Disclosure Timing, Information Asymmetry, and Stock Returns: Evidence from 8-K Filing Texts

Itay Goldstein

Di Wu *

First Draft: January 2015

Current Draft: June 2015

ABSTRACT

Using a unique sample of unscheduled disclosures of material corporate events constructed from over 500,000 8-K filings, we find significant heterogeneity in the *timeliness* of information disclosure, both across and within firms. We demonstrate that this timing dimension of information structure has first-order effects on information asymmetry and stock prices. Taking 10 more business days to disclose is associated with an average filing period bid-ask spread ranging from 13.78% to 34.31% wider, compared to an immediate disclosure. Contrary to predictions in existing models, the effect of timeliness on stock returns varies by the type of the event. For good news, longer disclosure lags is associated with significantly lower abnormal returns during the filing period, whereas bad news exhibit the opposite relation. Furthermore, we find evidence consistent with firms strategically choosing the level of disclosure timeliness according to the type and business category of the event.

*Itay Goldstein is the Joel S. Ehrenkranz Family Professor, Professor of Finance, The Wharton School, University of Pennsylvania, Phone: 215-746-0499, Email: itayg@wharton.upenn.edu. Di Wu is a Ph.D. Candidate of Finance, The Wharton School, University of Pennsylvania, Phone: 215-808-4835, Email: wudi1@wharton.upenn.edu.

1. Introduction

Theoretical models on the role of information in financial markets have predominantly treated information as a single-dimensional object. Specifically, Glosten and Milgrom (1985) and a vast array of subsequent studies generally suggest that, as the *quantity* of disclosed information increases, the level of information asymmetry in the securities market should decrease, because the relative advantage of informed traders is diminished due to this increased disclosure. Such predictions, however, have not been well supported by empirical evidence. For example, Fang and Peress (2009) demonstrate that the quantity of media coverage on a stock does not have a statistically significant relation to the bid-ask spread of the stock. Similarly, Zhao (2014) shows that the intensity of information disclosures in the form of current report frequency is not significantly related to common measures of information asymmetry.

Therefore it is natural to ask: is information asymmetry also affected by some other, *qualitative* dimensions of disclosure, such as timing, tone, etc.? If this is the case, then the relation between disclosure quantity and information asymmetry could be complicated by these additional dimensions. This paper empirically explores this question. Contrary to most existing research, we posit that information consists of two dimensions, quantity *and timeliness*, both of which have first-order effects on information asymmetry and stock prices. We hypothesize several channels for these effects and empirically test our hypotheses using a novel, large dataset of *unscheduled* disclosures of *materially important* corporate events.

In addition to regularly scheduled disclosures such as 10-K/Qs, public companies in the United States are required to report a wide variety of material corporate events on a more current basis, in the form of 8-K filings, or "current reports". We collect the texts of over 1 million 8-Ks filed between 1994 and 2012 and, using a simple textual analysis algorithm, construct a sample of 8-K filings that represent truly ad-hoc disclosures of unscheduled major events. For these events, we extract the exact dates when they take place, as well as the dates when they are first publicly disclosed. Surprisingly, we find significant heterogeneity in the event-to-disclosure time interval, both across and within firms, despite the SEC requirement that most 8-Ks be filed in a timely manner.[1] Taking advantage of these large variations in timing, we construct a measure of *disclosure timeliness* as the lag between the event date and the first disclosure date for each event in our sample. We use this measure as our main metric of the timing dimension of information in our empirical analysis. In addition, we conduct additional textual analyses on the filing texts, obtaining a host of other useful information such as the type of each event (e.g. good/bad news) and the business category that the event belongs to (e.g. corporate governance, financial matters, etc.). Our paper is one of the first in finance to systematically study 8-K filings, and to our knowledge, is the first paper to synthesize a timing dimension from the texts of corporate disclosures.

Armed with this new dataset, we first ask whether the observed large variations in disclosure timeliness across and within firms are economically meaningful, i.e. are these variations just random noise, or an endogenous outcome of firms' strategic decisions? If the latter, then what firm and event characteristics drive this timeliness decision? We demonstrate that the decision is predominantly affected by the characteristics of the events, and not significantly related to any particular firm characteristics. First, disclosure timeliness is significantly related to the level of *uncertainty* about the event, which we measure using the fraction of uncertain words in the filings. Firms that are most uncertain about the impact of the event, on average disclose 7 business days later compared to those that are least uncertain. Second, disclosure timeliness varies significantly by the business type of the events: there is little variation in event types with little disclosure flexibility, e.g. stock delisting events or shareholder votes. By contrast, for events with more "wiggle room" in disclosure timing, such as asset acquisitions/disposals, new projects, and personnel changes, the variation in

[1] The SEC requires that most 8-Ks be filed within 4 business days of the event. However, for many events the filing firm can ask for additional grace periods. They can also file amendments after the original deadline. In our sample, the actual lag between event and filing ranges from zero to 150 business days.

timeliness is significantly more pronounced. Furthermore, using a subsample consisting of multiple filings for the same event, we rule out firm-level attributes e.g. litigation risk as drivers of the timeliness decision. Overall, our results provide partial support for the small body of existing theories on disclosure timing by 1) establishing disclosure timeliness as a strategic decision rather than random noise, but 2) attributing heterogeneities in timeliness to event-related rather than firm-related attributes, consistent with models such as Acharya, DeMarzo, and Kremer (2011) but inconsistent with other models such as Marinovic and Varas (2014).

Next, we link disclosure timeliness to information asymmetry (hereafter referred to as IA). Theoretically, the information processing channel implies a negative relation between disclosure timeliness and IA, while the strategic timing channel impose no clear relations. Consistent with the first channel, we demonstrate that disclosure timeliness is significantly negatively related to IA around the filing period. For example, depending on regression specifications, disclosing 10 business days later is associated with a filing period bid-ask spread ranging from 13.78% to 34.31% wider compared to an immediate disclosure. This finding is robust to a wide range of empirical specifications and alternative IA measures such as the Amihud (2002) measure. In addition, this relation persists for both good and bad news types and across many business subcategories. Furthermore, in contrast to previous empirical results, the *quantity* of disclosure, when used in conjunction with the timeliness measure, becomes significantly negatively related to IA as predicted by the theoretical models. We provide further evidence for our hypothesized mechanism by showing that informed trading significantly increases on average during quarters with longer average disclosure lags.

Our next set of tests link disclosure timeliness to stock returns. Similar to IA, previous literature have divergent predictions on the timeliness-stock returns relation. Empirical literature, such as Easley and O'hara (2004) and Kelly and Ljungqvist (2012), while not directly addressing the timing dimension, do demonstrate a negative relation between IA and (expected) stock returns. Because our results link lower disclosure timeliness to higher IA, this line of research predicts a positive relation between disclosure timeliness and subsequent returns. By contrast, recent theoretical work by Guttman, Kremer, and Skrzypacz (2015) predict a negative relation between the timeliness and returns. Somewhat surprisingly, we show that the sign of this relation depends on the type of the event. For good news, our result is consistent with Easley and O'hara (2004): longer disclosure lags is associated with significantly lower (-0.43% for a 10-day lag) abnormal returns during the four-day window around the filing date. For bad news, however, longer lags is related to higher abnormal returns (0.37% for a 10-day lag), consistent with Guttman et al. (2015). This finding suggests that existing mechanisms are only partially successful in explaining the timeliness-return relation, and offers a motivation for future theoretical research. We briefly propose and test a mechanism that generates the observed relation in the upcoming version of the paper.

Our paper is related to a small but rapidly growing theoretical literature on the strategic disclosure of information *over time*. Guttman et al. (2015) propose a multi-period, multi-signal framework where uncertainty about the information endowment structure leads to a surprisingly negative relation between disclosure timeliness and stock prices. Our paper provide partial empirical support for this mechanism while motivates an extension with different event types. Additionally, we provide evidence consistent with Acharya et al. (2011), who, using a dynamic model with random information acquisition, predict that bad news tend to be disclosed earlier. Marinovic and Varas (2014) derive similar results using a model featuring litigation risk. The information processing channel outlined in our paper is motivated by Kim and Verrecchia (2001).

On the empirical front, our paper is the first in finance to examine disclosure timeliness–IA–contemporaneous stock returns link. Our paper is also related to the literature on strategic information disclosure, such as Goto, Watanabe, and Xu (2009) and Edmans, Goncalves-Pinto, Wang, and Xu (2014). The important difference is that, while these papers focus on the timing of disclosure relative to some specific events such as cross listing or the vesting of managerial equity, our paper focuses on the timeliness dimension in a general sense. Our study is one of the two papers in finance to use data on 8-K filings, the other being Zhao (2014). While that paper focuses in the intensity and quantity of 8-K filings, we focus on the timing dimension of these filings. In addition, we examine the textual content of these filings and extract additional information dimensions based on the language used in the texts.

1.1. Institutional Background

The finance and accounting literature has traditionally focused on the quantitative data (and recently textual data) contained in *regularly scheduled* corporate reports such as 10-K/Qs. However, between these regular disclosures, public companies experiencing material corporate events are required by the SEC to file a Form 8-K shortly after such events. Consequently, the 8-K filings vastly outnumber that of the 10-Ks, as we have collected over 2 million filings from SEC's EDGAR database. Despite this, little research on 8-Ks exists in finance. Utilizing this large, more timely database of a diverse set of corporate events can potentially yield new insights on a variety of finance topics from optimal disclosure to uncovering economic links.

8-K filings are defined by two key features. The first feature is the *diversity in timeliness*. The SEC requires that most 8-Ks be filed within 4 business days of the event and are therefore much more timely than 10-K/Qs. However, in reality, firms have a much larger leeway in terms of filing timing, as the SEC has a long list of exemption events and safe harbors where the firm gets more time.[2] Second, over 20% of the 8-Ks are filed in a voluntary basis (8-K Item 8.01) with no filing deadlines. Third, although late filings theoretically constitute a violation of Exchange Act and the SEC could institute an administrative proceedings against the late filer including registration revocation, these proceedings are extremely rare and are

[2]See https://www.sec.gov/rules/final/33-8400.htm for a list of exemptions.

typically aimed at recurring and egregious violations.[3] Fourth, firms can file an 8-K within the initial deadline and then disclose additional information via filing amendments later with 8-K/As, and over 15% of the filings are amendments. All of the above result in a large variation of 8-K filing timeliness both within and across firms, which we document in detail in Section 3. These variations are the main motivations for this research.

The second feature of 8-Ks is the *diversity of event types* covered by these filings. The SEC mandates over 30 types of "triggering events" after which an 8-K must be filed. These events encompass a large variety of business activities including the financing (e.g. new projects and asset acquisitions), operating (e.g. formation/termination of business and supply chain relationships), trading (e.g. stock issuances and listings), and governance (e.g. personnel changes and director nominations) aspects. These events are represented by a series of numerical codes in the filing header. This diversity of events aids our research efforts by allowing for the construction of subsamples by event type and further isolate the effect of different types on disclosure decisions.

2. Economic Mechanisms

This section develops our main hypotheses that drive the three-way relations between disclosure timeliness, information asymmetry, and stock returns. We start with intuitions on the origins of the timeliness variations and propose several economic channels through which different levels of timeliness may arise. We then outline the specific mechanisms where these channels can further affect information asymmetry in the markets. Finally, we discuss the economic link between timeliness and stock returns.

[3]See this legal communique for a discussion on consequences of late filings.

2.1. Determinants of Disclosure Timeliness

In order for disclosure timeliness to matter, it must contain some informational value itself. The vast majority of theoretical literature does not consider this value at all. The conclusion under these theories, which serves as our null hypothesis, is that timeliness is randomly determined and not systematically related to anything. However, this view has been challenged by a new and rapidly expanding breed of theories that explicitly consider the timing dimension of disclosure. These models are usually information-theoretic in that the manager receives one or more signals related to future profitability or stock prices, then decide not only on whether to disclose the signal, but whether to disclose it early or late. These signals are akin to the events in our sample. While the specific mechanisms proposed by these models differ, a common theme is that some factors related to either the 1) event, 2) firm, or 3) macroeconomy is systematically related to the manager's decision on disclosure timeliness.

First, Kim and Verrecchia (2001) and Clinch and Verrecchia (1997) relate disclosure timeliness to *event-specific* factors, namely the degree of uncertainty of the event. They propose a Cournot competition setting with two firms and show that in equilibria, firms disclose a realized signal (regardless of type) early only if it falls within their anticipations, i.e. within a certain interval around the mean. The signal is disclosed late if it falls outside the range. Their justification for this model is that, more unanticipated signals create more uncertainty and may require more scrutiny by managers, lawyers, etc before disclosure. Therefore, assuming that we can measure the degree of uncertainty or "unexpectedness" of the events in our sample, the above mechanism implies the following hypothesis:

Hypothesis 1 *On average, unanticipated 8K events with higher levels of uncertainty should take longer to disclose, regardless of type.*

Second, Marinovic and Varas (2014) relate timeliness of *bad* news disclosure to to both *event- and firm-specific* factors. Their model is based on litigation risks that a firm faces: Concealing bad news is risky because it might trigger a costly litigation process. Therefore, ex ante the firm will voluntarily disclose bad news on time. Moreover, litigation costs crowd out the disclosure of good news: firms will often choose to remain "silent" which itself serves as a good signal to the marke. The model thus suggests that bad news in our 8K sample events tend to be disclosed sooner than good news. The additional implication of this mechanism is that disclosures will be more timely for firms facing higher litigation risks, e.g. those in financial industries or those who have faced similar lawsuits before. We summarize these mechanisms in the hypothesis below:

Hypothesis 2 *In the presence of litigation risks, bad 8K events should on average be disclosed more promptly than good events. In addition, the disclosure of bad news should be clustered at the firm and industry levels: those facing higher litigation risks disclose bad events earlier.*

Third, Acharya et al. (2011) implies that the timeliness of disclosures in general is related to *macroeconomic* factors. In this setting, managers observe public signals and are endowed with the ability to delay disclosure of a private signal. In equilibrium, whether to strategically delay the disclosure depends on whether the prevailing public news is positive or negative. With good public news, the market's expectation of the stock price without disclosure might be higher than that with disclosure, and vice versa after bad public news. Therefore disclosure is on average more delayed after good public news than after bad public news. For our 8K event sample, this mechanism implies a "business-cycle" effect of disclosure clustering around bad times:

Hypothesis 3 *When the prevailing public news is bad, e.g. during economic busts, all 8K events, both good and bad, should be disclosed more promptly; during good times all type of events should take longer to disclose.*

2.2. Timeliness and Information Asymmetry

It is conceptually easier to link timeliness to information asymmetry and adverse selection in the markets. We hypothesize two channels. The first channel is related to informed trader risk first outlined in Glosten and Milgrom (1985). Consistent with this model, we assume that informed traders are more cognizant of the occurrence and the nature of an event as of the event date. We then posit a simple and intuitive extension: as more time passes without an disclosure of the event, the longer the informed traders can enjoy their information advantage. Such advantage can further incentivize informed traders to engage in additional information-gathering activities, thereby further increasing their advantage. Correspondingly, participation and trading intensity by informed traders increases during this period. When the information is finally revealed, however, market participants, aware of this heightened level of advantage, demand higher compensation for this informed trading risk. This leads to higher bid-ask spreads and illiquidity measures. We summarize the above discussion in the following hypothesis:

Hypothesis 4 *A longer disclosure lag leads to a longer period of information advantage and higher information seeking by informed traders. This leads to higher participation by informed traders and higher IA measures during the filing period.*

Although not yet explored by existing theories, an alternative hypothesis is that a less timely disclosure increases the probability of leaks. For example, someone inside the firm might talk to the media, and the longer it has been since the event date, the more likely that the media is able to find something. Such leaks can be viewed as partial public disclosures of the event. Therefore, when the event is eventually disclosed, compared to the case of immediate disclosure, the public is more informed due to the interim leaks. This reduces the informed trader risk, thereby reducing IA measures.

Hypothesis 5 *A longer disclosure lag leads to a higher probability of leaks, which reduces the information advantage of informed traders. This leads to lower IA measures during the filing period.*

The null against both hypotheses above is that timeliness of information disclosure does not have any informational value whatsoever. In this case, our measure of timeliness would not be significantly related to filing period IA measures.

2.3. Timeliness and Filing Period Abnormal Returns

We develop three competing hypotheses regarding the timeliness-returns relation. The first hypothesis is derived from previous empirical research: Easley and O'hara (2004) and Kelly and Ljungqvist (2012) demonstrate that information asymmetry is negatively related to expected stock returns, possibly because stocks with higher IA have higher costs of capital. Then, if Hypothesis 4 is true, this line of research predicts a positive relation between disclosure timeliness and stock returns: more prompt disclosures reduce adverse selection, leading to lower costs of capital demanded by market participants and therefore higher ex-post prices.

By contrast, recent theoretical work by Guttman et al. (2015) predict a negative relation between the timeliness and ex-post returns. This is due to a multi-signal framework where both the manager and the market is uncertain about the information endowment structure. Here, later disclosures are interpreted more favorably because they lead to a reduced outcome set. This result is true even if the signal has no information value (i.e. the event is meaningless and does not change any existing or expected cash flow of the firm). Therefore, in this setting, firms that take longer to disclose events have higher subsequent realized stock returns.

A third hypothesis is that, because most empirical research that links disclosure quantity to returns has failed to link it with IA, perhaps our timeliness measure has zero informational value or is mismeasured. In this case, the relation between our timeliness measure and returns should not be statistically different from zero. We summarize all three views in the following hypothesis:

Hypothesis 6 *If a longer disclosure lag leads to higher IA and IA is positively related to the cost of capital, then a longer disclosure lag leads to lower ex-post abnormal stock returns. Alternatively, in a multi-signal framework with uncertainty about the information endowment structure, later disclosure leads to higher ex-post returns.*

We first discuss our data and variable construction methodology in detail in the next Section. Using the new dataset, we then develop several schemes to test Hypotheses 1 to 6 in Section 4 below.

3. Data and Methodology

We obtain all 8-Ks filed from January 1994 through December 2012 from SEC's EDGAR database using a customized web crawling algorithm. Beside the main texts, most 8-K filings are accompanied by a highly informative HTML header, which contains the date that the event has taken place, the filing date, as well as the numeric codes identifying the type of the triggering events. Using a customized textual analysis algorithm, we parse these data from the headers and merge them back with the main texts to create one unique date-stamped, item-coded text string per 8-K filing. We then use the following criteria to construct our sample of 8-Ks:

1. EDGAR identifies firms that file 8-Ks using Central Index Key (CIK). We use the WRDS CIK-PERMNO file to match CIK with PERMNO from the CRSP-COMPUSTAT Merged database. We exclude all firms for which we are not able to match CIK to PERMNO's.
2. Many firms file multiple 8-Ks and amendments for the same event. We coalesce all filings for the same event into a single 8-K text string.
3. Our tests use a series of control variables such as book-to-market ratio, turnover, etc. Descriptions of control variables are included in Appendix **??**. We exclude all firms for which we do not have these data for the years when the data are not available.
4. In order to mitigate the effect of bid-ask bounces on the spread and returns, the stock price should be at least $3.00 on the filing date.

Although the majority of 8-Ks filings describe unscheduled events, firms are required to file an 8-K when they announce or revise their quarterly earnings. In addition, some 8-Ks are related to other events that are scheduled in advance, such as annual shareholder meetings. We eliminate pre-scheduled events from the sample because market participants are likely to be informed about these events prior to the filing date. Specifically, we drop all 8-Ks filed with an item code of 2.02, as well as filings in the "other events" catetory that contain keywords from a list of words associated with pre-scheduled events, particularly earnings announcements.[4] We also drop all filings with the event date reported as the end of fiscal quarters.

[Insert Table 1 here]

The final sample contains 395,711 filings between 1994 and 2012, and 10,827 unique firms. Table 1 presents a summary of the sample that we use in our analysis. The mean market

[4]The keyword list contains words and phrases such as *scheduled*, *earnings announcement*, etc and is available at available at https://fnce.wharton.upenn.edu/profile/1661/. Not deleting filings with matching keywords does not change our results.

value is \$2.818 billion and the book-to-market ratio has a mean value of 0.603. This sample represents ad hoc disclosures of material events that do *not* take place in the ordinary course of business.[5] This quasi-random information arrival and disclosure process is therefore ideally suited for our analysis of timeliness and other dimensions of information disclosure.

An additional advantage of using the HTML headers is that they reveal the detailed business category that a disclosed event belongs to, i.e. exactly what the event is about. This is categorized by a series of numerical codes. For example, a firm that signs a new supplier contract would file an 8-K under the content code 1.01 (business–new contract), and a sudden departure of a key board member would trigger a filing under content code 5.02 (governance–personnel changes). There are eight major categories, which further divide into 31 subcategories encompassing a wide variety of business scenarios. We briefly describe each subcategory in Appendix A. This pre-classification of events represents an uniquely attractive feature of 8-Ks, as it eliminates the need for the researcher to classify the content of the filings with machine learning-based methods such as Naïve Bayes or Latent Dirichlet Analysis, thereby greatly increasing the depth of the analysis by directly quantifying the business content of the texts.

3.1. Construction of Disclosure Timeliness

For each firm i that discloses an event of type j, our textual analysis yields the date when the j has taken place ($EventDate_{i,j}$) and the date that the 8-K is filed ($FilingDate_{i,j,t}$). We

[5] For example, according to SEC's web site: if a company takes out a five-year loan with a bank or signs a long-term lease, and the loan or lease is material to the company, the agreement must be reported. But if a retailer already has a chain of stores and signs a lease for one more, the new lease generally would be in the ordinary course of business and would not be reported. Similarly, if an agreement simply expires according to its terms, that termination would not need to be reported. However, if the agreement terminates prior to the date on which it would otherwise expire, that event would need to be reported.

therefore compute the timeliness measure as the lag between the two dates in business days, which we term as the Disclosure Lag (DL):

$$\widehat{DL}_{i,j,t} = FilingDate_{i,j,t} - EventDate_{i,j}$$

In addition, many firms file multiple 8-Ks or 8-K amendments for the same event. This can be seen as either a way to circumvent the SEC filing deadlines or as disclosure of signals that are revised over time. We coalesce all filings at the firm-event level, with one filing date per firm-event. In case of multiple filings, we adjust the filing date to reflect the *first* date that the event is disclosed to the public.[6] Specifically,

$$DL_{i,j,t} = \min_t \{FilingDate_{i,j,t}\} - EventDate_{i,j} \tag{1}$$

Fig. 1 displays the distribution histogram of DL_j across all firm-events. Here it is evident that, while most firms do comply with the 10-business-day deadline mandated by the SEC, there is significant variation in DL within the 10-day window. The mean value is DL is 4.05 business days and the standard deviation is 11.10 business days. In addition, despite the deadline, close to 10% of the filings takes longer than 10 business days after the event. This heterogeneity in disclosure timeliness motivates our empirical studies in the next Section.

Furthermore, Table 2 reports average distribution statistics both within firms and within each business category. We first compute the mean and standard deviation of $DL_{j,k}$ for each firm k and event j, then compute the average value of the means and standard deviations across k. To ensure the robustness of the result, we partition our dataset into different subsamples according to 1) total number of filings, 2) maximum value DL, 3) market cap, and 4) book-to-market ratio, and report the average statistics for each subsample in Panel

[6] An alternative adjustment method is to assume that later filings contain information that are already known at the first filing but are withheld by the firm. In this case, an event with multiple filings should have a longer disclosure lag compared to an event with a single filing, e.g. $DL_{i,j} = \text{Average}_i(FilingDate_{i,j}) - EventDate_j$. Adjustments using either the mean or median of filing dates has little effect on our result.

A of the Table. The average within-firm standard deviation of DL is large at 16.13 business days, and this large variation persists even when we restrict the samples to firms that file very frequently (more than 100 8-K filings) and firms with smaller maximum DL (less than 100 business days). In addition, the variation in DL is similarly large for both large firms (market cap greater than \$6 billion) and value firms (book-to-market ratio greater than 1.00). This result also indicates that event-specific heterogeneities probably affects variations in DL more so than pure firm-level heterogeneities, an observation we revisit in the following subsection.

Panel B of Table 2 reports average distribution statistics by major business categories. Here again, the variation in DL is large, with the average standard deviation of around 15 business days. A more interesting fact is that this variation is not uniform across categories. For example, the average values of both mean and standard deviation of DL in the category related to trading status of a firm's stock (2.96 and 4.92 respectively) are much lower than those for the category related to financial matters such as asset acquisition and disposal (3.97 and 10.20). Intuitively, a firm has much more leeway in timing in the disclosure of financial matters than of trading-related matters, because the latter is dictated by the exchange that the firm's stock trades on. We revisit this observation in Section 4.1.

3.2. Measures of Filing Period Information Asymmetry and Abnormal Returns

The 8-K filings are available to investors who pay a fee through a qualified provider within ten minutes after filing and for free to other investors within the next one or two days after the filing date (see Griffin (2003) for details about EDGAR access). Most studies define the filing period as a four-day window from days 0 through +3 relative to the filing dates. To facilitate comparability, we also define filing period information asymmetry measure, $IA_{i,t}$, as the average bid-ask spread or the Amihud (2002) illiquidity measure in this window. We compute the spread and illiquidity measures for each date t as follows:

- $Spread_{i,t} = 100 \times \dfrac{ask_{i,t} - bid_{i,t}}{(ask_{i,t} + bid_{i,t})/2}$
- $Amihud_{i,t} = \ln\left(1 + \dfrac{|R_{i,t}|}{P_{i,t} \times \text{shares outstanding}_{i,t}/100000}\right).$

We compute filing period abnormal return $r^e_{i,t}$ as follows:

$$r^e_{i,t} = \prod_{t=0}^{3} ret_{i,t} - \prod_{t=0}^{3} ret_{vw,t}\ , \tag{2}$$

where $ret_{i,t}$ and $ret_{vwi,t}$ are the returns on stock i and on the CRSP value-weighted index on date t.

4. Empirical Tests and Results

We first test Hypotheses 1 to 3 advanced in Section 2 to explore the systematic drivers of the large variation in disclosure timeliness across and within firms. We then test Hypotheses 4 and 5 and examine the relation between timeliness and information asymmetry, both in the full sample and in a variety of subsamples. Finally, we examine the relation between timeliness and filing period abnormal returns, and the extent to which this relation holds for different event types and categories.

4.1. Determinants of Disclosure Timeliness

Under Hypotheses 1 to 3, DL is strategically chosen. The decision can be driven by factors related to 1) the firm, 2) the event, or 3) the economy. In the data, economic mechanisms outlined in these hypotheses imply that variations in DL reveals information about the firm and/or the events. In other words, if a factor indeed affects a firm's timing decisions, then it should be related to DL in a statistically significant fashion.

To test Hypotheses 1 and 2, we first use textual analysis on each 8-K filing to infer 1) whether each event is good/bad news and 2) the level of uncertainty about the event.

Specifically, we infer the nature of the events using a bag-of-word approach first employed by Loughran and McDonald (2011) and Jegadeesh and Wu (2013) for 10-K filings and IPO prospectuses. We start from the Loughran and McDonald (2011) lists of Positive, Negative and Uncertain words that applies specifically to financial disclosures.[7] We then compute the tone and uncertainty Scores for each 8-K filing i as follows:

1. First for each word v, assign a weight w_v such that $w_v = \ln \frac{N}{df_v}$ where df_j is the number of 8-Ks where word v occurs at least once. This inverse document frequency weighting scheme is first proposed by Manning and Schütze (1999) and is found to be useful in a variety of document retrieval applications.[8]

2. Collect the lexicons of J_{POS} positive words, J_{NEG} of negative words, and J_{UNC} of uncertain words. Also collect the total number of words T_i for each 8-K i.

3. The Tone Score, which measures whether the event is positive/negative, is

$$Score_i^{TONE} = \frac{1}{T_i}\left(\sum_{j=1}^{J_{POS}} w_{i,j} F_{i,j}^{POS}\right) - \left(\sum_{j=1}^{J_{NEG}} w_{i,j} F_{i,j}^{NEG}\right),$$

where $F_{i,j}^{NEG}$ and $F_{i,j}^{POS}$ are the frequency of each negative/positive word within each 8-K filing.

4. The Uncertainty Score, which measures how uncertain the firm is about the event, is similarly computed as

$$Score_i^{UNC} = \frac{1}{T_i}\left(\sum_{j=1}^{J_{UNC}} w_{i,j} F_{i,j}^{UNC}\right),$$

where $F_{i,j}^{UNC}$ is the frequency of each uncertain word within each 8-K filing.

[7] Available at http://www3.nd.edu/∼mcdonald/Word_Lists.html

[8] Alternative term-weighting methods include equal weights and market-based weights used in Jegadeesh and Wu (2013). The weighting scheme used has little impact on our results.

We then examine the validity of Hypotheses 1 and 2 by fitting the following regression:

$$DL_{i,j,t} = a + bScore_t^{TONE} + cScore_t^{UNC} + dITEM_t + \gamma X_{i,t} + f\mathcal{F}_{i,j,t} + \epsilon_{i,t}, \tag{3}$$

where $ITEM_t$ is a collection of dummies of 8 major business item codes that we capture from the HTML headers. In addition, we include a variety of fixed effects in the vector $\mathcal{F}_{i,j,t}$:

- Firm (PERMNO) fixed effects
- Industry fixed effects constructed at the 2-digit SIC code level
- Three time fixed effects constructed as the year, day of the week, and month of the year of the filing dates
- Business category fixed effects constructed at the level of detailed subcategories described in Appendix A

$X_{i,t}$ is our list of firm-specific characteristics described in Appendix A. $Size$ and $Turnover$ are intuitively related to IA: large firms and liquid stocks have more trading activity, thus less adverse selection and smaller spreads. BM, PE and ROA relates firm's fundamental performance. Firms with better cash-flow performance attracts more trading activity, lowering IA measures. *ED_Ret* and *Accruals* reflect recent events. The abnormal return around the event date and acts as an additional proxy for the type of the event. For example, if informed traders are aware that an event is bad news, then they might trade around the event dates, lowering returns. High values of *Accruals* are generally considered bad news either because it indicates an increase in working capital that may be due to bad business conditions or due to earnings manipulation. $Inst\_Ownership$ and $Info\_Intensity$ reflects the information structure of the stock. Firms with high institutional ownership tend to be larger and more liquid, and thus have lower IA measures. Moreover, $Info\_Intensity$ proxy for the *quantity* of information disclosure. Zhao (2014) demonstrates that the quantity of disclosures is not significantly related to IA measures.

[Insert Table 3 here]

Table 3 reports regression coefficients. First, timeliness is significantly related to the type of the event. The coefficient for $Score^{TONE}$ is significantly negative. Therefore, on average firms take longer to disclose negative news than positive news. This result is somewhat surprising and is not consistent with Hypothesis 2, Marinovic and Varas (2014) and Beyer and Dye (2012), who show that, in the presence of some external concerns such as litigation costs or reputation, it is preferable to disclose bad news earlier than good news. This result can also be compared to Edmans et al. (2014), who find that firms are strategic with the disclosure timing, in order to coincide these disclosures with the vesting of managerial equity. Current theoretical models do not address this strategic delay of bad news disclosure that we document. Therefore, the exact mechanism that generates our result is an interesting theoretical research topic, with interesting questions such as whether firms are deliberately trying to manipulate the information, or whether such delays are due to some genuine information processing constraints for bad news.

More importantly, the coefficient for $Score^{UNC}$ is significantly positive, indicating that more unanticipated and uncertain events do take longer to disclose. For example, an one-standard-deviation increase in uncertainty is associated with a DL that is over 7 business days longer. This result is consistent with theoretical predictions of Kim and Verrecchia (2001) who, using a multi-period signaling model, show that firms facing more extreme and uncertain events disclose in later periods. Empirically, our result is also consistent with the fact that firms facing a more uncertain event tend to spend more time with legal and compliance teams on the wording of the disclosure, in order to minimize the possibility of misrepresenting the events. This coefficient is significant even after controlling for $ToneScore$ in the regressions, indicating that the effect of uncertainty on firms disclosure timing decision is incremental to that of the type of the events.

Next, to firmly establish DL as a strategic decision, we explore the relation between DL and the business content of the events. Interestingly, this relation is heterogeneous across business contents. For example, the relation is not statistically different from zero for disclosures of trading-related events (item code 4.xx), but is significantly positive for financial matters (item code 2.xx), and is significantly negative for Reg FD disclosures (item code 7.xx). A more intuitive explanation can be found in Panel B of the table, where we regress DL on the full set of detailed item subcategory dummies. Here for categories with little disclosure flexibility, e.g. stock delisting events or shareholder votes, there is lower variation of DL. By contrast, for events with more "wiggle room" in disclosure timing, such as asset acquisitions/disposals, new projects, and personnel changes, the variation in timeliness is significantly more pronounced and the average DL is significantly higher. The interpretation is straightforward: it is easier to manipulate the content or timing of the events that are more within a firm's control, such as new projects, than events that are dictated by external parties such as stock exchanges. The former group has more information value than the latter, because, regardless of the firm's disclosure policy, the latter group is likely to be disclosed anyway by external parties, whose timing decisions is outside the control of the firm. This result again suggests the possibility that our timing dimension can be further interacted with other information dimensions obtained through textual content analysis, and highlights the usefulness of such methods in broadening the scope of financial research.

Then, to test Hypothesis 3, we first plot the average DL over time and across different industry groups (first digit of the SIC codes) in Figure 3. The top panel displays the time series of average DL across all firms, as well as the subsample average DL for good and bad news. Each filing i is classified as good news if its tone score is positive, i.e. $Score_i^{TONE} > 0$, and bad news if the score is negative. For these subplots, documents with zero tone scores are eliminated from the sample. For the bottom panel, the sample is divided into 8 subsamples based on the first digit of the filing firm's SIC code. As a proxy for the tone of the prevailing public news, real price levels of the S&P 500 index are plotted on the secondary axis.

The top panel of this Figure seems to be consistent with some clustering of disclosures around bad public news. Specifically, average disclosure takes noticibly longer during the economic booms of the 1990s, and again during the recovery from the latest financial crisis. In addition, disclosures become more timely during the last two recessions, and the drop in DL happened before the passage of the Sarbanes-Oxley Act of 2002. This pattern is consistent with both good and bad events, with bad news on average taking significantly longer to disclose. Taken together, these evidence provide support for Acharya et al. (2011) and Hypothesis 3, namely that disclosure timeliness is affected by prevailing public news, which is in turn affected by macroeconomic factors.

This result is further confirmed by noting the rightmost column of Table 3, which inserts two additional macro-related variables into Regression (3):

- $UnEmp$: latest monthly rate of unemployment obtained from the Bureau of Labor Statistics.
- $Recession$: a dummy variable which equal to one if meeting date t falls within a NBER-designated recession period.

This column demonstrate a strong procyclicality associated with average DL: the estimate for $UnEmp$ is significantly negative, indicating that during periods of bad employment (busts), disclosure is on average more prompt. Moreover, during economic recessions, an average disclosure is 4.71 days sooner than normal times.

Lastly, the bottom panel of Figure 3 shows no evidence of disclosure clustering around particular industries. The same pattern discussed above holds for all major industry groups, and industries that are particularly succeptible to litigation risk, e.g. the financial service industry (group 6 in the figure), do not exhibit any markly different pattern. Therefore, our sample is not consistent with the litigation risk mechanism on the industry level either.

4.2. Timeliness and Information Asymmetry

We have established DL as a strategic decision and related it to a host of event-specific characteristics in the previous subsection. Now we examine the relation between filing period IA and disclosure timeliness using the following regression. For firm i, event type j and filing date t:

$$IA_{i,t} = a + b \times DL_{i,j,t} + \gamma X_{i,t} + f \times \mathcal{F}_{i,j,t} + \epsilon_{i,t}, \tag{4}$$

where $IA_{i,t}$ is the measure of filing period information asymmetry, which is either the average bid-ask spread or the Amihud (2002) illiquidity measure in the four-day window from days 0 through +3 relative to the filing dates. $X_{i,t}$ and $\mathcal{F}_{i,j,t}$ are the same controls and fixed effects used in Regression (3) before.

[Insert Table 4 here]

Table 4 reports the coefficient estimates. All standard errors are heteroskadasdicity-robust (White), and clustered around the firm level. The numbers in brackets are t-Statistics. The coefficient for DL is significantly positive across all specifications. For example, in Column (6) of Panel A, which reports the results with full set of controls and fixed effects, the coefficient for DL is 0.0148 with a t-Statistic of 4.60. In this setting, a 10-day increase in DL (roughly a 0.6-standard-deviation change) is associated with a 13.78% increase (0.1480) in the filing period bid-ask spread relative to the long-term average of 1.0740. Similar magnitudes and statistical significance can be found for other specifications, as well as the Amihud (2002) measure. This result is consistent with Hypothesis 4 above and inconsistent with Hypothesis 5, and suggests that longer disclosure lags are possibly associated with higher information advantages of informed traders.

To further reconcile our evidence with Hypothesis 4, we directly assess the level of informed trading during the disclosure lag period. We report the results for this analysis in the upcoming version of the paper.

It is also worth noting that, when used in conjunction of our DL measure, the coefficient for $Info\_Intensity$ becomes significantly negative, instead of zero as shown in previous empirical studies. Here in our setting, the *quantity* of disclosure behaves exactly as predicted by the theoretical models. This evidence supports our hypothesis of the multi-dimensional nature of information: more disclosure is beneficial for reducing adverse selection, but only if these disclosures are made in a timely fashion. The trade-off between quantity and timing is an important managerial decision and an interesting topic for further theoretical exploration.

4.3. Timeliness and Filing Period Abnormal Returns

This subsection examines the relation between DL and subsequent stock returns. We develop three competing hypotheses regarding this relation. The first hypothesis is derived from previous empirical research: Easley and O'hara (2004) and Kelly and Ljungqvist (2012) demonstrate that information asymmetry is negatively related to expected stock returns, possibly because stocks with higher IA have higher costs of capital. Because our results link a higher DL to higher IA, this line of research predicts a negative relation between DL and stock returns.

By contrast, recent theoretical work by Guttman et al. (2015) predict a positive relation between the DL and returns. This is due to a multi-signal framework where both the manager and the market is uncertain about the information endowment structure. Here, later disclosures are interpreted more favorably because they lead to a reduced outcome set. This result is true even if the signal has no information value (i.e. the event is meaningless and does not change any existing or expected cash flow of the firm). Therefore, in this setting, firms with longer DL have higher subsequent realized stock returns.

A third hypothesis is that, because most empirical research that links disclosure quantity to returns has failed to link it with IA, perhaps our timeliness measure has zero informational value or is mismeasured. In this case, the relation between our DL measure and returns should not be statistically different from zero.

We test the above hypotheses by fitting the following regression of filing period abnormal return from days 0 through +3 relative to the filing dates, on DL and a variety of controls:

$$r^e_{i,t} = a + bDL_{i,j,t} + cScore^{TONE}_{i,t} + \gamma X_{i,t} + f\mathcal{F}_{i,j,t} + \epsilon_{i,t}, \tag{5}$$

where $r^e_{i,t}$ is the filing period abnormal return computed according to Equation (2).

The regression results are reported in Table 5. First, Panel C of this table reports coefficient estimates for the combined sample of both positive and negative news. The first three columns report estimates without additional variables and the three rightmost columns display results with control variables. The unconditional mean filing period abnormal return is not statistically different from zero in all specifications. Moreover, the coefficient estimate for DL is not statistically significant either. This indicates that, although variations in DL is associated with variations in filing period IA measures, they are not necessarily so with abnormal returns when the nature if the events are not fully accounted for.

To further examine the return-timeliness relation at different event types, we fit Regression (5) separately for good and bad news using the Tone Scores computed in the previous subsection. Here we denote good news as filings with a positive Tone Score and bad news as filings with a negative Tone Score, and report the respective results in Panels A and B. Loughran and McDonald (2011) and Jegadeesh and Wu (2013) demonstrate that the Tone Score reliably predicts the type of the news, and here our result further confirms this for

8-K filings. The coefficient for $Score^{TONE}$ is indeed significantly positive. More importantly and perhaps surprisingly, we show that the DL-return relation is different for different news types. For good news, our result is consistent with Easley and O'hara (2004): longer disclosure lags is associated with significantly lower (-0.43% for a 10-day lag) abnormal returns during the four-day window around the filing date. For bad news, however, longer lags is related to higher abnormal returns (0.37% for a 10-day lag). This evidence is consistent with Guttman et al. (2015). Moreover, the opposite signs of the relation does not depend on the particular good/bad news sample cut. Regressions in Panel C of the table also include an interactive term of $DL \times Score^{TONE}$, and the coefficient estimate for this term is also significantly negative at -0.025. This means that, the DL-return relation is increasingly more negative as the news gets more positive. This finding suggests that existing mechanisms are only partially successful in explaining the timeliness-return relation, thus substantiating the existence of the timing dimension of information beside the quantity dimension. This result also offers a motivation for future theoretical research, where models such as Guttman et al. (2015) can be extended to specifically incorporate the type of events.

5. Event Endogeneity Concerns and Robustness Checks

Because our sample consists of only unscheduled disclosures of corporate events outside the firms' normal line of business, it is safe to assume that for each firm, the arrival process for each event/signal is random and uncorrelated with factors unobservable to researchers. However, one can still argue that the endogeneity issue is not completely resolved, because in this setting, an airtight identification strategy requires another random component: In addition to the the arrival process of an event, a firm's decision to choose the disclosure lag following the event's arrival should also be random. However, this process is endogenous in our setting, as firms are likely to strategically choose the disclosure lag in order to optimize market responses or satisfy some other strategic objectives. If there are factors that are both firm- and time-varying, *and* are correlated in the same direction with both the disclosure lag decision and IA measures around the filing date, then the coefficient estimates could still be positively biased.

Nevertheless, we argue that this case is unlikely. Recall that Table 2 shows significant variation of DL within each firm *and* within each type of the event. This suggests that the decision to optimally choose DL is based on unobservable *event-level* heterogeneity rather than unobservable firm heterogeneities. To address this issue, our textual analysis step classifies each event into over 30 categories as described in the previous Section. We then include these event category-specific fixed effects in our regressions. Because these fixed effects result from a very granular classification scheme, they probably control for the event-level heterogeneities satisfactorily. In addition, thanks to the large number of filings per firm, we control for other heterogeneities by including firm, industry and a variety of time fixed effects in our regressions. The combined effect of the above steps probably serves to diminish endogeneity concern in our analysis.

6. Conclusion and Next Steps

We demonstrate the multi-dimensionality of information structure using a unique sample of unscheduled disclosures of material corporate events constructed from over 500,000 8-K filings. Specifically, we demonstrate that the *timeliness* of information disclosure has first-order effects on both information asymmetry and abnormal returns during the disclosure period. Longer disclosure lags are associated with higher information asymmetry during the filing period. We further posit two channels for this effect and isolate one particular channel: the longer it takes for an informative event to be publicly disclosed, the longer informed traders can enjoy their information advantage. Such advantage can further incentivize informed traders to engage in additional information-gathering activities, thereby further increasing their advantage. Market participants therefore demand higher compensation for this informed trading risk, widening the bid-ask spreads. In general, this feedback loop should result in a negative relation between disclosure timeliness and measures of information asymmetry in the markets.

On the return front, we provide reconciling evidence for two divergent predictions in the previous literature. For good news, our result is consistent with Easley and O'hara (2004): longer disclosure lags is associated with significantly lower abnormal returns during the four-day window around the filing date. For bad news, however, longer lags is related to higher abnormal returns, which is consistent with Guttman et al. (2015). This finding suggests that existing mechanisms are only partially successful in explaining the timeliness-return relation, and offers a motivation for future theoretical research. We briefly propose and test a mechanism that generates the observed relation in the upcoming version of the paper.

The 8-K sample that we construct has several unique advantages not available in other textual disclosure sources. First, the date of the event is always disclosed. Second, the firm is also required to pre-categorize the type of the event according to a detailed classification scheme, which provides an excellent training source for machine learning-based topic classification schemes. We take advantage of these unique features in several concurrent studies in finance and operations management.

References

Acharya, V. V., DeMarzo, P., Kremer, I., 2011. Endogenous information flows and the clustering of announcements. American Economic Review 101, 2955–2979.

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

Beyer, A., Dye, R. A., 2012. Reputation management and the disclosure of earnings forecasts. Review of Accounting Studies 17, 877–912.

Easley, D., O'hara, M., 2004. Information and the cost of capital. Journal of Finance 59, 1553–1583.

Edmans, A., Goncalves-Pinto, L., Wang, Y., Xu, M., 2014. Strategic news releases in equity vesting months. Working Paper, London Business School and NBER .

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

Glosten, L. R., Milgrom, P. R., 1985. Bid, ask and transaction prices in a specialist market with heterogeneously informed traders. Journal of Financial Economics 14, 71–100.

Goto, S., Watanabe, M., Xu, Y., 2009. Strategic disclosure and stock returns: Theory and evidence from us cross-listing. Review of Financial Studies 22, 1585–1620.

Griffin, P., 2003. Investor response to form 10-k and form 10-q edgar filings. Review of Accounting Studies 8, 433–460.

Guttman, I., Kremer, I., Skrzypacz, A., 2015. Not only what but also when: A theory of dynamic voluntary disclosure. American Economic Review, Forthcoming .

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

Kelly, B., Ljungqvist, A., 2012. Testing asymmetric-information asset pricing models. Review of Financial Studies 25, 1366–1413.

Kim, O., Verrecchia, R. E., 2001. The relation among disclosure, returns, and trading volume information. Accounting Review 76, 633–654.

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

Manning, C. D., Schütze, H., 1999. Foundations of Statistical Natural Language Processing. MIT Press.

Marinovic, I., Varas, F., 2014. No news is good news: Voluntary disclosure in the face of litigation. Working Paper, Stanford University .

Sloan, R. G., 1996. Do stock prices fully reflect information in accruals and cash flows about future earnings? Accounting Review 71, 289–315.

Zhao, X., 2014. Not only what but also when: A theory of dynamic voluntary disclosure. Working Paper, University of Texas at Dallas .

Appendix A Variable Definitions

We use the following control variables in our regressions. We obtain accounting data from Compustat, price and return data from CRSP, and data on institutional trading and ownership from Thomson Reuters.

- Turnover: Natural logarithm of the number of shares traded during the period from six to 252 trading days before the filing date divided by the number of shares outstanding on the filing date.
- BM: The ratio of the book value of equity as of the previous fiscal year end.
- PE: The ratio of price to book earnings as of the previous fiscal year end.
- ROA: The ratio of (net income + interest expense) to average total assets.
- ED-Ret: The return over the three-day window [t-1, t+1] around the event date minus the CRSP value-weight index return over the same period.
- Accruals: We compute accruals as in Sloan (1996). Specifically, accruals is one-year change in current assets excluding cash minus change in current liabilities excluding long-term debt in current liabilities and taxes payables minus depreciation divided by average total assets.
- IdVol: The standard deviation of the firm-specific component of returns estimated using up to 60 months of data as of the end of the month before the filing date. We estimate volatility for all firms with at least 12 months of data during this 60-month period.
- Inst_Ownership: Number of shares owned by 13F institutions divided by total shares outstanding as of the previous quarter end.
- Info_Intensity: We compute information intensity as in Zhao (2014) as the average number of 8-Ks filed in the past 6 months as of the last month end.

Figure 1. Distribution of Mean Disclosure Lag for All 8-K Filings

[Insert description here]

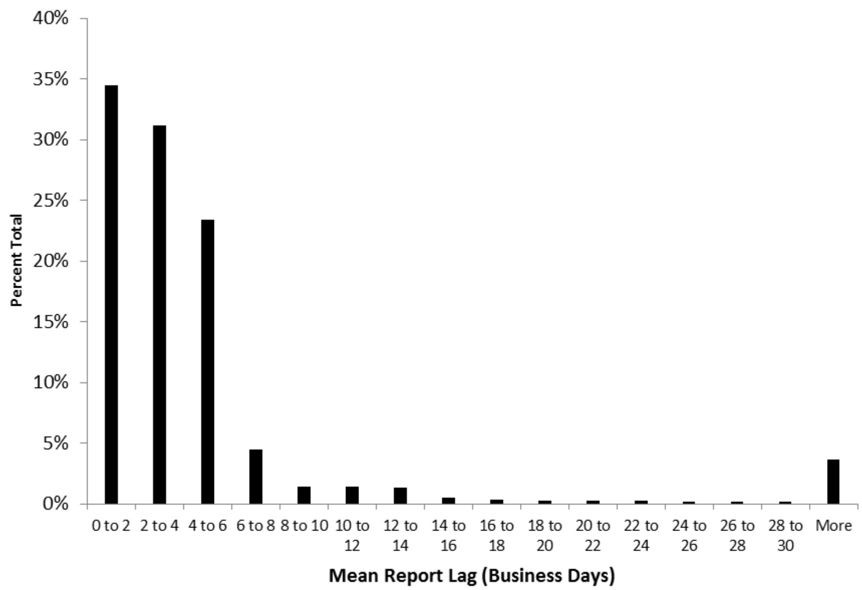


Figure 2. Distribution of 8-K Textual Properties

[Insert description here]

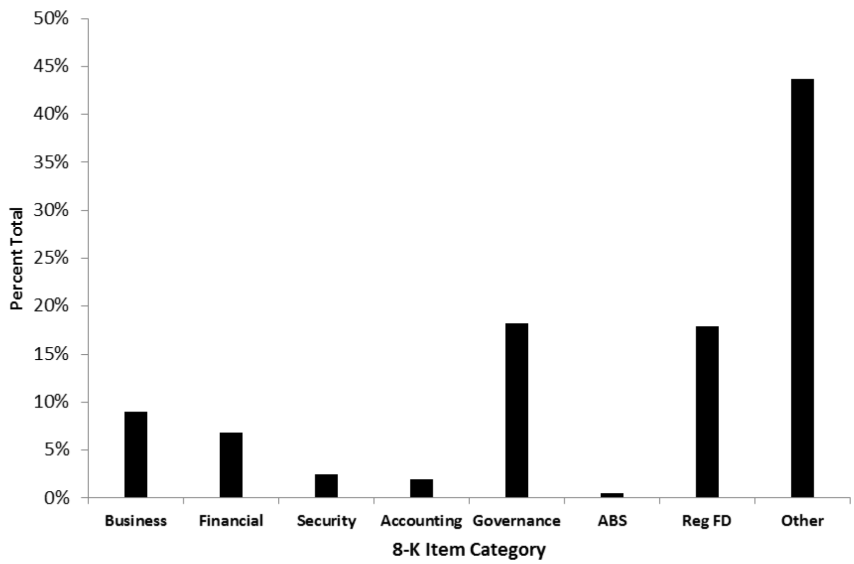
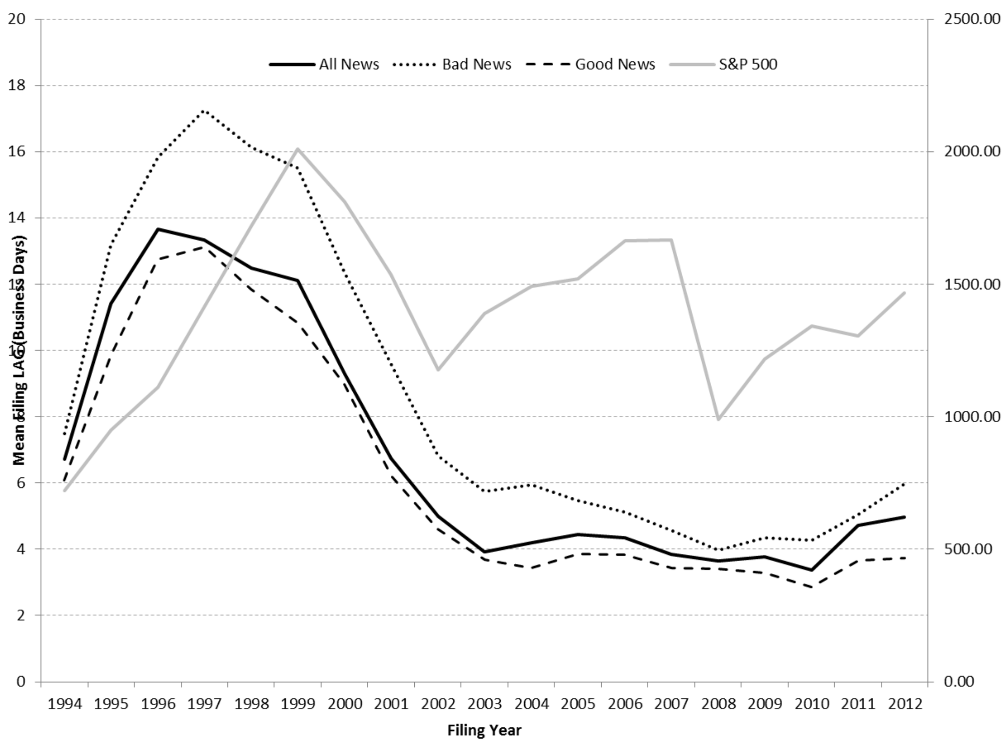


Figure 3. Mean Disclosure Lag Over Time and By Industry

[Insert description here]



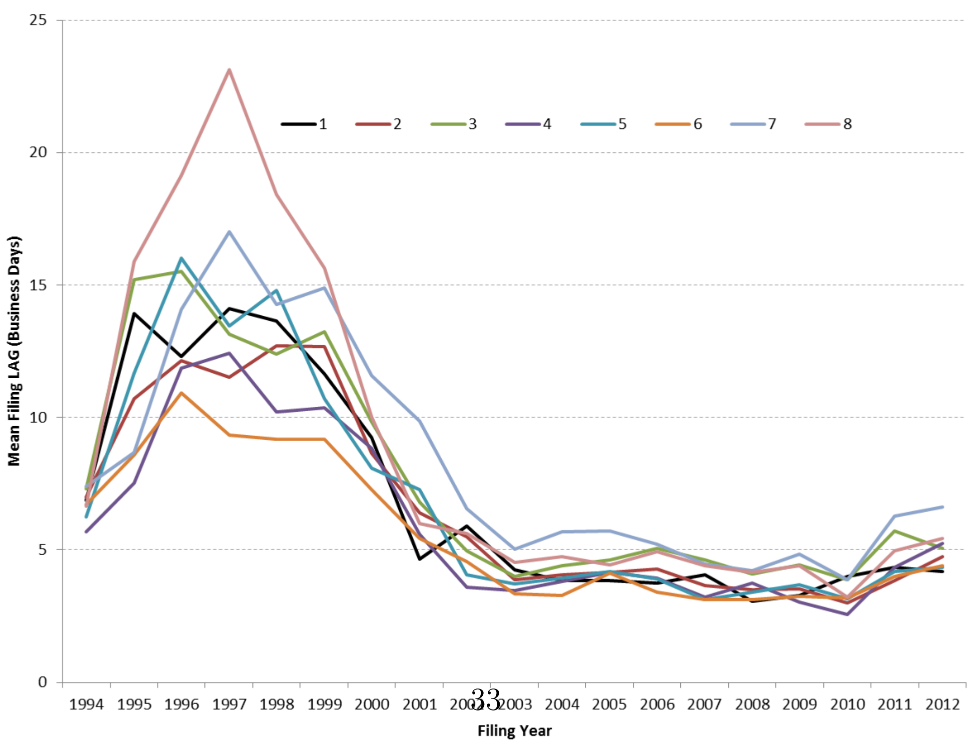


Table 1
Summary Statistics

[Insert description here]

Year	# of Firms	# of Filings	Statistics													
			Spread		Amihud		Log(MktCap)		BM		Log(Turnover)		Info_Intensity		Disclosure Lag	
			Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
1994	726	1962	2.683	3.438	0.026	0.128	13.361	1.798	0.542	0.485	6.316	0.855	6.173	6.767	8.224	13.359
1995	1480	3466	4.438	7.006	0.064	0.188	12.368	2.139	0.589	0.511	6.331	1.000	5.258	5.923	13.254	25.990
1996	3146	7291	4.544	5.495	0.072	0.192	11.949	1.997	0.523	0.485	6.637	1.041	4.477	5.022	16.313	31.839
1997	3857	10077	4.345	6.261	0.085	0.227	11.943	2.025	0.483	0.447	6.667	1.009	4.510	4.892	15.782	34.123
1998	4055	11177	3.958	5.847	0.095	0.241	12.011	2.067	0.450	0.437	6.698	0.997	4.849	4.944	15.584	38.440
1999	3732	10518	3.561	4.660	0.077	0.188	12.144	2.146	0.555	0.565	6.772	1.078	5.251	4.974	14.902	36.089
2000	3669	10994	3.745	5.584	0.081	0.189	12.355	2.155	0.578	0.619	6.931	1.117	5.765	5.151	11.568	27.485
2001	3589	12330	3.239	4.913	0.096	0.253	12.160	2.243	0.736	0.801	6.744	1.105	6.264	5.205	8.759	22.651
2002	4067	16806	2.606	3.731	0.095	0.244	12.147	2.187	0.685	0.714	6.585	1.150	6.512	4.920	6.469	17.257
2003	4642	25022	1.583	2.548	0.063	0.186	12.239	2.052	0.762	0.711	6.559	1.151	6.439	4.609	4.571	12.952
2004	4405	26179	0.908	1.403	0.033	0.096	12.696	1.943	0.513	0.461	6.816	1.143	6.974	4.650	4.919	15.013
2005	4555	38427	0.828	1.398	0.028	0.080	12.788	1.926	0.462	0.382	6.842	1.150	7.031	4.472	5.044	11.390
2006	4458	37398	0.652	1.057	0.023	0.071	12.940	1.920	0.461	0.384	6.897	1.152	7.245	4.436	5.145	16.521
2007	4379	35543	0.745	1.266	0.024	0.068	13.009	1.927	0.447	0.353	6.974	1.173	7.312	4.490	4.410	13.685
2008	4170	32673	1.842	3.259	0.062	0.150	12.669	2.027	0.561	0.473	7.100	1.208	7.429	4.502	4.138	10.015
2009	3961	29181	1.844	3.525	0.075	0.185	12.436	2.096	0.925	0.853	7.122	1.230	7.581	4.523	4.333	13.832
2010	3865	29235	1.010	2.137	0.036	0.107	12.821	2.051	0.753	0.715	7.093	1.142	7.620	4.530	4.243	20.164
2011	3752	29023	0.833	1.747	0.032	0.101	13.000	2.061	0.650	0.587	7.045	1.137	7.471	4.603	6.039	14.890
2012	3628	28409	0.953	2.062	0.032	0.105	13.027	2.094	0.752	0.699	7.002	1.161	7.329	4.705	4.603	14.280
1994-2012	10827	395711	2.107	3.952	0.057	0.168	12.511	2.086	0.603	0.600	6.836	1.143	6.490	4.898	7.914	22.533

Table 2
Disclosure Lag Heterogeneity within Firms and Textual Content Categories

[Insert description here]

Panel A: Summary of Disclosure Lag Heterogeneity Within Firms

Summary Statistics	Sample			
	(1)	(2)	(3)	(4)
Avg within-firm SD(Filing Lag)	16.126	14.665	12.976	11.391
Avg within-firm Mean(Filing Lag)	8.718	7.884	7.776	5.421
Avg No. of Filings per Firm	39.452	145.351	37.132	144.754
No. of Firms in Sample	10013	927	9234	788
Sample Restrictions				
No. of Filings	No Rest.	≥ 100	No Rest.	≥ 100
Maximum Filing Lag	No Rest.	No Rest.	≤ 150	≤ 150

Panel B: Summary of Disclosure Lag Within Content Categories
(Top Number=Average, Bottom Number=Std. Dev)

Content Type	Sample			
	(1)	(2)	(3)	(4)
Business	4.352	4.014	3.857	3.610
	15.473	13.415	9.499	8.361
Financial	6.617	5.100	5.103	3.973
	22.647	23.636	11.863	10.202
Security	3.590	2.955	3.341	2.957
	6.611	4.907	5.901	4.919
Accounting	5.554	5.384	4.528	3.545
	21.138	28.211	11.958	9.760
Governance	4.746	4.092	4.056	3.518
	17.564	14.955	10.569	9.072
Reg FD	1.859	1.286	1.379	0.862
	8.166	8.999	5.112	3.568
Other	3.725	2.705	3.119	2.304
	16.144	12.441	8.740	6.591
Sample Restrictions				
No. of Filings	No Rest.	≥ 100	No Rest.	≥ 100
Maximum Filing Lag	No Rest.	No Rest.	≤ 150	≤ 150

Table 3
Determinants of Disclosure Lag

[Insert description here]

Panel A: Disclosure Lag and Firm and Event Characteristics

	Models			
	(1)	(2)	(3)	(4)
Event Characteristics				
Tone	-21.2242***	-14.7650***	-13.5934***	-19.0771***
	(-13.99)	(-9.79)	(-7.27)	(-10.02)
Uncertainty	7.7346***	8.7683***	8.1285**	8.0198***
	(3.36)	(4.41)	(3.17)	(4.14)
Macro Economy				
Unemployment				9.9103***
				(4.66)
Recession				-4.7132***
				(-3.88)
Event Type				
Financial		0.9844***	0.9480***	
		(20.91)	(16.79)	
Security		-0.2911***	-0.3615***	
		(-6.37)	(-6.60)	
Auditing		-0.6607***	-0.5127***	
		(-7.14)	(-4.22)	
Governance		0.0143	-0.0283	
		(0.45)	(-0.72)	
ABS		-0.1231	0.3636	
		(-0.21)	(0.45)	
Reg FD		-1.8340***	-1.8649***	
		(-51.87)	(-43.56)	
Other		-1.2802***	-1.2877***	
		(-38.52)	(-31.41)	
Firm Characteristics				
Turnover			-0.0301	
			(-1.31)	
BM			0.0026	
			(0.85)	
PE			-0.0000	
			(-0.17)	
ROA			-0.1602*	
			(-2.28)	
Accruals			0.2115	
			(1.34)	
IdVol			0.8040*	
			(2.46)	
Inst_Ownership			0.0009*	
			(1.99)	
Info_Intensity			-0.0100**	
			(-3.07)	
Fixed Effects	✓	✓	✓	✓

Panel B: Detailed Content Items and Disclosure Lag

Content Details (Select Descriptions)	Models (1)	(2)
1.01	0.7037***	0.6319***
(Entry into material agreements)	(23.49)	(17.26)
1.02	0.0996	0.1058
	(1.82)	(1.45)
1.03	-1.0069***	-1.0780***
(Bankruptcy or receivership)	(3.85)	(3.35)
1.04	0.3585	0.2651
	(1.44)	(1.14)
2.01	2.4862***	2.4743***
(Asset acquisition & disposal)	(30.94)	(24.90)
2.02	-0.0830*	-0.1438**
	(-2.07)	(-2.95)
2.04	0.2614	0.1695
	(1.67)	(0.83)
2.05	1.2152***	1.0907***
(Asset disposal costs)	(6.66)	(5.24)
2.06	0.0517	-0.2001
	(0.38)	(-1.12)
3.01	-0.3058***	-0.2512***
(Delisting notices)	(-5.30)	(-3.53)
3.02	-0.0883	-0.1327
	(-1.15)	(-1.41)
3.03	-0.6788	-0.7056
	(-0.61)	(-0.45)
4.01	-0.2342*	-0.2710*
(Accountant change)	(-2.47)	(-2.11)
4.02	0.9783***	0.8940**
(Nonreliance on previous statements)	(3.97)	(2.97)
5.01	0.2743	0.2853
	(1.45)	(1.15)
5.02	0.5617***	0.5080***
(Personnel departures and appointments)	(20.61)	(14.98)
5.03	0.3995***	0.3128***
(Articles/bylaws change)	(7.81)	(5.06)
5.04	0.4029	0.1950
	(1.57)	(0.57)
5.05	0.0729	0.0977
	(1.75)	(1.47)
5.06	-2.8606***	-2.5517***
(Shell company status change)	(-6.27)	(-3.85)
5.07	-0.9021***	-1.0546***
(Submission of matters to vote)	(-12.57)	(-10.77)
5.08	-0.0981	0.0169
	(-0.23)	(0.04)
7.01	-1.0902***	-1.1309***
(Reg FD)	(-38.66)	(-32.27)
Control Variables	No	✓
Fixed Effects	✓	✓

Table 4
Disclosure Lag and Filing Period Information Asymmetry

This table reports the estimates of the regression of filing period information asymmetry, defined as a firm's average bid-ask spread and Amihud (2001) liquidity measure, over the five-day window of [filing date-1, filing date+3], against the disclosure lag measure, computed per Eq. (xyz) of the text, and various control variables. Panels A and B reports the coefficients for the bid-ask spread and Amihud (2001) measures, respectively. Control variables are defined in Appendix **??**. The coefficients estimated using a variety of fixed effects and all standard errors are robust and clustered at the firm level. The estimates use a sample of 375,438 8-Ks over 1994 to 2012. All independent variables except for disclosure lag are standardized to a mean of 0 and standard deviation of 1.

Panel A: Disclosure Lag and Filing Period Bid-Ask Spread

	Model					
	(1)	(2)	(3)	(4)	(5)	(6)
Lag	0.0419***	0.0155***	0.0138***	0.0156***	0.0153***	0.0148***
	(15.41)	(6.32)	(5.54)	(5.30)	(4.63)	(4.60)
Controls						
Turnover				-0.1424***	-0.1499***	-0.1489***
				(-6.32)	(-6.24)	(-6.50)
BM				-0.6808	0.4872	0.4101
				(-1.11)	(0.65)	(0.56)
PE				-0.0083***	-0.0449*	-0.0483*
				(-5.21)	(-2.07)	(-2.28)
ROA				-0.0599	-0.8639***	-0.8394***
				(-1.08)	(-7.37)	(-7.78)
ED_Ret					-0.0435***	-0.0402***
					(-5.32)	(-5.13)
Accruals					-0.1039*	-0.1159*
					(-2.22)	(-2.55)
IdVol					-0.0919***	-0.0983***
					(-3.81)	(-4.17)
Inst_Ownership						-0.1430***
						(-3.55)
Info_Intensity						-0.0081**
						(-2.83)
No. Obs	375438	363771	363771	285574	237599	235330
Within R^2	0.0029	0.0747	0.0770	0.0830	0.0955	0.1009
Overall R^2	0.0199	0.0790	0.0843	0.1113	0.1240	0.1405
Fixed Effects						
Firm	✓	✓	✓	✓	✓	✓
Year	No	✓	✓	✓	✓	✓
Industry	No	✓	✓	✓	✓	✓
Day-of-Week	No	No	✓	✓	✓	✓
Month-of-Year	No	No	✓	✓	✓	✓
Text Content	No	No	✓	✓	✓	✓

Panel B: Disclosure Lag and Filing Period Amihud (2002) Illiquidity Measure

	Model					
	(1)	(2)	(3)	(4)	(5)	(6)
Lag	0.0002	0.0005***	0.0005***	0.0005***	0.0004**	0.0004**
	(1.80)	(4.03)	(3.94)	(3.58)	(2.67)	(2.59)
Controls						
Turnover				0.0164***	0.0166***	0.0158***
				(11.69)	(11.28)	(11.25)
BM				-0.0094	0.0842	0.0782
				(-0.21)	(1.31)	(1.23)
PE				-0.0001	-0.0002	-0.0005
				(-1.19)	(-0.23)	(-0.50)
ROA				-0.0044	-0.0594***	-0.0574***
				(-1.11)	(-9.61)	(-9.66)
ED_Ret					-0.0007	-0.0007
					(-1.17)	(-1.10)
Accruals					-0.0078**	-0.0083**
					(-2.58)	(-2.91)
IdVol					-0.0025	-0.0035*
					(-1.44)	(-2.02)
Inst_Ownership						-0.0181***
						(-10.06)
Info_Intensity						-0.0002*
						(-2.02)
No. Obs	377769	366036	366036	287512	239116	236833
Within R^2	0.0000	0.0197	0.0249	0.0297	0.0403	0.0445
Overall R^2	0.0060	0.0114	0.0186	0.0133	0.0366	0.0602
Fixed Effects						
Firm	✓	✓	✓	✓	✓	✓
Year	No	✓	✓	✓	✓	✓
Industry	No	✓	✓	✓	✓	✓
Day-of-Week	No	No	✓	✓	✓	✓
Month-of-Year	No	No	✓	✓	✓	✓
Text Content	No	No	✓	✓	✓	✓

Table 5
Disclosure Lag and Filing Period Abnormal Returns

This table reports the estimates of the regression of filing period abnormal return, defined as a firm's buy-and-hold return minus the CRSP value-weighted index return over the four-day window of [filing date, filing date+3] against the disclosure lag measure, computed per Eq. (xyz) of the text, document tone scores computed using the LM lexicon of both positive and negative words, and various control variables. Panels A and B reports the coefficients for the samples of positive events and negative events, respectively. Control variables are defined in Appendix **??**. The coefficients estimated using a variety of fixed effects and all standard errors are robust and clustered at the firm level. The estimates use a sample of 285,052 8-Ks over 1994 to 2012. All independent variables except for disclosure lag are standardized to a mean of 0 and standard deviation of 1.

Panel A: Positive Events (Tone Score > 0)

	Models					
	(1)	(2)	(3)	(4)	(5)	(6)
Lag	-0.0414**	-0.0460***	-0.0449**	-0.0402*	-0.0467**	-0.0433*
	(-3.28)	(-3.54)	(-3.29)	(-2.52)	(-2.64)	(-2.46)
Tone Score	0.1721*	0.1646*	0.1789*	0.1816*	0.1938*	0.1874
	(2.47)	(2.34)	(2.51)	(2.12)	(1.98)	(1.91)
Controls						
BM				0.1858*	0.1620*	0.1618*
				(2.11)	(2.20)	(2.22)
Turnover				-0.5876***	-0.6192***	-0.5426***
				(-7.28)	(-6.54)	(-5.62)
PE				-0.0000	-0.0000	-0.0001
				(-0.94)	(-0.20)	(-0.27)
ROA				-0.0537	-1.3857***	-1.1920**
				(-1.36)	(-3.74)	(-3.18)
Accruals					1.0220	0.5983
					(1.53)	(0.93)
IdVol					-5.0462	-5.7604*
					(-1.75)	(-1.99)
Inst_Ownership						-0.0053***
						(-6.81)
Info_Intensity						-0.0018
						(-0.22)
Alpha	1.1872	-1.1062	-1.4151	2.8567	1.7857	0.9027
	(1.31)	(-0.78)	(-1.03)	(1.86)	(0.89)	(0.44)
No. Obs	154150	151969	151969	117069	97797	97088
Within R^2	0.0001	0.0005	0.0012	0.0033	0.0047	0.0051
Overall R^2	0.0002	0.0002	0.0004	0.0016	0.0013	0.0013
Fixed Effects						
Industry	No	✓	✓	✓	✓	✓
Year	No	✓	✓	✓	✓	✓
Day-of-Week	No	No	✓	✓	✓	✓
Month-of-Year	No	No	✓	✓	✓	✓
Text Content	No	No	✓	✓	✓	✓

Panel B: Negative Events (Tone Score < 0)

	Models					
	(1)	(2)	(3)	(4)	(5)	(6)
Lag	0.0357*	0.0351*	0.0338*	0.0389*	0.0428*	0.0446**
	(2.35)	(2.25)	(2.08)	(2.48)	(2.53)	(2.62)
Tone Score	0.1789***	0.1757***	0.1660***	0.1364***	0.0957*	0.0821*
	(5.40)	(5.28)	(4.95)	(3.88)	(2.44)	(2.10)
Controls						
BM				-0.3014***	-0.4011***	-0.3165***
				(-3.41)	(-4.50)	(-3.48)
Turnover				0.0092	0.0090	0.0104
				(0.38)	(0.47)	(0.53)
PE				-0.0000**	-0.0002	-0.0001
				(-2.80)	(-0.75)	(-0.65)
ROA				-0.0288	-0.3066	-0.3393
				(-0.56)	(-1.02)	(-1.10)
Accruals					-0.1013	-0.3723
					(-0.17)	(-0.65)
IdVol					-1.4479	-1.0412
					(-1.17)	(-0.85)
Inst_Ownership						-0.0048***
						(-6.53)
Info_Intensity						0.0003
						(0.03)
Alpha	-0.2003	-1.4816	-1.5286	2.1131	2.3116	0.6676
	(-0.60)	(-0.52)	(-0.53)	(0.88)	(0.91)	(0.23)
No. Obs	130902	129299	129299	105665	87960	86781
Within R^2	0.0003	0.0008	0.0021	0.0026	0.0030	0.0032
Overall R^2	0.0002	0.0001	0.0007	0.0011	0.0010	0.0004
Fixed Effects						
Industry	No	✓	✓	✓	✓	✓
Year	No	✓	✓	✓	✓	✓
Day-of-Week	No	No	✓	✓	✓	✓
Month-of-Year	No	No	✓	✓	✓	✓
Text Content	No	No	✓	✓	✓	✓

Panel C: Both Positive and Negative Events

	Models					
	(1)	(2)	(3)	(4)	(5)	(6)
Lag	-0.0071	-0.0091	-0.0112	-0.0049	-0.0069	-0.0038
	(-0.74)	(-0.91)	(-1.07)	(-0.44)	(-0.56)	(-0.31)
Tone Score	0.2754***	0.2784***	0.2995***	0.2772***	0.2801***	0.2691***
	(10.08)	(10.08)	(10.61)	(9.12)	(8.18)	(7.83)
Lag×Tone Score	-0.0170*	-0.0170*	-0.0186**	-0.0210**	-0.0235**	-0.0253**
	(-2.55)	(-2.55)	(-2.78)	(-2.88)	(-2.94)	(-3.12)
Controls						
BM				-0.4497***	-0.5135***	-0.4364***
				(-7.33)	(-7.96)	(-6.61)
Turnover				0.0313	0.0382	0.0373
				(0.85)	(1.55)	(1.47)
PE				-0.0000**	-0.0001	-0.0001
				(-2.60)	(-0.93)	(-0.91)
ROA				-0.0464	-0.9030***	-0.8426***
				(-1.00)	(-4.09)	(-3.74)
Accruals					0.4005	0.0818
					(0.88)	(0.19)
IdVol					-3.6234*	-3.5079*
					(-2.33)	(-2.20)
Inst_Ownership						-0.0051***
						(-8.71)
Info_Intensity						-0.0023
						(-0.34)
Alpha	0.0174	-0.5795	-0.7216	1.2675	3.0298	1.8203
	(0.90)	(-0.40)	(-0.49)	(0.71)	(1.93)	(1.20)
No. Obs	285052	281268	281268	222734	185757	183869
Within R^2	0.0006	0.0008	0.0015	0.0026	0.0034	0.0036
Overall R^2	0.0008	0.0003	0.0008	0.0018	0.0017	0.0010
Fixed Effects						
Industry	No	✓	✓	✓	✓	✓
Year	No	✓	✓	✓	✓	✓
Day-of-Week	No	No	✓	✓	✓	✓
Month-of-Year	No	No	✓	✓	✓	✓
Text Content	No	No	✓	✓	✓	✓