and finds a -49% decrease in 8-K traffic in response to the presence of a press release, even after controlling for the 8-K's sentiment. The coefficient on the 8-K net sentiment, $\beta_2$, is also negative with a one standard deviation increase in tone decreasing traffic by 11%. Filings released on Fridays also experience low initial search traffic, with a 38% decrease, which is consistent with the inattention hypothesis from Dellavigna and Pollet (2009) and also serves as a benchmark for the magnitude of the effect of the press release variable. Column 2 adds an interaction term between the press release and the 8-K sentiment to statistically test the relationship observed in Figure 8 where the effect of press release seems dependent on the underlying 8-K tone. The interaction is negative, which is consistent with the graph, but not statistically significant. The point estimate is also small, suggesting only a 3.7% incremental decline for each standard deviation increase in the tone between press release and non-press release related 8-Ks. Column 3 restricts the sample to the set of 8-Ks for which the filing date equals the reporting date with the coefficient $\beta_1$, much larger at -78%. Finally to control for unobserved heterogeneity within firms fixed over time, firm fixed effects are added in Columns 4 to 6, but otherwise mirroring the specifications from Columns 1 to 3. The coefficient on press release, $\beta_1$ is negative in every column and large, between 34-54% with firm fixed effects.

With respect to the regression specification, since my dependent variable is a count measure, a negative binomial specification may be more appropriate.[40] Re-estimating the

[40] A Poission model is also possible but imposes the equality of mean and variance which is violated in

equation via the alternative specification, in un-tabulated results, I obtain qualitatively similar results, namely a statistically significant negative $\beta_1$.

More broadly, these results are also consistent with a sequential search cost model (Stahl, 1989) where investors first read the press release and contingent on that information, decide whether to search further for additional information since investors do not ex-ante know the contents of the 8-K unless they search. Note that my findings do not suggest investors are irrational since search is obviously costly.[41] As an additional robustness check (untabulated), I also re-estimate the equation but replace the dependent variable with overall search traffic for the firm (which subsumes 8-K specific search traffic) on the day of the 8-K filing and obtain similar results albeit with smaller coefficients.

In terms of the generalizability of these results, it is certainly true that EDGAR is not the only source where investors can retrieve 8-Ks. Many other services download and rebroadcast the same information. However this concern should not create a bias for my estimated results unless the availability of the press release itself is systematically correlated with the investor's propensity to search on other websites for 8-Ks instead of EDGAR.

# 5 Robustness Discussion And Additional Results

## 5.1 Sample Selection of Press Releases

In order to understand potential biases which may be associated with the firm's decision to issue a press release, but fail to attach it as an exhibit in the 8-K, the following paragraph describes the SEC's filing instructions with respect to the 8-K[42]:

> If the registrant makes available to its stockholders or otherwise publishes, within the period prescribed for filing the report, a press release or other document or statement containing information meeting some or all of the requirements of this form, the information called for may be incorporated by reference to such published document or statement, in answer or partial answer to any item or items of this form, provided copies thereof are filed as an exhibit to the report on this form.

Based on the instructions, firms appear to have the option, but not the obligation to provide press releases if they materially relate to the 8-K. A concern arises if firms systemat-

the data. The negative binomial specification is a generalization which allows for over-dispersion which is consistent with the data-the variance of traffic is larger than the mean.

[41] The other general caveat in interpretation these results is that the issuance of the press release is non-random, and I may be omitting important covariates which are influencing the decision to issue a press release.

[42] Form OMB Number: 3235-0060, Expires: April 30, 2015.

ically avoid embedding related press releases. Alternatively, a benign scenario could be that 8-Ks are prepared by the general counsel whereas press releases are generally assigned to the public relations department. The failure to bundle the appropriate press release may reflect mis-communications between the two departments rather than a more strategic motive. If however the latter case dominates, then the direction of the bias would depend on the model which describes the firm's incentives to issue a press release, but subsequently fail to bundle it as an attachment in the 8-K and the associated empirical question being asked.

In order to examine the potential magnitude and bias of unreported press releases, I collect all firm issued press releases through NewsBank, which is a large data repository for media content for a random sample of 50 firms between years 2006 and 2011. In accordance with Bushee and Miller (2012), I use PR Newswire, Market Wire, Business Wire sourced articles to proxy for firm issued press releases; the limitation on the starting period stems from Newsbank's data availability. More details concerning the procedure and the actual list of firms are available in Appendix B.

Figure A.1 plots the daily frequency of any firm issued press releases as a function of calendar time relative to the 8-K filing date, which is denoted at time=0. The cyclicality in the figure arises from the inclusion of weekends. The plot is separately generated for the sample of 8-Ks which have or do not have an attached press release.

If firm systematically fail to attach actual press releases around the filing date, then it might be possible to detect via an examination of the time series variation of press releases. As expected, frequency of press releases mechanically spike at time 0 for the set of firms which release 8-Ks with attached press releases. There is also a small bump in the frequency of press releases for the sample of 8-Ks which did not report a press release, suggesting that some press releases may have gone unreported, or alteratively (but less likely) firms issue other unrelated press releases.[43] In general however, the frequency of press releases for the attached and unattached sample are similar, and excluding the 8-K filing date at time zero, I fail to reject the null of no difference in frequency of press releases between the attached and non-attached sample with a *t-stat* of 1.02.

The test for sample selection is admittedly limited since I am only exploring time series variation in total newswires as a form of non-random selection. However it does alleviate concerns that firms disproportionally issue but fail to bundle press releases pertaining to the 8-K prior or after the filing date, if we believe that the frequency of overall press releases proxies for 8-K related press releases. This strategy however does not address selection in terms of the underlying contents of the press releases. For example, it is possible that 8-Ks which do not bundle press releases do in fact have press releases which are systematically

[43]It may be possible to provide an upper bound on the magnitude of the under-reporting, but this will require additional assumptions.

different from attached press releases along the dimensions of tone. In order to examine the actual contents of the press releases to determine selection bias, it would be necessary to devise a matching engine which semantically and contextually links 8-Ks to press releases.

To ascertain whether firms issue press releases after the 8-K, in lieu of the matching engine, I focus on a set of 8-K disclosures which I can more plausibly link to press releases without advanced contextual matching techniques. Specifically I examine director resignations which is a subset of Item 5.02. Using the name of the board of director who resigned plus the company name, I manually search for press releases around the filing date of the 8-K for a random sample of 20 Item 5.02 8-Ks which do not have an attached press release. I fail to find any press release related to the resignation either before or after the 8-K filing date.[44]

As another robustness check, I re-examine the differences in tone across the entire Newsbank sample of newswires related to 8-Ks with and without attached press releases. I do not find any difference in the average net tone in the sample of all newswires either before or after the 8-K filing date.

## 5.2 2008 Regulation FD Interpretative Guidance

After August 2008, the SEC issued an interpretative guidance which allowed firms to disseminate Regulation Fair disclosure (Reg FD) related information via their websites in lieu of filing an 8-K based on certain conditions.[45] This guidance might differentially influence the availability of the press release as an attached exhibit for Reg FD related disclosures, which are coded in the 8-K category as Item 7.01(9) in the new(old) form. As a robustness check, in un-tabulated results, I condition on filings prior to the interpretive guidance, and my tests of market reaction do not change qualitatively.

## 5.3 The Effect of Sentiment May be Nonlinear

Tonal measurements are synthetic and atheoretical, and it is possible that their effects on returns behave nonlinearly. While this possibility does not alter the general conclusions of this paper, I explore this possibility and perform market response tests using quintiles of the 8-K and press release sentiment instead of their continuous counterparts. In un-tabulated results, I find that returns are monotonically increasing in the quintiles of the textual sentiments, which give credence for the linearly specified models.

[44] There was one case of a Dow Jones Newswire alert which is a premium service that rebroadcasts the 8-K filing as an alert to subscribers on the same day as the 8-K filing. Li *et al.* (2011)

[45] [Release Nos. 34-58288, IC-28351; File No. S7-23-08] Date: August 7, 2008

### 5.4 Measurement of Information in 8-K and Press Release

One potential measurement concern is that the 8-K itself reflects no additional information but refers the reader to the press release. There are several approaches in response to this concern- one could remove these 8-Ks by conditioning on the length of the 8-K disclosure since they would be necessarily short by construction. Another more general approach to identify "empty" 8-Ks would be to determine if the textual analysis recognizes any words as positive or negative. If the 8-K only contains information such as "refer to press release dated Jan 1, 2011" then these 8-Ks would by construction be both short and also devoid of any positive or negative tonal words. In lieu of filtering these 8-Ks from the sample entirely, one could also mechanically assign the press release sentiment to equal the 8-K sentiment since by the firm's choice, they are equal to each other. Table VIII reports the results of each of these specifications. Column 1 represents the baseline specification. If we restrict to 8-Ks which have non-zero tone (and therefore are selected to contain information), we find higher 8-K and lower press release coefficients. In Column 3, I restrict to the sample of 8-Ks which contain at least 100 words, which again should select out empty references 8-Ks and find nearly identical coefficients to Column 2. Finally in Column 4, I identify the set of 8-Ks which did not contain tone and were excluded in Column 2 and code them as having the same sentiment as the press release. This specification again does not change the coefficients materially. The robustness results confirm that investors reacted to incremental information in the press release absent in the 8-K.

### 5.5 Scaling of Textual Sentiment Measure

Li (2008) raises the concern that managers may use more words to explain or obfuscate bad news. Therefore scaling by word count may cause underestimation of bad news and overestimation of good news. It is also unclear if scaling by the word count is appropriate when analyzing a wide variety of disparate disclosures such as 8-Ks, as opposed to a single class of relatively more homogenous documents such as earnings press releases. As a robustness check, I also use either the raw word count of positive and negative words, or quintiles of the raw word counts. The implied magnitudes of the effect of tone are similar compared to estimates using the scaled word count measures, but standard errors are actually estimated more precisely.

## 6 Conclusion

This study examines whether investors correctly distinguish qualitative information from promotional language in press releases related to material events of US public firms. It offers

new evidence of the magnitude of investors' response to qualitative information released through the 8-K. It also finds evidence that press releases contain incremental information above and beyond the contents of the 8-K. The initial positive reaction but subsequent negative drift after the 8-K disclosure in response to the tonal difference between the press release and the 8-K suggests that investors did not fully distinguish qualitative information from promotional language.

Identifying a channel through which investors may have overreacted to information in the press release, but subsequently failed to de-bias themselves through the 8-K, I use search traffic micro-data from the SEC EDGAR website, and find lower 8-K search intensity in the presence of a press release. Substitution behavior between the press release and the 8-K is consistent with a sequential search model in which search costs may have contributed to investors' over-reliance on the press release. This finding is important on its own because we have limited knowledge of investors' information acquisition behavior, especially in response to a biased information signal (DellaVigna and Gentzkow, 2009).

This work relates to active research concerning the role of the media as an information intermediary for investors (Bushee *et al.*, 2010). Much of this literature treats the underlying "news" as the firm-issued press release, and analyzes how the media (selectively) repackages this underlying news and the subsequent consequences of such repackaging (Dyck and Zingales, 2003). My study provides caution that firm-generated bias already exists, even prior to any subsequent media bias. Whether the media or other information intermediaries de-bias firm news or simply serves as a conduit for retransmission for non-routine disclosures such as the 8-K is an area for future research.

The paper contributes to our understanding of the complementarity and substitutability between mandatory and voluntary disclosures in a setting largely dominated by heterogeneously qualitative information. The apparent tradeoff investors (and consumers in general) face between information bias and search costs (Stigler, 1961), and whether investors search optimally, are other potential topics for future research.

# References


Antweiler, W. and Frank, M. Z. (2004). Is all that talk just noise? the information content of internet stock message boards. *The Journal of Finance*, **59** (3), 1259–1294.

Barber, B. M. and Lyon, J. D. (1997). Detecting long-run abnormal stock returns: The empirical power and specification of test statistics. *Journal of Financial Economics*, **43** (3), 341–372.

Benabou, R. and Laroque, G. (1992). Using privileged information to manipulate markets: Insiders, gurus, and credibility. *The Quarterly Journal of Economics*, **107** (3), 921–958.

Beyer, A., Cohen, D. A., Lys, T. Z. and Walther, B. R. (2010). The financial reporting environment: Review of the recent literature. *Journal of Accounting and Economics*, **50** (2), 296–343.

Bhojraj, S. and Lee, C. (2002). Who is my peer? a valuation-based approach to the selection of comparable firms. *Journal of Accounting Research*, **40** (2), 407–439.

Bradshaw, M. T. and Sloan, R. G. (2002). GAAP versus the street: An empirical assessment of two alternative definitions of earnings. *Journal of Accounting Research*, **40** (1), 4166.

Bushee, B. J., Core, J. E., Guay, W. and Hamm, S. J. (2010). The role of the business press as an information intermediary. *Journal of Accounting Research*, **48** (1), 1–19.

— and Miller, G. S. (2012). Investor relations, firm visibility, and investor following. *The Accounting Review*, **87** (3), 867–897.

Cain, D., Loewenstein, G. and Moore, D. (2005). The dirt on coming clean: Perverse effects of disclosing conflicts of interest. *The Journal of Legal Studies*, **34** (1), 1–25, ArticleType: research-article / Full publication date: January 2005 / Copyright 2005 The University of Chicago.

Carter, M. E. and Soo, B. S. (1999). The relevance of form 8-k reports. *Journal of Accounting Research*, **37** (1), 119–132.

Cheng, Q. and Lo, K. (2006). Insider trading and voluntary disclosures. *Journal of Accounting Research*, **44** (5), 815–848.

Cohen, L. and Frazzini, A. (2008). Economic links and predictable returns. *The Journal of Finance*, **63** (4), 1977–2011.

Crawford, V. P. and Sobel, J. (1982). Strategic information transmission. *Econometrica: Journal of the Econometric Society*, p. 14311451.

Dambra, M., Wasley, C. E. and Wu, J. S. (2012). Soft-talk management cash flow forecasts: Bias, quality, and stock price effects*. *Contemporary Accounting Research*.

Daniel, K. and Titman, S. (1997). Evidence on the characteristics of cross sectional variation in stock returns. *The Journal of Finance*, **52** (1), 1–33.

Davis, A. K., Piger, J. M. and Sedor, L. M. (2012). Beyond the numbers: Measuring the information content of earnings press release language. *Contemporary Accounting Research*, **29** (3), 845–868.

DellaVigna, S. (2009). Psychology and economics: Evidence from the field. *Journal of Economic Literature*.

DellaVigna, S. and Gentzkow, M. (2009). *Persuasion: Empirical Evidence*. Tech. rep., National Bureau of Economic Research.

Dellavigna, S. and Pollet, J. M. (2009). Investor inattention and friday earnings announcements. *The Journal of Finance*, **64** (2), 709–749.

Demers, E. A. and Vega, C. (2011). Linguistic tone in earnings announcements: News or noise? *SSRN eLibrary*.

Dichev, I. D. and Piotroski, J. D. (2001). The long-run stock returns following bond ratings changes. *The Journal of Finance*, **56** (1), 173–203.

Drake, M., Roulstone, D. and Thornock, J. (2012a). The informativeness of stale financial disclosures. *Available at SSRN 2083812*.

Drake, M. S., Roulstone, D. T. and Thornock, J. R. (2012b). What investors want: Evidence from investors use of the EDGAR database. *SSRN eLibrary*.

Dyck, A. and Zingales, L. (2003). The media and asset prices. *Unpublished http://www. nber. org/ confer/2003/si2003/papers/cf/dyck. pdf*.

Dye, R. A. and Sridhar, S. S. (2004). Reliability-relevance trade-offs and the efficiency of aggregation. *Journal of Accounting Research*, **42** (1), 51–88.

Engelberg, J. (2008). Costly information processing: Evidence from earnings announcements. In *AFA 2009 San Francisco Meetings Paper*.

Eyster, E. and Rabin, M. (2009). Rational and naive herding.

Fama, E. F. and French, K. R. (1992). The cross-section of expected stock returns. *the Journal of Finance*, **47** (2), 427–465.

Fang, L. and Peress, J. (2009). Media coverage and the cross-section of stock returns. *The Journal of Finance*, **64** (5), 2023–2052.

Francis, J., Schipper, K. and Vincent, L. (2002). Expanded disclosures and the increased usefulness of earnings announcements. *The Accounting Review*, **77** (3), 515–546.

Frederickson, J. R. and Miller, J. S. (2004). The effects of pro forma earnings disclosures on analysts' and nonprofessional investors' equity valuation judgments. *The Accounting Review*, **79** (3), 667–686.

Gandy, O. H. (1982). *Beyond agenda setting: Information subsidies and public policy*. Ablex Publishing Company Norwood, NJ.

Gentzkow, M. and Shapiro, J. M. (2006). *What drives media slant? Evidence from US daily newspapers*. Tech. rep., National Bureau of Economic Research.

Giglio, S. and Shue, K. (2012). No news is news: Do markets underreact to nothing? *SSRN Electronic Journal*.

Gow, I. D., Ormazabal, G. and Taylor, D. J. (2010). Correcting for cross-sectional and time-series dependence in accounting research. *The Accounting Review*, **85** (2), 483–512.

Gurun, U. G. and Butler, A. W. (2012). Don't believe the hype: Local media slant, local advertising, and firm value. *The Journal of Finance*, **67** (2), 561–598.

Hand, J. R. M. (1990). A test of the extended functional fixation hypothesis. *The Accounting Review*, **65** (4), 740–763.

Henry, E. (2008). Are investors influenced by how earnings press releases are written? *Journal of Business Communication*, **45** (4), 363–407.

Hirshleifer, D., Hou, K., Teoh, S. H. and Zhang, Y. (2004). Do investors overvalue firms with bloated balance sheets? *Journal of Accounting and Economics*, **38**, 297–331.

—, Lim, S. S. and Teoh, S. H. (2009). Driven to distraction: Extraneous events and underreaction to earnings news. *The Journal of Finance*, **64** (5), 2289–2325.

Hoskin, R. E., Hughes, J. S. and Ricks, W. E. (1986). Evidence on the incremental information content of additional firm disclosures made concurrently with earnings. *Journal of Accounting Research*, **24**, 1–32, ArticleType: research-article / Issue Title: Studies on Alternative Measures of Accounting Income / Full publication date: 1986 / Copyright 1986 Accounting Research Center, Booth School of Business, University of Chicago.

Huberman, G. and Regev, T. (2001). Contagious speculation and a cure for cancer: A nonevent that made stock prices soar. *The Journal of Finance*, **56** (1), 387–396.

Hutton, A. P., Miller, G. S. and Skinner, D. J. (2003). The role of supplementary statements with management earnings forecasts. *Journal of Accounting Research*, **41** (5), 867–890.

Jagolinzer, A. D. (2009). SEC rule 10b5-1 and insiders' strategic trade. *Management Science*, **55** (2), 224–239.

Jegadeesh, N. and Wu, D. (2011). Word power: A new approach for content analysis. In *AFA 2012 Chicago Meetings Paper*.

Karpoff, J. M., Koester, A., Lee, D. S. and Martin, G. S. (2012). A critical analysis of databases used in financial misconduct research. *SSRN Electronic Journal*.

Kogan, S., Levin, D., Routledge, B. R., Sagi, J. S. and Smith, N. A. (2009). Predicting risk from financial reports with regression. In *Proceedings of Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, NAACL '09, Stroudsburg, PA, USA: Association for Computational Linguistics, p. 272280.

Kothari, S. P., Li, X. and Short, J. E. (2009a). The effect of disclosures by management, analysts, and business press on cost of capital, return volatility, and analyst forecasts: A study using content analysis. *The Accounting Review*, **84** (5), 1639–1670.

—, Shu, S. and Wysocki, P. D. (2009b). Do managers withhold bad news? *Journal of Accounting Research*, **47** (1), 241–276.

Lang, M. H. and Lundholm, R. J. (2000). Voluntary disclosure and equity offerings: Reducing information asymmetry or hyping the stock?*. *Contemporary Accounting Research*, **17** (4), 623–662.

Larcker, D. F. and Zakolyukina, A. A. (2012). Detecting deceptive discussions in conference calls. *Journal of Accounting Research*, **50** (2), 495–540.

Lee, C., Ma, P. and Wang, C. (2012). Identifying peer firms: Evidence from edgar search traffic.

Lerman, A. and Livnat, J. (2010). The new form 8-k disclosures. *Review of Accounting Studies*, **15** (4), 752–778.

Li, E. X., Ramesh, K. and Shen, M. (2011). The role of newswires in screening and disseminating value-relevant information in periodic SEC reports. *The Accounting Review*, **86** (2), 669–701.

Li, F. (2008). Annual report readability, current earnings, and earnings persistence. *Journal of Accounting and economics*, **45** (2), 221–247.

— (2010). The information content of forward-looking statements in corporate FilingsA nave bayesian machine learning approach. *Journal of Accounting Research*, **48** (5), 1049–1102.

Loughran, T. and McDonald, B. (2011). When is a liability not a liability? textual analysis, dictionaries, and 10-ks. *The Journal of Finance*, **66** (1), 35–65.

Malmendier, U. and Shanthikumar, D. (2007a). Are small investors naive about incentives? *Journal of Financial Economics*, **85** (2), 457–489.

— and Shanthikumar, D. M. (2007b). *Do security analysts speak in two tongues?* National Bureau of Economic Research Cambridge, Mass., USA.

McNichols, M. and Manegold, J. G. (1983). The effect of the information environment on the relationship between financial disclosure and security price variability. *Journal of Accounting and Economics*, **5**, 49–74.

Padfield, S. J. (2007). Is puffery material to investors-maybe we should ask them. *U. Pa. J. Bus. & Emp. L.*, **10**, 339.

Petersen, M. A. (2009). Estimating standard errors in finance panel data sets: Comparing approaches. *Review of Financial Studies*, **22** (1), 435–480.

Rogers, J. L. and Van Buskirk, A. (2011). Bundled forecasts in empirical accounting research. *Journal of Accounting & Economics (JAE), Forthcoming*.

—, — and Zechman, S. L. C. (2011). Disclosure tone and shareholder litigation. *SSRN Electronic Journal*.

Solomon, D. H. (2012). Selective publicity and stock prices. *The Journal of Finance*, **67** (2), 599–638.

Stahl, D. O. (1989). Oligopolistic pricing with sequential consumer search. *The American Economic Review*, pp. 700–712.

Stigler, G. J. (1961). The economics of information. *The journal of political economy*, pp. 213–225.

Stocken, P. C. (2000). Credibility of voluntary disclosure. *The RAND Journal of Economics*, pp. 359–374.

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

— (2011). All the news that's fit to reprint: Do investors react to stale information? *Review of Financial Studies*, **24** (5), 1481–1512.

—, Saar-Tsechansky, M. and Macskassy, S. (2008). More than words: Quantifying language to measure firms' fundamentals. *The Journal of Finance*, **63** (3), 1437–1467.

Turk, J. V. S. (1986). Public relations influence on the news. *Newspaper Research Journal*, **7** (4), 15–27.

Wysocki, P. D. (1998). Cheap talk on the web: The determinants of postings on stock message boards. *SSRN eLibrary*.

You, H. and Zhang, X.-j. (2009). Financial reporting complexity and investor underreaction to 10-k information. *Review of Accounting Studies*, **14** (4), 559–586.

# Appendix A: Textual Analysis

## 6.1 Sentiment Construction

The universe of 8-Ks from 1994-2011 is downloaded from SEC EDGAR excluding amendments. A script extracts the reported 8-K categories (called item numbers) from the raw text of each filing. This gives different results from commercially available datasets such as WRDS because they only parse the index page for item numbers of the filing, which has a tendency to misreport the actual item numbers prior to 2004.

In order to identify whether the 8-K contained a press release first a text keyword search for the words "press release" or "news release" is conducted. Note that this will overwhelmingly generate type 1[46] relative to type 2 errors which is preferable given the research design. The lack of a text match indicates high certainty that no press release exists[47]. If there is no match, I mark the 8-K as not having a press release.

An alternative method is to search for string patterns of "Exhibit 99" which is the additional exhibit that nearly all press releases are filed under according to Regulation S-K. While this will reduce the incidence of type 1 errors, it will also increase type 2 errors because firms sometimes embed the text of the press release within the 8-K instead of following the instructions to file the press release separately as an Exhibit 99. Thus I adopt the former approach.

Next if a match occurs I then look for the date of the press release which is done through searching for particular patterns such as "press release ... dated [datestring]" or "news release.... on...[datestring]". If a pattern matches, I save the datestring[48] associated with the press release. The determination of the search patterns was an iterative process whereby I would go back and check failed matches, which enabled me to add new matching patterns if possible. Multiple validly tests are concurrently done to reduce the possibility of a mismatch and I also retain information when the parser is unable to make a definitive match because of multiple dates, for example because the 8-K referenced two different press release attachments with different dates, which are rare and I code them as missing.

In total, I match 91% of 8-Ks which had a positive hit on the keyword trigger to a press release date. Including type 1 errors, other reasons for a failed match include the firm attaching the press release in pdf format instead of a text file, which I did not parse.

I delete a small number of observations for which there seems to be obvious reporting errors. This includes incidences of when the filing date was earlier than the reporting date. To alleviate measurement error, I also code as missing any press release date that preceded the filing date by 30 calendar days. Filings which contain zero word count for either the press release or the 8-K are also removed (900 observations).

Next to generate the actual press release text and 8-K text for suitable analysis, I download the complete submission file of each 8-K. I then remove all HTML tags and all information embedded within tables. To extract the 8-K text language, I truncate the text of the 8-K to be between the start of the first item number and the required signature at the bottom of the document which removes large amounts of boilerplate language. This truncated text is referred to as the 8-K text in this paper.

In order to extract the press release text, I start with the same methodology for identifying press releases. In addition, I look for string patterns similar to "exhibit-99... press release" which signify the start of the press release. The end of the press release signifies the closing of the text document that the press release is embedded within. Finally, as an additional step, I look for string patterns of the general[49] form "forward looking statements" or "safe harbor statement" and remove those from the analysis whenever possible. A manual random check also reveals that this procedure will remove most of the company descriptions which are towards the bottom of the press release below the forward looking and safe harbor statements. The resulting truncated text serve as the basis for the press release text described in this paper.

Finally, as described in the main text, the two cleaned texts are tokenized into two sets of words which are matched against the word lists provided by Loughran and McDonald (2011).

[46] For example the 8-K could reference long-dated historical press releases. In addition, attached exhibits could also reference other press releases.

[47] With the caveat that post 2008: firms could theoretically release the press release through their company website given certain conditions.

[48] The date string itself can take on many formats as well which I flexibly capture.

[49] I allow for hyphens and other irregular characters within a certain character range to flexibly capture variants.

## 6.2 External Contact Person Identification

The following example can be viewed[50] at the SEC.

> EX-99.1 2 dex991.htm PRESS RELEASE
> Exhibit 99.1
> FOR IMMEDIATE RELEASE
> Contact:
> Patrick Spratt
> KVH Industries
> 401-847-3327
>
> Eric Boyriven
> Financial Dynamics
> 212-850-5600

The firm phone number is listed as 401-847-3327 (extracted separately from the title page) in the 8-K. Note that the second contact number does not match the firm's phone and so we conclude that the firm is using an external contact person for this filing. An internet search for "Financial Dynamics" reveals it to be the strategic communications group of FTI Consulting.

[50] *http://sec.gov/Archives/edgar/data//1007587/0001193125-05-204635.txt-index.htm*:

# Appendix B: Sample of All Press Releases

I take a random[51] sample of 50 firms from the S&P 500 index and search for all press wires related to them through News Bank, a news archive database for the sample period[52] 2006 to 2011. In accordance with Bushee and Miller (2012), I search for press wires from the three major newswires (PR Wire, Market Wire, and Business Wire). The names supplied to the search engine was sourced from Compustat firm names and standardized to remove common suffixes (CO, INC, etc) but may not have been the most commonly used name, which explains why some firm names returned low (some firms returned 0 and were not included) search counts. As a final restriction, I also check if the article headlines mention the firm name (same keyword used in the search engine), otherwise I do not count the press release (since I am using the firm's key word and not an unique firm identifier).

**Random Sample of Firm Issued Press Releases: 2006 to 2011**

| Company Name | Number of Newswires |
|---|---|
| 3M CO | 847 |
| ABBOTT LABORATORIES | 225 |
| AMERICAN ELECTRIC POWER CO | 544 |
| BAKER HUGHES INC | 362 |
| BRISTOL-MYERS SQUIBB CO | 156 |
| BROWN SHOE CO INC | 138 |
| CAMPBELL SOUP CO | 188 |
| CHEVRON CORP | 871 |
| CONOCOPHILLIPS | 891 |
| CROWN HOLDINGS INC | 123 |
| CVS CAREMARK CORP | 735 |
| DANA HOLDING CORP | 640 |
| DOW CHEMICAL | 1110 |
| DTE ENERGY CO | 391 |
| EATON CORP | 978 |
| EDISON INTERNATIONAL | 175 |
| ENTERGY CORP | 351 |
| EXELON CORP | 333 |
| EXXON MOBIL CORP | 393 |
| FMC CORP | 505 |
| HALLIBURTON CO | 524 |
| HEINZ (H J) CO | 20 |
| HERSHEY CO | 321 |
| HONEYWELL INTERNATIONAL INC | 180 |
| INTL BUSINESS MACHINES CORP | 1442 |
| KIMBERLY-CLARK CORP | 63 |
| MCGRAW-HILL COMPANIES | 32 |
| MERCK & CO | 788 |
| NAVISTAR INTERNATIONAL CORP | 166 |
| NVIDIA CORP | 582 |
| ONEOK INC | 298 |
| PENNEY (J C) CO | 85 |
| PEPSICO INC | 695 |
| PFIZER INC | 1028 |
| PITNEY BOWES INC | 598 |
| RAYTHEON CO | 1108 |
| ROCKWELL AUTOMATION | 226 |
| SEMPRA ENERGY | 205 |
| TEXAS INSTRUMENTS INC | 736 |
| TIMKEN CO | 322 |
| UNION PACIFIC CORP | 363 |
| UNITED TECHNOLOGIES CORP | 480 |
| WHIRLPOOL CORP | 374 |
| XCEL ENERGY INC | 302 |
| XEROX CORP | 673 |

[51] The sample of firms was randomly selected after restricting the sample to firms with at most two unique words, e.g. Exxon Mobil Corp. This choice was meant to minimize the occurrence of irrelevant results since the news search engine is based on keywords and not unique firm identifiers.

[52] The start date was chosen due to NewsBank's data coverage. Prior to 2006, there was insufficient coverage of all newswires.

# Table I.
# Variable Description

This table reports the construction of variables used in the regressions and their data sources.

| Variable | Description | Source(Year) |
|---|---|---|
| **8-K Filing Characteristics** | | SEC EDGAR |
| $PR_j$ | dummy=1 if the keyword "press release" is found in the main body of the 8-K text. | 1994-2011. |
| $8Kwords_j$ | Number of words in the 8-K text of filing $j$ excluding exhibits and attachments and boilerplate language. | 1994-2011. |
| $PRwords_j$ | Number of words in the press release text of filing $j$ | 1994-2011. |
| $8KSentiment_j$ | Positive-negative word count according to Loughran and McDonald (2011) scaled by word count for the 8-K text only. | 1994-2011. |
| $PRSentiment_j$ | Positive-negative word count according to Loughran and McDonald (2011) scaled by word count for the press release text only. | 1994-2011. |
| $ReportingDate_j$ | The event date that triggered the 8-K filing $j$. | 1994-2011. |
| $FilingDate_j$ | The date EDGAR received the 8-K filing $j$. | 1994-2011. |
| $PressReleaseDate_j$ | The event date that triggered the 8-K filing $j$. | 1994-2011. |
| $NumberofItems_j$ | Number of unique items on the 8-K form | 1994-2011. |
| $8KTimestamp_j$ | Timestamp EDGAR disseminated the 8-K. | 2004-2011. |
| $Quotes_j$ | Number of quotes in the press release associated with filing $j$. | 1994-2011. |
| $MultipleItem_j$ | Dummy=1 if there are more than 3 different categories under filing $j$ (ignoring financial statements) | 1994-2011. |
| $ExternalIR_j$ | Dummy=1 if an external contact person is listed in the press release associated with filing $j$. | 1994-2011. |
| **SEC Traffic Characteristics** | | |
| $HumanTraffic_{j,t}$ | Number of unique humanlike IPs visiting filing $j$ on calendar day $t$ where $t = 0$ indexes the first day of availability. | SEC EDGAR Traffic 2008-2011. |
| $HumanTraffic2_{j,t}$ | Number of unique humanlike IPs visiting firm $j$ on calendar day $t$ where $t = 0$ indexes the first day of availability and firm $j$ on date $t$ released an 8-K. | SEC EDGAR Traffic 2008-2011. |
| **Firm Characteristics** | | |
| $returns_{i,t}$ | Abnormal returns for firm $i$ on day $t$ defined as characteristically adjusted returns based quintiles of size and book to market. | CRSP 1994-2011. |
| $nanalysts_{i,t}$ | Number of analysts reporting earnings for the point in time quarter prior to current filing quarter $t$ for firm $i$. | IBES 1994-2011. |
| $IO_{i,t}$ | Institutional ownership at reporting date $t-2$ for filing quarter $t$ and firm $i$ | Thompson Reuters 1994-2011. |
| $NASDAQ_i$ | dummy which equals 1 when the stock is listed on the NASDAQ exchange. | CRSP 1994-2011. |
| $GICS2_i$ | 2-digit GICS industry classification as of 2011. | Compustat 2011. |
| $mcap_{i,t}$ | Log market cap at end of previous calendar quarter of filing quarter $t$ for firm $i$. CRSP 1994-2011. | |
| $shareturnover_{i,t}$ | Average trading volume scaled by average shares outstanding over the prior quarter of the filing quarter for firm $i$ | 1994-2011. |
| $momentum_{i,t}$ | The cumulative abnormal return [-12,-2] months prior to the filing date $t$ for firm $i$. | CRSP 1994-2011 |
| $b2m_{i,t}$ | The book value of equity divided by the market value of equity at end of previous calendar year $t$ for firm $i$. | Compustat 1994-2011. |

## Table II.
## Summary Statistics of Contents of 8-K and Press Release

Panel A reports summary statistics which describe the textual content of the entire 8-K sample filed between 1994-2011. The 8-K word count is drawn from the main text of the 8-K starting at the first item number and ends at the signature at the bottom of the main document. The count of good and bad words are matched with the Loughran and McDonald (2011) dictionary. The sentiment measure is the difference between the number of good and bad words scaled by the word count and multiplied by 100. Panel B describes the subsample of 8-Ks where the firm also reported a press release within the 8-K. In addition to the variables reported in Panel A, Panel B also reports the sentiment bias, which is defined as the difference between the press release and the 8-K sentiment. Inhouse IR is a dummy which equals 1 if the contact number reported in the press release matches the number of the firm listed in the 8-K. Number of quotes in PR is the number of times the press release uses a text quote. Panel C reports the pair-wise correlation between each of the individual sentiment measures.

### Panel A: All 8-Ks

| | N | Mean | Median | Std. Dev. | Q1 | Q3 |
|---|---|---|---|---|---|---|
| Press Release Bundled with 8-K | 794035 | 0.37 | 0.00 | 0.48 | 0.00 | 1.00 |
| 8-K Word Count | 794035 | 365.86 | 199.00 | 510.31 | 94.00 | 423.00 |
| 8-K Positive Word Percent | 794035 | 0.31 | 0.00 | 0.48 | 0.00 | 0.47 |
| 8-K Negative Word Percent | 794035 | 0.62 | 0.33 | 0.89 | 0.00 | 0.93 |
| Net 8-K Sentiment | 794035 | -0.31 | 0.00 | 0.92 | -0.63 | 0.00 |

### Panel B: All 8-K With Press Releases

| | N | Mean | Median | Std. Dev. | Q1 | Q3 |
|---|---|---|---|---|---|---|
| 8-K Word Count | 256782 | 328.61 | 152.00 | 487.89 | 78.00 | 367.00 |
| 8-K Positive Word Percent | 256782 | 0.28 | 0.00 | 0.45 | 0.00 | 0.42 |
| 8-K Negative Word Percent | 256782 | 0.51 | 0.00 | 0.76 | 0.00 | 0.80 |
| Press Release Word Count | 256782 | 712.66 | 503.00 | 861.75 | 332.00 | 772.00 |
| Press Release Positive Word Percent | 256782 | 0.87 | 0.77 | 0.63 | 0.40 | 1.23 |
| Press Release Negative Word Percent | 256782 | 0.62 | 0.48 | 0.61 | 0.20 | 0.87 |
| Net 8-K Sentiment | 256782 | -0.23 | 0.00 | 0.82 | -0.51 | 0.00 |
| Net Press Release Sentiment | 256782 | 0.25 | 0.24 | 0.92 | -0.26 | 0.81 |
| Sentiment Bias | 256782 | 0.48 | 0.42 | 1.03 | -0.10 | 1.07 |
| Number of Quotes in PR | 256782 | 0.24 | 0.00 | 0.68 | 0.00 | 0.00 |

### Panel C: Correlation Matrix Between Sentiment Measures

| | PR Pos Word % | PR Neg Word % | 8-K Pos Word % | 8-K Neg Word % |
|---|---|---|---|---|
| PR Pos Word % | 1 | -0.116*** | 0.183*** | -0.00918*** |
| PR Neg Word % | -0.116*** | 1 | -0.0101*** | 0.379*** |
| 8-K Pos Word % | 0.183*** | -0.0101*** | 1 | 0.149*** |
| 8-K Neg Word % | -0.00918*** | 0.379*** | 0.149*** | 1 |

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

**Table III. Text Sentiments By Type of 8-K Filing**

This table reports the raw 8-K and press release sentiment scores by different categories of 8-Ks and by whether the 8-K was also accompanied by a press release. The positive and negative variables are calculated as the number of positive and negative words in either the 8-K or press release text as identified through the word lists of Loughran and McDonald (2011) and scaled by the word count of the respective text. Panel A is restricted to the set of 8-Ks which report Item 1.01 "Entry Into a Definitive Material Agreement." Panel B is restricted to the set of 8-Ks which report either Item 5.02 (Departure of Directors or Principal Officers; Election of Directors; Appointment of Principal Officers) in the new 8-K, or Item 6 (Resignation of Directors) in the old 8-K.
Panel C is restricted to the following items in the new 8-K: Item 1.02 (Termination of a Material Definitive Agreement), Item 1.03 (Bankruptcy or Receivership), Item 2.04 (Trigger Events That Accelerate or Increase a Direct Financial Obligation under an Off-Balance Sheet Agreement), Item 2.06(Material Impairments), Item 3.01 (Notice of Delisting or Failure to Satisfy a Continued Listing Rule or Standard; Transfer of Listing), Item 4.01 (Changes in Registrants Certifying Accountant), Item 4.02 (Non-Reliance on Previously Issued Financial Statements or a Related Audit Report or Completed Interim Review), Item 5.03 (Amendments to Articles of Incorporation or Bylaws; Change in Fiscal Year). The corresponding items in the old 8-K reflecting the same heading are also included: Item 1 (Change in Control), Item 3 (Bankruptcy), Item 4 (Change of Auditor), and Item 10 (Amendment to Code of Ethics). Finally the samples in all panels are restricted to 8-Ks which do not report multiple item numbers (inclusion of item 9.01 financial statements and exhibits is permitted).

**Panel A: 8-K Entry Into A Material Definitive Agreement (Item 1.01)**

| 8-K | 8-K Pos Words | 8-K Neg Words | PR Pos Words | PR Neg Words | N |
|---|---|---|---|---|---|
| No Press Release | 0.373% | 0.665% | | | 55470 |
| Press Release | 0.365% | 0.686% | 0.985% | 0.600% | 9500 |

**Panel B: 8-K Change of Directors or Officers (Item 5.02 or Item 6)**

| 8-K | 8-K Pos Words | 8-K Neg Words | PR Pos Words | PR Neg Words | N |
|---|---|---|---|---|---|
| No Press Release | 0.607% | 0.962% | | | 62872 |
| Press Release | 0.456% | 0.504% | 1.286% | 0.400% | 18999 |

**Panel C: 8-K Pooled Ex-Ante Expected Negative Events**

| 8-K | 8-K Pos Words | 8-K Neg Words | PR Pos Words | PR Neg Words | N |
|---|---|---|---|---|---|
| No Press Release | 0.400% | 1.254% | | | 49096 |
| Press Release | 0.525% | 1.200% | 0.778% | 0.941% | 10642 |

## Table IV.
## What Determines Press Releases?

Coefficients reported in this table are results of Equation 1 which is a logit regression where the outcome variable equals one if the firm issued a press release in addition to the 8-K. Each observation is an 8-K filing from the sample period June 1996 to December 2011. Column 2 reports both the effects of positive and negative sentiment scores in the 8-K. Columns 3 and 4 include firm fixed effects and are estimated as a linear probability model. The net 8-K and Press sentiments are measured as the number of positive - minus negatives scaled by the respective word count of the texts. The positive and negative word percent variables are the number of good and bad words scaled by the word content, for each of the respective texts. All sentiment words are identified through Loughran and McDonald (2011)'s word lists. All sentiment measures are further normalized to be $\mu = 0$ and $\sigma = 1$. The Press=8-K indicator in columns 5-7 refer to the sub sample where the press release date equals the 8-K filing date. All specifications include 8-K item numbers, year and industry fixed effects (2-digit GICS), and firm controls. 8-K item number dummies refer to a vector of dummies which equal one if the filing contains a particular item number as listed in the appendix. Firm controls include: 5 quintiles of log market cap, log book to market, log number of analysts, institutional ownership, log share turnover, and a Nasdaq dummy. Standard errors are reported in square brackets and are clustered by firm and 8-K filing date as otherwise stated. The symbols $+, *, **$ represent significant at the 10%, 5% and 1% level respectively.

| | Logistic | | LPM | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| Net 8-K Sentiment | 0.1450** | | 0.0250** | |
| | [0.0098] | | [0.0007] | |
| 8-K Pos Word % | | -0.0169$^{+}$ | | -0.0023** |
| | | [0.0095] | | [0.0007] |
| 8-K Neg Word % | | -0.1897** | | -0.0321** |
| | | [0.0113] | | [0.0008] |
| Multiple Item Numbers in 8-K | 0.3477** | 0.3540** | 0.0166** | 0.0196** |
| | [0.0489] | [0.0490] | [0.0043] | [0.0043] |
| Last Week Form 4 | -0.1890** | -0.1909** | -0.0320** | -0.0323** |
| | [0.0150] | [0.0151] | [0.0019] | [0.0019] |
| Next Week Form 4 | 0.0072 | 0.0046 | -0.0012 | -0.0017 |
| | [0.0149] | [0.0149] | [0.0018] | [0.0018] |
| Momentum [-12 Months,-2 Months] | 0.0177* | 0.0163$^{+}$ | 0.0019$^{+}$ | 0.0017 |
| | [0.0089] | [0.0088] | [0.0011] | [0.0011] |
| Friday Filing Date | -0.0728** | -0.0712** | -0.0123** | -0.0120** |
| | [0.0166] | [0.0166] | [0.0022] | [0.0022] |
| Constant | -2.5385** | -2.5139** | 0.0917** | 0.0989** |
| | [0.1492] | [0.1499] | [0.0136] | [0.0136] |
| Observations | 307291 | 307291 | 307294 | 307294 |
| 8-K Item Number Dummies | Yes | Yes | Yes | Yes |
| Firm Characteristics Controls | Yes | Yes | Yes | Yes |
| Year and Industry Fixed Effects | Yes | Yes | Yes | Yes |
| Firm Fixed Effects | No | No | Yes | Yes |
| Twoway Clustered SE | Yes | Yes | Date Only | Date Only |

## Table V.
## What Determines the Sentiment Bias?

Coefficients reported in this table are results of Equation 2 where the dependent variable is the sentiment bias. Each observation is an 8-K filing from the sample period June 1996 to December 2011. The net 8-K and Press sentiments are measured as the number of positive - minus negatives scaled by the respective word count of the texts. The positive and negative word percent variables are the number of good and bad words scaled by the word content, for each of the respective texts. All sentiment words are identified through Loughran and McDonald (2011)'s word lists. All sentiment measures are further normalized to be $\mu = 0$ and $\sigma = 1$. The Press=8-K indicator in columns 5-7 refer to the sub sample where the press release date equals the 8-K filing date. All specifications include 8-K item numbers, year and industry fixed effects (2-digit GICS), and firm controls. 8-K item number dummies refer to a vector of dummies which equal one if the filing contains a particular item number as listed in the appendix. Firm controls include: momentum[-12 months, -2 months], 5 quintiles of log market cap, log book to market, log number of analysts, institutional ownership, log share turnover, and a Nasdaq dummy. Standard errors are reported in square brackets and are clustered by firm and 8-K filing date as otherwise stated. The symbols $+, *, **$ represent significant at the 10%, 5% and 1% level respectively.

| | Sentiment Bias | | | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| Net 8-K Sentiment | -0.6527** | -0.6540** | -0.6565** | -0.6601** |
| | [0.0072] | [0.0099] | [0.0037] | [0.0058] |
| External IR | 0.0985** | 0.0964** | 0.1037** | 0.1009** |
| | [0.0138] | [0.0171] | [0.0066] | [0.0101] |
| Number of Quotes in PR | 0.1692** | 0.1722** | 0.1690** | 0.1741** |
| | [0.0111] | [0.0089] | [0.0113] | [0.0072] |
| Multiple Item Numbers in 8-K | 0.0748** | 0.0913** | 0.0715** | 0.0834** |
| | [0.0142] | [0.0209] | [0.0115] | [0.0185] |
| Last Week Form 4 | 0.0300** | 0.0333** | 0.0114* | 0.0115 |
| | [0.0069] | [0.0091] | [0.0055] | [0.0079] |
| Next Week Form 4 | 0.0295** | 0.0340** | 0.0211** | 0.0211** |
| | [0.0068] | [0.0093] | [0.0053] | [0.0076] |
| Momentum [-12 Months,-2 Months] | 0.0032 | 0.0035 | 0.0051+ | 0.0101* |
| | [0.0039] | [0.0055] | [0.0028] | [0.0045] |
| Log Book to Market | -0.0304** | -0.0414** | -0.0202** | -0.0153* |
| | [0.0068] | [0.0090] | [0.0041] | [0.0064] |
| Log Number of Analysts | 0.0079** | 0.0072** | -0.0011 | -0.0038+ |
| | [0.0020] | [0.0026] | [0.0013] | [0.0020] |
| IO Percent | 0.0212 | -0.0121 | 0.0367* | -0.0038 |
| | [0.0255] | [0.0349] | [0.0170] | [0.0274] |
| NASDAQ Listed | 0.0403* | 0.0108 | 0.0320+ | 0.0457 |
| | [0.0160] | [0.0211] | [0.0186] | [0.0311] |
| Log Share Turnover | -0.0340** | -0.0336** | -0.0321** | -0.0321** |
| | [0.0061] | [0.0080] | [0.0036] | [0.0056] |
| Friday Filing Date | -0.0270** | -0.0730** | -0.0311** | -0.0719** |
| | [0.0061] | [0.0108] | [0.0055] | [0.0088] |
| Constant | -0.3402** | -0.3468** | -0.2303** | -0.1707+ |
| | [0.0494] | [0.0819] | [0.0453] | [0.0916] |
| Observations | 129609 | 58213 | 129609 | 58213 |
| Adjusted $R^2$ | 0.350 | 0.354 | 0.476 | 0.499 |
| 8-K Item Number Dummies | Yes | Yes | Yes | Yes |
| Firm Characteristics Controls | Yes | Yes | Yes | Yes |
| PR=8-K Date | All | Yes | All | Yes |
| Year and Industry Fixed Effects | Yes | Yes | Yes | Yes |
| Firm Fixed Effects | No | No | Yes | Yes |
| Twoway Clustered SE | Yes | Yes | Date Only | Date Only |

## Table VI.
## CAR[-15,1] Around 8-K Filing Date- Baseline

This table reports regression results where the dependent variable is the cumulative abnormal returns around the -15 to 1 trading day window surrounding the filing date of an 8-K and the explanatory variables are measures of textual sentiment in the 8-K and press release when available, along with other controls as specified in Equation 3. Abnormal return is calculated as the raw firm return minus a size(5 quantiles) and book to market(5 quantiles) matched value weighted portfolio return. The sample encompasses 8-Ks filed between June 1996 and December 2011 excluding those identified as Item 2.02 "Results of Operations and Financial Conditions" in the post August 2004 8-K or Item 12 "Results of Operation" in the pre August 2004 8-K. It also excludes any 8-Ks which are filed concurrently with other confounding events as listed in the text. The dummy "Issued Press Release" equals one if the 8-K has an attached press release in the SEC filings. The net 8-K and Press sentiments are measured as the number of positive - minus negatives scaled by the respective word count of the texts. The positive and negative word percent variables are the number of good and bad words scaled by the word content, for each of the respective texts. All sentiment words are identified through Loughran and McDonald (2011)'s word lists. All sentiment measures are further normalized to be $\mu = 0$ and $\sigma = 1$. The Press=8-K indicator refer to the sub sample where the press release date equals the 8-K filing date. All specifications include 8-K item numbers, year and industry fixed effects (2-digit GICS), and firm controls. 8-K item number dummies refer to a vector of dummies which equal one if the filing contains a particular item number as listed in the appendix. Firm controls include: momentum[-12 months, -2 months], 5 quintiles of log market cap, log book to market, log number of analysts, institutional ownership, log share turnover, and a Nasdaq dummy. Standard errors are reported in square brackets and are clustered by firm and 8-K filing date as otherwise stated. The symbols $+, *, **$ represent significant at the 10%, 5% and 1% level respectively.

| | Cumulative Abnormal Returns On Filing Date[-15,1] | | | | | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Issued Press Release | 0.0022** | | | | | |
| | [0.0007] | | | | | |
| Net 8-K Sentiment | 0.0033** | 0.0059** | 0.0112** | 0.0117** | 0.0121** | 0.0126** |
| | [0.0003] | [0.0006] | [0.0007] | [0.0010] | [0.0007] | [0.0010] |
| Sentiment Bias | | | 0.0088** | 0.0091** | 0.0106** | 0.0109** |
| | | | [0.0006] | [0.0009] | [0.0006] | [0.0009] |
| Constant | 0.0032 | 0.0082 | 0.0088+ | 0.0008 | 0.0048 | 0.0064 |
| | [0.0033] | [0.0050] | [0.0050] | [0.0106] | [0.0196] | [0.0218] |
| Observations | 302592 | 135271 | 135271 | 59397 | 135271 | 59397 |
| Adjusted $R^2$ | 0.378 | 0.342 | 0.344 | 0.314 | 0.380 | 0.363 |
| 8-K Item Number Dummies | Yes | Yes | Yes | Yes | Yes | Yes |
| Firm Characteristics Controls | Yes | Yes | Yes | Yes | Yes | Yes |
| Year and Industry Fixed Effects | Yes | Yes | Yes | Yes | Yes | Yes |
| PR=8-K Date | All | All | All | Yes | All | Yes |
| Firm Fixed Effects | No | No | No | No | Yes | Yes |
| Twoway Clustered SE | Yes | Yes | Yes | Yes | Date Only | Date Only |

## Table VII.
## CAR[-15,1] Around Filing Date- Decomposed Sentiment Specifications

This table reports regression results where the dependent variable is the cumulative abnormal returns around the -15 to 1 trading day window surrounding the filing date of an 8-K and the explanatory variables are measures of textual sentiment in the 8-K and press release when available, along with other controls as specified in Equation 3. Abnormal return is calculated as the raw firm return minus a size(5 quantiles) and book to market(5 quantiles) matched value weighted portfolio return. The sample encompasses 8-Ks filed between June 1996 and December 2011 excluding those identified as Item 2.02 "Results of Operations and Financial Conditions" in the post August 2004 8-K or Item 12 "Results of Operation" in the pre August 2004 8-K. It also excludes any 8-Ks which are filed concurrently with other confounding events as listed in the text. The dummy "Issued Press Release" equals one if the 8-K has an attached press release in the SEC filings. The net 8-K and Press sentiments are measured as the number of positive - minus negatives scaled by the respective word count of the texts. The positive and negative word percent variables are the number of good and bad words scaled by the word content, for each of the respective texts. All sentiment words are identified through Loughran and McDonald (2011)'s word lists. All sentiment measures are further normalized to be $\mu = 0$ and $\sigma = 1$. The Press=8-K indicator refer to the sub sample where the press release date equals the 8-K filing date. All specifications include 8-K item numbers, year and industry fixed effects (2-digit GICS), and firm controls. 8-K item number dummies refer to a vector of dummies which equal one if the filing contains a particular item number as listed in the appendix. Firm controls include: momentum[-12 months, -2 months], 5 quintiles of log market cap, log book to market, log number of analysts, institutional ownership, log share turnover, and a Nasdaq dummy. Standard errors are reported in square brackets and are clustered by firm and 8-K filing date as otherwise stated. The symbols $+, *, **$ represent significant at the 10%, 5% and 1% level respectively.

| | Cumulative Abnormal Returns On Filing Date[-15,1] | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| Issued Press Release | 0.0022** | 0.0022** | | | | | | |
| | [0.0007] | [0.0007] | | | | | | |
| Net 8-K Sentiment | 0.0033** | | 0.0033** | | 0.0036** | | 0.0029** | |
| | [0.0003] | | [0.0006] | | [0.0008] | | [0.0008] | |
| Net Press Sentiment | | | 0.0079** | | 0.0082** | | 0.0098** | |
| | | | [0.0005] | | [0.0008] | | [0.0008] | |
| 8-K Pos Word % | | 0.0014** | | 0.0022** | | 0.0023** | | 0.0019* |
| | | [0.0003] | | [0.0005] | | [0.0007] | | [0.0008] |
| 8-K Neg Word % | | -0.0033** | | -0.0024** | | -0.0026** | | -0.0018+ |
| | | [0.0003] | | [0.0006] | | [0.0009] | | [0.0010] |
| PR Pos Word % | | | | 0.0040** | | 0.0039** | | 0.0046** |
| | | | | [0.0005] | | [0.0007] | | [0.0008] |
| PR Neg Word % | | | | -0.0068** | | -0.0073** | | -0.0087** |
| | | | | [0.0006] | | [0.0008] | | [0.0009] |
| Constant | 0.0032 | 0.0033 | 0.0095+ | 0.0090+ | 0.0015 | 0.0010 | 0.0073 | 0.0071 |
| | [0.0033] | [0.0033] | [0.0050] | [0.0050] | [0.0106] | [0.0106] | [0.0218] | [0.0219] |
| Observations | 302592 | 302592 | 135271 | 135271 | 59397 | 59397 | 59397 | 59397 |
| Adjusted $R^2$ | 0.378 | 0.378 | 0.344 | 0.344 | 0.314 | 0.314 | 0.363 | 0.363 |
| 8-K Item Number Dummies | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Firm Characteristics Controls | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Year and Industry Fixed Effects | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| PR=8-K Date | All | All | All | All | Yes | Yes | Yes | Yes |
| Firm Fixed Effects | No | No | No | No | No | No | Yes | Yes |
| Twoway Clustered SE | Yes | Yes | Yes | Yes | Yes | Yes | Date Only | Date Only |

## Table VIII.
## CAR[-15,1] Around Filing Date- Robustness Checks

This table reports regression results where the dependent variable is the cumulative abnormal returns around the -15 to 1 trading day window surrounding the filing date of an 8-K and the explanatory variables are measures of textual sentiment in the 8-K and press release when available, along with other controls as specified in Equation 3. Column 1 represents the baseline sample of all 8-Ks and press releases. Column 2 restricts the sample to 8-Ks which do not contain empty positive or negative tone. Column 3 restricts the sample of 8-Ks to those with at least a 100 word count, which represents the 25% percentile of word count. Column 4 identifies the 8-Ks with zero tones (those restricted out in Column 2) and code them as having the same sentiment as the press release. Abnormal return is calculated as the raw firm return minus a size(5 quantiles) and book to market(5 quantiles) matched value weighted portfolio return. The sample encompasses 8-Ks filed between June 1996 and December 2011 excluding those identified as Item 2.02 "Results of Operations and Financial Conditions" in the post August 2004 8-K or Item 12 "Results of Operation" in the pre August 2004 8-K. It also excludes any 8-Ks which are filed concurrently with other confounding events as listed in the text. The dummy "Issued Press Release" equals one if the 8-K has an attached press release in the SEC filings. The net 8-K and Press sentiments are measured as the number of positive - minus negatives scaled by the respective word count of the texts. The positive and negative word percent variables are the number of good and bad words scaled by the word content, for each of the respective texts. All sentiment words are identified through Loughran and McDonald (2011)'s word lists. All sentiment measures are further normalized to be $\mu = 0$ and $\sigma = 1$. The Press=8-K indicator refer to the sub sample where the press release date equals the 8-K filing date. All specifications include 8-K item numbers, year and industry fixed effects (2-digit GICS), and firm controls. 8-K item number dummies refer to a vector of dummies which equal one if the filing contains a particular item number as listed in the appendix. Firm controls include: momentum[-12 months, -2 months], 5 quintiles of log market cap, log book to market, log number of analysts, institutional ownership, log share turnover, and a Nasdaq dummy. Standard errors are reported in square brackets and are clustered by firm and 8-K filing date as otherwise stated. The symbols $+, *, **$ represent significant at the 10%, 5% and 1% level respectively.

| | Baseline | Non-Zero 8-K Sentiment | 8-K Word Count>100 | Set 8-K=Press Sentiment |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| Net 8-K Sentiment | 0.0033** | 0.0044** | 0.0052** | 0.0039** |
| | [0.0006] | [0.0006] | [0.0008] | [0.0005] |
| Net Press Sentiment | 0.0079** | 0.0067** | 0.0067** | 0.0065** |
| | [0.0005] | [0.0006] | [0.0006] | [0.0006] |
| Constant | 0.0095+ | 0.0160* | 0.0159* | 0.0101* |
| | [0.0050] | [0.0066] | [0.0067] | [0.0050] |
| Observations | 135271 | 80215 | 86923 | 135271 |
| Adjusted $R^2$ | 0.344 | 0.353 | 0.343 | 0.344 |
| 8-K Item Number Dummies | Yes | Yes | Yes | Yes |
| Firm Characteristics Controls | Yes | Yes | Yes | Yes |
| Year and Industry Fixed Effects | Yes | Yes | Yes | Yes |
| PR=8-K Date | All | All | All | All |
| Firm Fixed Effects | No | No | No | No |
| Twoway Clustered SE | Yes | Yes | Yes | Yes |

## Table IX.
## CAR[2,20] After Filing Date

This table reports regression results where the dependent variable is the cumulative abnormal returns around the 2 to 20 trading day window after the filing date of an 8-K and the explanatory variables are measures of textual sentiment in the 8-K and press release when available, along with other controls as specified in Equation 3. Abnormal return is calculated as the raw firm return minus a size(5 quantiles) and book to market(5 quantiles) matched value weighted portfolio return. The sample encompasses 8-Ks filed between June 1996 and December 2011 excluding those identified as Item 2.02 "Results of Operations and Financial Conditions" in the post August 2004 8-K or Item 12 "Results of Operation" in the pre August 2004 8-K. It also excludes any 8-Ks which are filed concurrently with other confounding events as listed in the text. The positive and negative word percent variables are the number of good and bad words scaled by the word content, for each of the respective texts. All sentiment words are identified through Loughran and McDonald (2011)'s word lists. All sentiment measures are further normalized to be $\mu = 0$ and $\sigma = 1$. The Press< 8-K indicator refers to the relative timing between the press release and the 8-K date. All specifications include 8-K item numbers, year and industry fixed effects (2-digit GICS), and firm controls. 8-K item number dummies refer to a vector of dummies which equal one if the filing contains a particular item number as listed in the appendix. Firm controls include: cumulative abnormal returns[-20,-5] prior to the 8-K filing date, momentum[-12 months, -2 months], 5 quintiles of log market cap, log book to market, log number of analysts, institutional ownership, log share turnover, and a Nasdaq dummy. Standard errors are reported in square brackets and are clustered by firm and 8-K filing date as otherwise stated. The symbols $+, *, **$ represent significant at the 10%, 5% and 1% level.

| | Abnormal Returns On Filing Date[2,20] | | | | | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| 8-K Pos Word % | 0.0002 | 0.0003 | 0.0004 | 0.0003 | -0.0009 | 0.0007 |
| | [0.0013] | [0.0017] | [0.0019] | [0.0012] | [0.0018] | [0.0020] |
| 8-K Neg Word % | -0.0006 | -0.0001 | -0.0010 | -0.0014 | 0.0003 | -0.0016 |
| | [0.0013] | [0.0018] | [0.0019] | [0.0012] | [0.0018] | [0.0020] |
| PR Pos Word % | -0.0013* | -0.0020** | -0.0001 | -0.0020** | -0.0024** | -0.0008 |
| | [0.0005] | [0.0007] | [0.0007] | [0.0005] | [0.0007] | [0.0008] |
| PR Neg Word % | 0.0001 | 0.0003 | -0.0004 | 0.0003 | 0.0007 | 0.0002 |
| | [0.0005] | [0.0006] | [0.0009] | [0.0006] | [0.0007] | [0.0010] |
| Constant | 0.0046 | 0.0041 | 0.0133 | 0.0113 | 0.0072 | 0.0140 |
| | [0.0056] | [0.0062] | [0.0106] | [0.0204] | [0.0074] | [0.0212] |
| Observations | 140357 | 75350 | 61365 | 140357 | 75350 | 61365 |
| Adjusted $R^2$ | 0.008 | 0.006 | 0.010 | 0.060 | 0.074 | 0.065 |
| 8-K Item Number Dummies | Yes | Yes | Yes | Yes | Yes | Yes |
| Firm Characteristics Controls | Yes | Yes | Yes | Yes | Yes | Yes |
| Year and Industry Fixed Effects | Yes | Yes | Yes | Yes | Yes | Yes |
| PR<Filing Date | All | Yes | Same Day | All | Yes | Same Day |
| Firm Fixed Effects | No | No | No | Yes | Yes | Yes |
| Twoway Clustered SE | Yes | Yes | Yes | Date Only | Date Only | Date Only |

## Table X.
## Market Reaction To 8-Ks Which Do Not Have Press Releases

This table reports regression results where the dependent variable is the cumulative abnormal returns for the specified window period in either Panel A or Panel B as estimated in Equation 3 for the sample of 8-Ks which do not have a press release . Abnormal return is calculated as the raw firm return minus a size(5 quantiles) and book to market(5 quantiles) matched value weighted portfolio return. The sample encompasses 8-Ks filed between June 1996 and December 2011 excluding those identified as Item 2.02 "Results of Operations and Financial Conditions" in the post August 2004 8-K or Item 12 "Results of Operation" in the pre August 2004 8-K. It also excludes any 8-Ks which are filed concurrently with other confounding events as listed in the text. The positive and negative word percent variables are the number of good and bad words scaled by the word content, for each of the respective texts. All sentiment words are identified through Loughran and McDonald (2011)'s word lists. All sentiment measures are further normalized to be $\mu = 0$ and $\sigma = 1$. All specifications include 8-K item numbers, year and industry fixed effects (2-digit GICS), and firm controls. 8-K item number dummies refer to a vector of dummies which equal one if the filing contains a particular item number as listed in the appendix. Firm controls include: cumulative abnormal returns[-20,-5] prior to the 8-K filing date, momentum[-12 months, -2 months], 5 quintiles of log market cap, log book to market, log number of analysts, institutional ownership, log share turnover, and a Nasdaq dummy. Standard errors are reported in square brackets and are clustered by firm and 8-K filing date as otherwise stated. The symbols $+, *, **$ represent significant at the 10%, 5% and 1% level.

### Panel A: Cumulative Abnormal Returns [-1,1]

| | Abnormal Returns On Filing Date[-1,1] | | | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| Net 8-K Sentiment | 0.0010** | | 0.0011** | |
| | [0.0002] | | [0.0002] | |
| 8-K Pos Word % | | 0.0013** | | 0.0013** |
| | | [0.0004] | | [0.0004] |
| 8-K Neg Word % | | -0.0012** | | -0.0012** |
| | | [0.0004] | | [0.0004] |
| Constant | 0.0042 | 0.0041 | 0.0105+ | 0.0104+ |
| | [0.0028] | [0.0028] | [0.0060] | [0.0060] |
| Observations | 156035 | 156035 | 156035 | 156035 |
| Adjusted $R^2$ | 0.002 | 0.002 | 0.049 | 0.049 |
| 8-K Item Number Dummies | Yes | Yes | Yes | Yes |
| Firm Characteristics Controls | Yes | Yes | Yes | Yes |
| Year and Industry Fixed Effects | Yes | Yes | Yes | Yes |
| Firm Fixed Effects | No | No | Yes | Yes |
| Twoway Clustered SE | Yes | Yes | Date Only | Date Only |

### Panel B: Cumulative Abnormal Returns [2,20]

| | Abnormal Returns On Filing Date[2,20] | | | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| Net 8-K Sentiment | 0.0015** | | 0.0015** | |
| | [0.0004] | | [0.0004] | |
| 8-K Pos Word % | | 0.0020** | | 0.0022** |
| | | [0.0007] | | [0.0007] |
| 8-K Neg Word % | | -0.0018* | | -0.0019* |
| | | [0.0008] | | [0.0008] |
| Constant | 0.0109 | 0.0107 | 0.0207* | 0.0205* |
| | [0.0068] | [0.0068] | [0.0104] | [0.0104] |
| Observations | 156035 | 156035 | 156035 | 156035 |
| Adjusted $R^2$ | 0.008 | 0.008 | 0.061 | 0.061 |
| 8-K Item Number Dummies | Yes | Yes | Yes | Yes |
| Firm Characteristics Controls | Yes | Yes | Yes | Yes |
| Year and Industry Fixed Effects | Yes | Yes | Yes | Yes |
| Firm Fixed Effects | No | No | Yes | Yes |
| Twoway Clustered SE | Yes | Yes | Date Only | Date Only |

## Table XI.
## EDGAR Search Traffic As Function of 8-K Characteristics

This table reports regression results which explain the amount of human search traffic for a particular 8-K filing on the date of the filing as a function of the characteristics of the 8-K as specified in Equation 4. The sample encompasses 8-Ks filed between Jan 2008 and December 2011 excluding those identified as Item 2.02 "Results of Operations and Financial Conditions" in the post August 2004 8-K or Item 12 "Results of Operation" in the pre August 2004 8-K. It also excludes any 8-Ks which are filed concurrently with other confounding events as listed in the text.Issued press release is a dummy if the 8-K contained a press release attachment. The positive and negative word percent variables are the number of good and bad words scaled by the word content, for each of the respective texts. All sentiment words are identified through Loughran and McDonald (2011)'s word lists. All sentiment measures are further normalized to be $\mu = 0$ and $\sigma = 1$. The Press< 8-K indicator refers to the relative timing between the press release and the 8-K date. All specifications include 8-K item numbers, year and industry fixed effects (2-digit GICS), and firm controls. 8-K item number dummies refer to a vector of dummies which equal one if the filing contains a particular item number as listed in the appendix. Firm controls include: cumulative abnormal returns[-20,-5] prior to the 8-K filing date, momentum[-12 months, -2 months], 5 quintiles of log market cap, log book to market, log number of analysts, institutional ownership, log share turnover, and a Nasdaq dummy. Additional controls include hourly dummies corresponding to the dissemination timestamp of the 8-K. Standard errors are reported in square brackets and are clustered by firm and 8-K filing date as otherwise stated. The symbols $+, *, **$ represent significant at the 10%, 5% and 1% level.

| | Log(Unique Humans) (8-K Traffic) | | | | | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Issued Press Release | -0.4895** | -0.4867** | -0.7815** | -0.3454** | -0.3447** | -0.5472** |
| | [0.0353] | [0.0351] | [0.0607] | [0.0249] | [0.0249] | [0.0460] |
| Net 8-K Sentiment | -0.1138** | -0.1016** | -0.1205** | -0.0814** | -0.0777** | -0.0701** |
| | [0.0121] | [0.0130] | [0.0213] | [0.0103] | [0.0118] | [0.0191] |
| Net 8-K Sentiment X Press | | -0.0376 | | | -0.0119 | |
| | | [0.0246] | | | [0.0201] | |
| Multiple Item Numbers in 8-K | 0.1907** | 0.1889** | 0.2681* | 0.0689+ | 0.0684+ | 0.0174 |
| | [0.0469] | [0.0468] | [0.1119] | [0.0414] | [0.0414] | [0.1003] |
| PRE CAR[20,5] | -0.1114 | -0.1115 | -0.0827 | -0.0884+ | -0.0884+ | -0.0852 |
| | [0.0694] | [0.0694] | [0.1115] | [0.0524] | [0.0524] | [0.1006] |
| Momentum [-12 Months,-2 Months] | -0.1358** | -0.1358** | -0.1228** | -0.0790** | -0.0790** | -0.0854** |
| | [0.0247] | [0.0247] | [0.0311] | [0.0166] | [0.0166] | [0.0236] |
| Friday Filing Date | -0.3828** | -0.3827** | -0.3228** | -0.3941** | -0.3941** | -0.3572** |
| | [0.0791] | [0.0791] | [0.0988] | [0.0770] | [0.0770] | [0.0905] |
| NASDAQ Listed | -0.1495** | -0.1493** | -0.1187+ | 0.0822 | 0.0819 | 0.1286 |
| | [0.0475] | [0.0475] | [0.0695] | [0.1431] | [0.1431] | [0.2454] |
| Log Number of Analysts | -0.0138* | -0.0139* | -0.0246* | -0.0063 | -0.0063 | -0.0026 |
| | [0.0063] | [0.0063] | [0.0098] | [0.0070] | [0.0070] | [0.0146] |
| IO Percent | -0.7820** | -0.7818** | -0.8422** | -0.1815+ | -0.1816+ | -0.0660 |
| | [0.0771] | [0.0771] | [0.1129] | [0.0983] | [0.0984] | [0.1879] |
| Log Share Turnover | 0.4697** | 0.4697** | 0.4957** | 0.2159** | 0.2159** | 0.1851** |
| | [0.0202] | [0.0202] | [0.0302] | [0.0185] | [0.0185] | [0.0348] |
| Log Book to Market | 0.0892** | 0.0888** | 0.0901* | 0.0883** | 0.0883** | 0.1541** |
| | [0.0226] | [0.0226] | [0.0365] | [0.0231] | [0.0231] | [0.0415] |
| Constant | -0.6583** | -0.6574** | -0.6787** | -0.3336* | -0.3330* | -0.3553 |
| | [0.1367] | [0.1367] | [0.1959] | [0.1401] | [0.1401] | [0.2351] |
| Observations | 105896 | 105896 | 34085 | 105896 | 105896 | 34085 |
| Adjusted $R^2$ | 0.118 | 0.118 | 0.131 | 0.206 | 0.206 | 0.229 |
| 8-K Item Number Dummies | Yes | Yes | Yes | Yes | Yes | Yes |
| Firm Characteristics Controls | Yes | Yes | Yes | Yes | Yes | Yes |
| Year and Industry Fixed Effects | Yes | Yes | Yes | Yes | Yes | Yes |
| Report Date=8-K Date | All | All | Yes | All | All | Yes |
| Firm Fixed Effects | No | No | No | Yes | Yes | Yes |
| Twoway Clustered SE | Yes | Yes | Yes | Date Only | Date Only | Date Only |

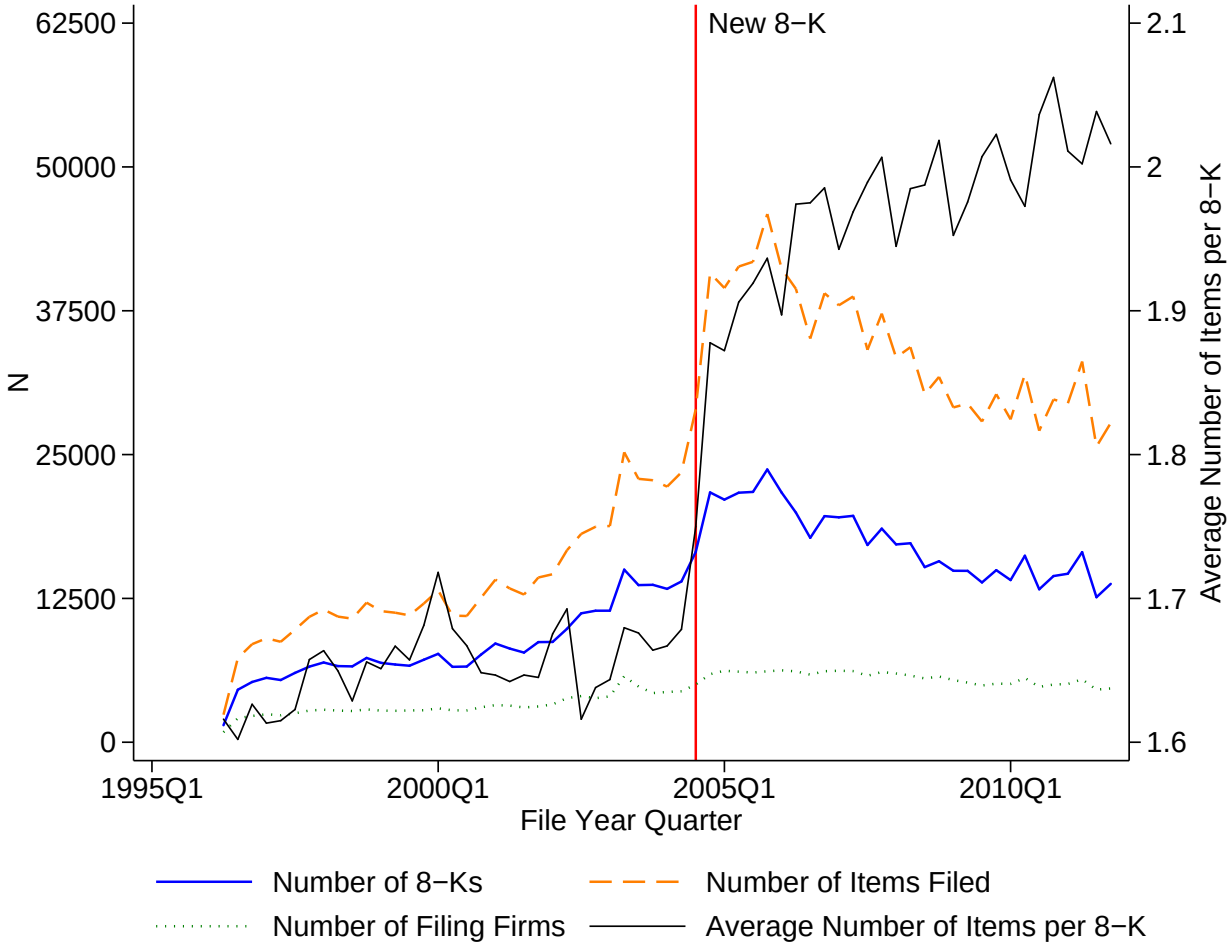



**Figure 1. Time Series Variation in 8-K Filings.** This figure plots the number of 8-Ks, the aggregated number of item categories within each 8-K, the number of filing firms, and the average number of item categories per 8-K over time. The sample excludes 8-Ks containing Item number 2.02(Item 12 in the old 8-K) "Results of Operations".

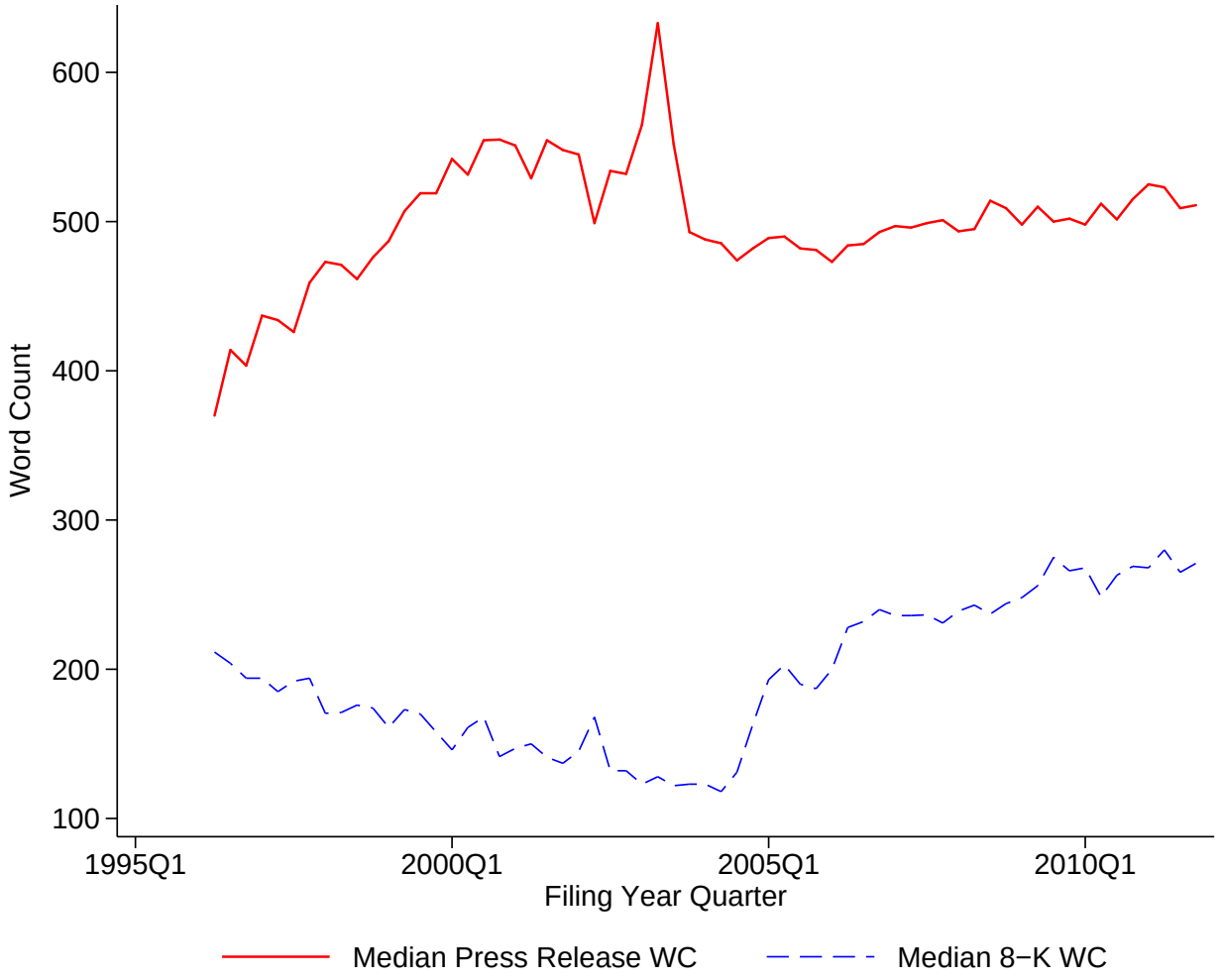
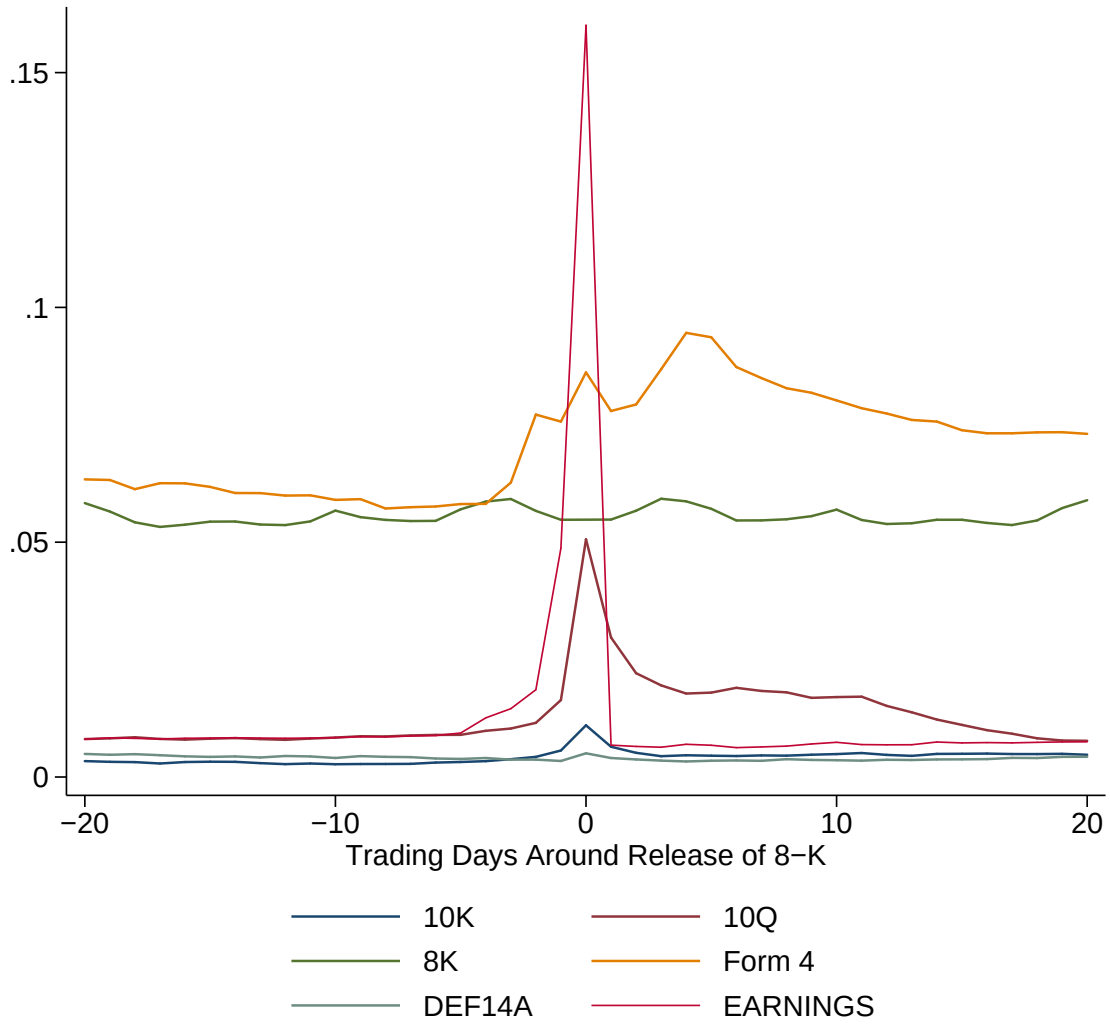
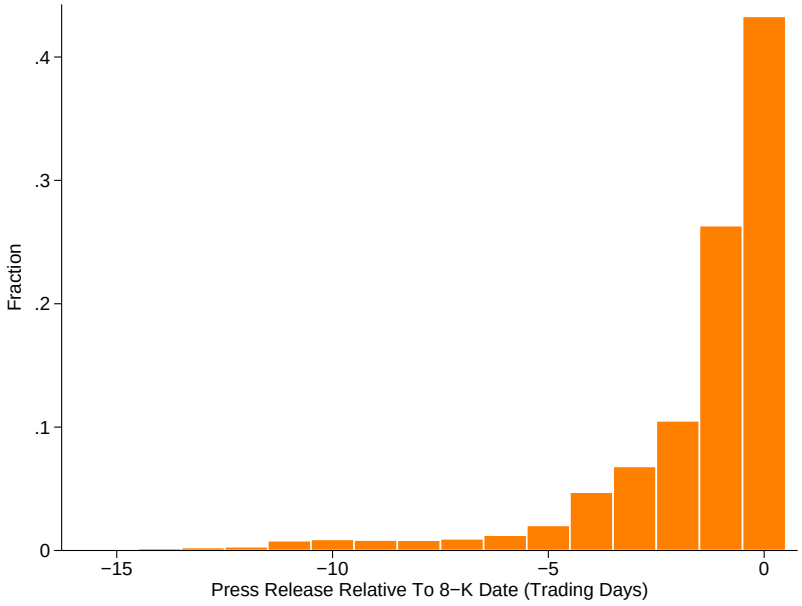



**Figure 2. Length of 8-Ks and Related Press Releases.** This figure plots the median word count of 8-Ks and press releases associated with 8-Ks a function of the filing time. Standard boilerplate SEC filing language, tables, as well as safe harbor and forward-looking statements are removed prior to the word count as described in Appendix A. The sample excludes 8-Ks containing Item number 2.02(Item 12 in the old 8-K) "Results of Operations".

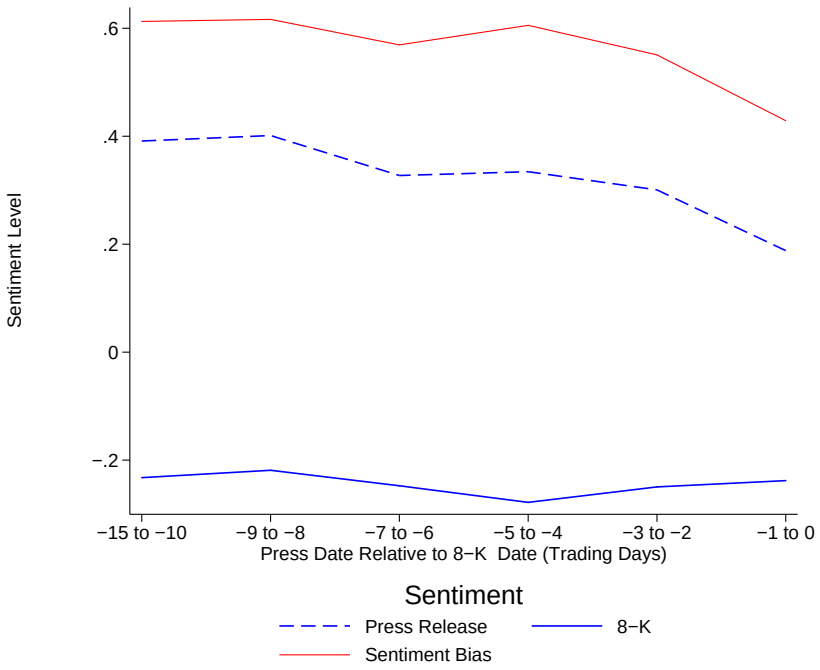



(b)

**Figure 3. Concurrency of 8-K With Other Events.** This figure plots the frequency of other common announcements and filings relative to the 8-K filing date for the universe(prior to filtering in the text) of 8-Ks filed between 1994 and 2011. Form 10-K and 10-Q are the firm's annual and quarterly reports, Form 4 refers to insider trading, DEF14A is the firm's annual proxy statement, and EARNINGS is firm's quarterly earnings announcement. The x-axis represents the number of trading days relative to the 8-K filing date which is centered at zero and the y-axis is the fraction $\in [0, 1]$ of correspondence between the 8-K date and the other announcement or filing date.




**(a) Press Date Relative to Filing Date**




**(b) Press Date Relative to Filing Date by Sentiment**

**Figure 4. 8-K Report and Press Release Date Relative to Filing Date.** Figure (a) plot the frequency distribution of 8-Ks which have a press release as a function of the number of trading days prior to the filing date. Figure (b) plots the average 8-K sentiment, the corresponding press release sentiment, and the difference between the two (sentiment bias) where the first two are measured by the net positive minus negative words scaled by word count as a function of the number of trading days prior to the 8-K filing date for the sample of 8-Ks with press releases.

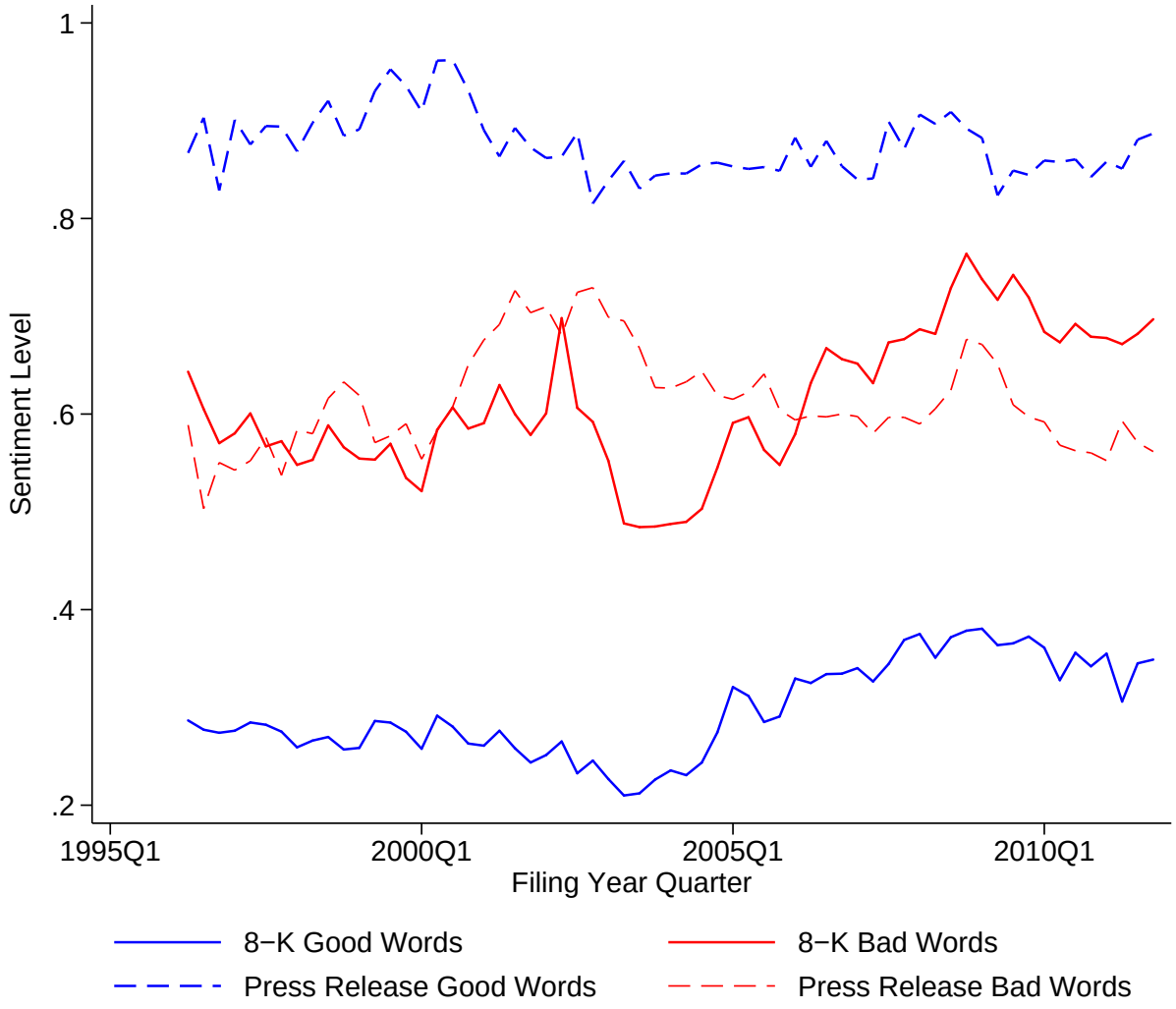
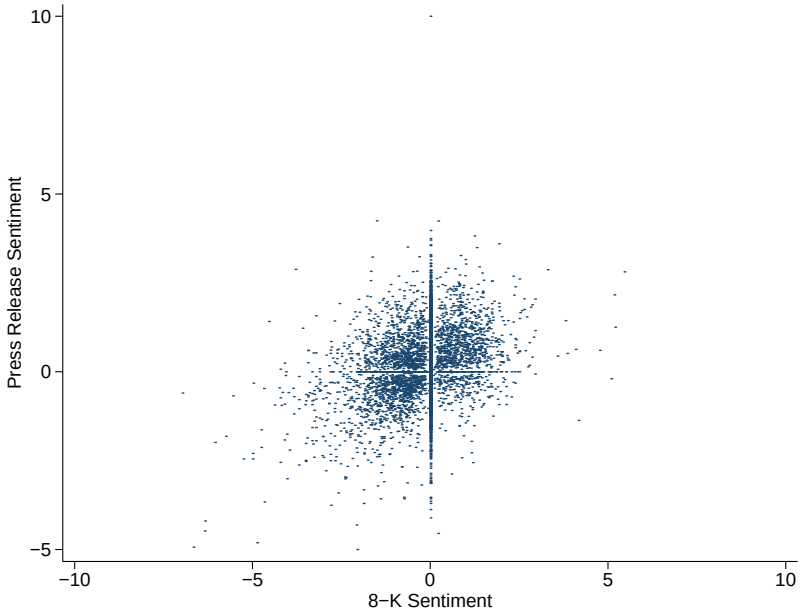
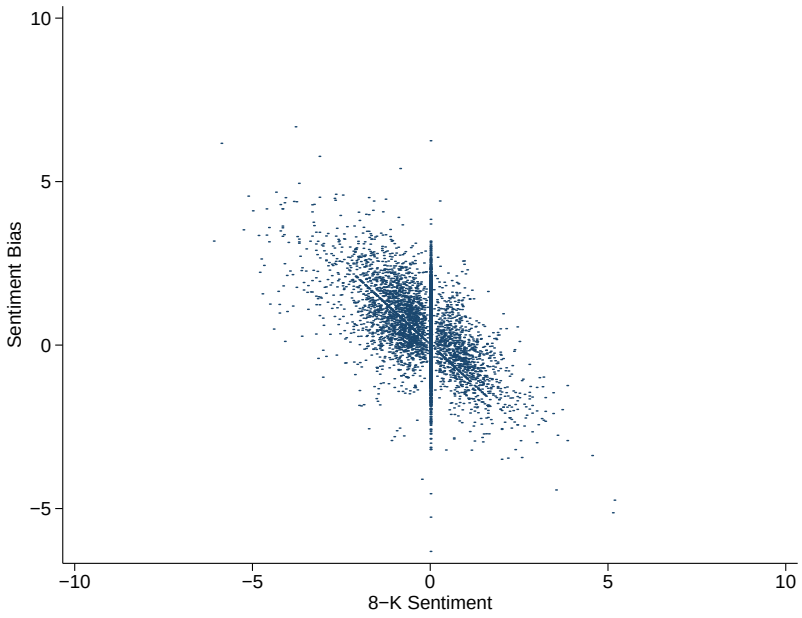



**Figure 5. Time Series Variation In Press Release and 8-K Sentiment.** This figure plots the time series variation in textual sentiment separately for the 8-K text and press releases associated with 8-Ks. Positive and negative sentiment are separately plotted and are defined as the fraction of positive or negative words in a particular text using the word lists of Loughran and McDonald (2011) and scaled by a factor of 100.

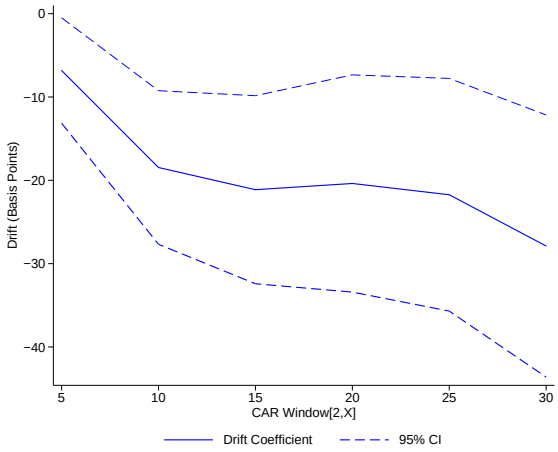



**(a) Net Press Release Sentiment vs Net 8-K Sentiment**

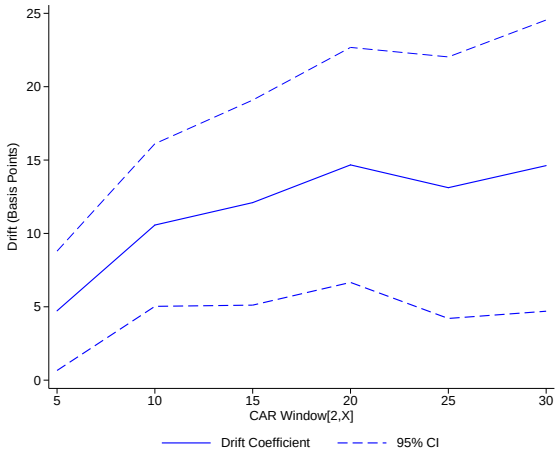



**(b) Sentiment Bias vs Net 8-K Sentiment**

**Figure 6. Correlation Between Textual Sentiment Measures.** Figure (a) plot the net press release sentiment versus the net 8-K sentiment. Net sentiment is measured as the difference in the number of positive and negative words scaled by the respective text length using the word lists of Loughran and McDonald (2011) and scaled by a factor of 100. Figure (b) plots the sentiment bias against the net 8-K sentiment, where sentiment bias is defined as the net press release sentiment minus the net 8-K sentiment. The sample is restricted to 8-Ks for which the reported event date equals the filing date and 8-Ks.

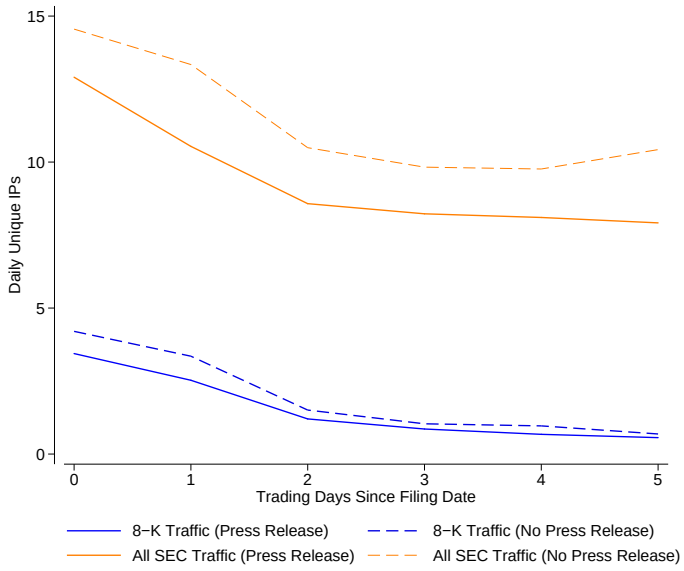



**(a) 8-Ks With Press Releases**

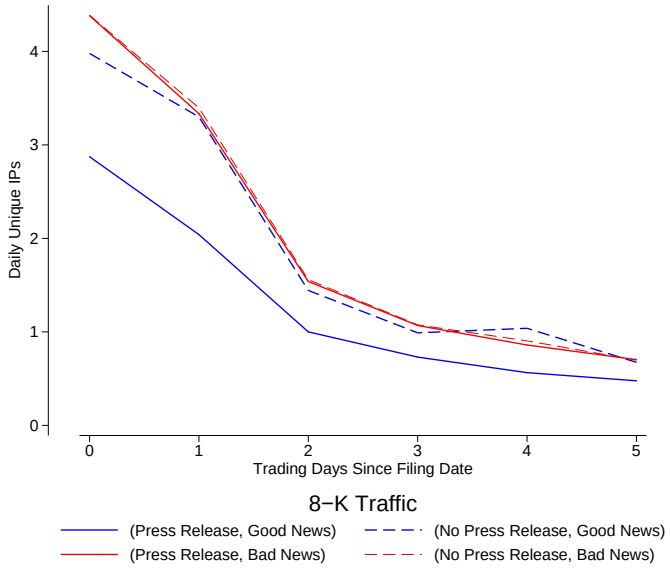



**(b) 8-Ks Without Press Releases**

**Figure 7. 8-K Price Drift Across Different Horizons.** Figure (a) plots the coefficient from the regression reported in Column 2 of Table IX which estimates Equation 3 in the text for the subsample of 8-Ks which have accompanied press releases. Figure (b) plots the coefficient from the regression reported in Column 1 of Table X which estimates Equation 3 in the text for the subsample of 8-Ks which do not have press releases. For each figure, the y-axis is measured in basis points of cumulative abnormal returns and the time horizons on the x-axis from days 2 to X after the 8-K filing date. The drift is measured in terms of a one standard deviation change in either the positive press release sentiment in Figure (a), or net 8-K sentiment in Figure (b).




**(a) Search Traffic By Availability of Press Release**




**(b) Search Traffic By Availability Press Release And Underlying 8-K Tone**

**Figure 8. EDGAR Search Traffic by Press Release Availability and Underlying 8-K Tone.** These figures plot the average number of unique daily human visitors downloading each 8-K as a function of the number of trading days since the filing date(denoted as time=0) for 8-Ks filed on EDGAR between 2008 and 2011. In Figure (a), the search traffic is plotted depending on whether the 8-K filed also contained a press release. The figure also plots overall search traffic for all documents of the firm that filed the 8-K. For Figure (b), search traffic for 8-Ks is plotted on whether there was a press release or not, and also whether the underlying 8-K negative word percent was above(good news) or below(bad news) the median using the negative word list of Loughran and McDonald (2011). The sample is restricted to 8-Ks for which the filing date equals the report date.

## Table A.I
## Frequency of 8-K Categories and Attached Press Releases

This table reports the number of 8-Ks which contain a particular item category for Panel A(the new 8-K form) and Panel B (the old 8-k form), where the new form was introduced in August 2004. The frequency of attached press releases in the 8-K is also reported. Note that the reported observation is at the item category level and not at the level of the 8-K.

### Panel A: New 8-K Aug 2004-2011

| Form 8-K Item Number | N | Press Release |
|---|---|---|
| 1.01 Entry into a Material Definitive Agreement | 132055 | 31.6% |
| 1.02 Termination of a Material Definitive Agreement | 12119 | 32.6% |
| 1.03 Bankruptcy or Receivership | 2007 | 34.7% |
| 2.01 Completion of Acquisition or Disposition of Assets | 18178 | 45.9% |
| 2.02 Results of Operations and Financial Condition | 157917 | 94.8% |
| 2.03 Creation of a Direct Financial Obligation or an Obligation under an Off-Balance Sheet Arrangement of a Registrant | 34500 | 24.2% |
| 2.04 Triggering Events That Accelerate or Increase a Direct Financial Obligation under an Off- Balance Sheet Arrangement | 2732 | 38.4% |
| 2.05 Cost Associated with Exit or Disposal Activities | 3912 | 60.3% |
| 2.06 Material Impairments | 2407 | 57.4% |
| 3.01 Notice of Delisting or Failure to Satisfy a Continued Listing Rule or Standard; Transfer of Listing | 9279 | 69.3% |
| 3.02 Unregistered Sales of Equity Securities | 27642 | 30.4% |
| 3.03 Material Modifications to Rights of Security Holders | 7089 | 44.3% |
| 4.01 Changes in Registrants Certifying Accountant | 11375 | 6.2% |
| 4.02 Non-Reliance on Previously Issued Financial Statements or a Related Audit Report or Completed Interim Review | 4740 | 36.8% |
| 5.01 Changes in Control of Registrant | 5336 | 30.2% |
| 5.02 Departure of Directors or Principal Officers; Election of Directors; Appointment of Principal Officers | 116880 | 33.2% |
| 5.03 Amendments to Articles of Incorporation or Bylaws; Change in Fiscal Year | 21279 | 26.9% |
| 5.04 Temporary Suspension of Trading Under Registrants Employee Benefit Plans | 839 | 8.0% |
| 5.05 Amendments to the Registrants Code of Ethics, or Waiver of a Provision of the Code of Ethics | 1303 | 19.3% |
| 5.06 Change in Shell Company Status | 789 | 27.9% |
| 5.07 Submission of Matters to a Vote of Security Holders | 10257 | 14.0% |
| 6.01 ABS Informational and Computational Material | 148 | 0.0% |
| 6.02 Change of Servicer or Trustee | 494 | 1.0% |
| 7.01 Regulation FD Disclosure | 102154 | 65.1% |
| 8.01 Other Events | 191965 | 50.4% |
| 9.01 Financial Statements and Exhibits | 512524 | 60.9% |

### Panel B: Old 8-K 1994-Aug 2004

| Item Number | N | Press Release |
|---|---|---|
| 1. Change in Control | 16251 | 26.1% |
| 2. Acquisition or Disposition of Assets | 29145 | 33.4% |
| 3. Bankrupcy | 3081 | 46.4% |
| 4. Change of Auditor | 13019 | 7.1% |
| 5. Other Events | 202748 | 45.2% |
| 6. Resignation of Director | 2457 | 20.7% |
| 7. Financial Statements | 223391 | 48.2% |
| 8. Change in Fiscal Year | 1647 | 19.1% |
| 9. Regulation FD Disclosure | 36390 | 67.2% |
| 11. Suspension of Trading of Employee Benefits | 206 | 6.8% |
| 12. Results of Operations | 23813 | 94.9% |

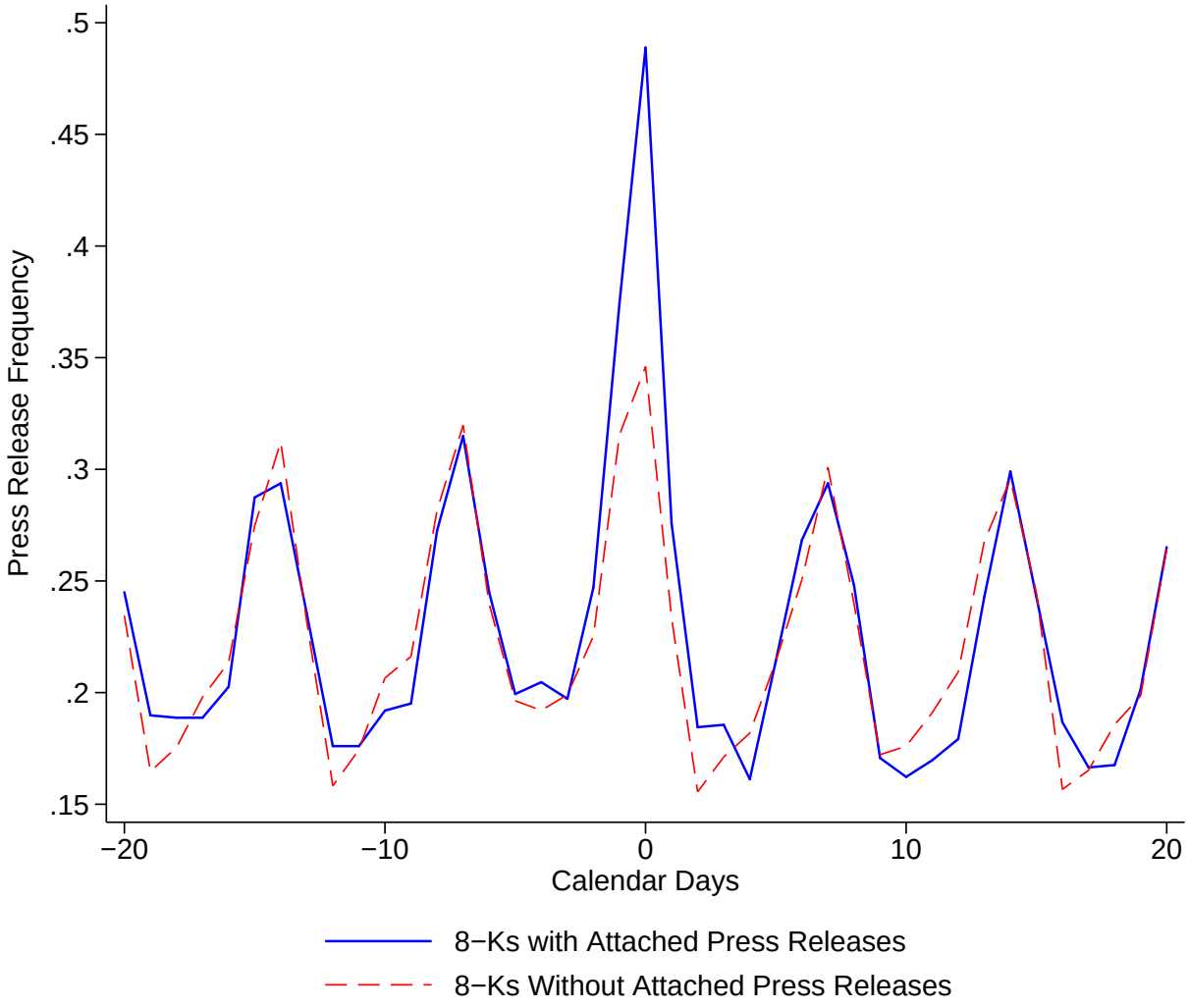



**Figure A.1. Testing for Time Series Evidence of Selection Bias In Bundling of Press Release in 8-Ks.** This figure plots the frequency of all newswires (sources: PR Newswire, Market Wire, Business Wire) relative to the 8-K filing date, which is denoted at time=0. The sample consists of 50 randomly chosen firms from the S&P 500 index for the period between 2006 to 2011. The figure separately plots the incidence of press releases for the set of 8-Ks which have or do not have an attached press release. The y-axis is the frequency $\in [0, 1]$ of a firm issued press release across the sample of 8-Ks for each day relative to the 8-K filing date.