Uncovering Information: Can Language Models Tell Us Where to Look?*

Anna Costello Bradford Levy Valeri Nikolaev

September 2024

Abstract

Amidst the wealth of content in the digital era, identifying information that is genuinely relevant is a challenging task. This study uses large language models to uncover new, or surprising, information in textual data. Specifically, we train LLMs using data from the financial domain to model investors' prior beliefs over narrative content and, thus, identify new information that a company communicates. Our models, which are tailored to each firm, allow us to determine when a sentence or a paragraph in a corporate filing contains a surprise. The models are trained with fixed knowledge cutoffs to eliminate lookahead bias. We also generate a summary measure of the amount of information released in each document, and we show that this measure explains a large portion of market response to disclosed information and predicts firms' future profitability. In sum, our study highlights the importance of identifying new and salient information in today's complex, data-driven landscape.

JEL Classifications: C45; D8; D84; G14; M41

Keywords: Beliefs; AI; large language models; GPT; information in markets; market efficiency

*We thank Ralph Koijen, Stefan Nagel, and seminar participants at Harvard University and the University of Chicago for helpful comments and suggestions. The authors gratefully acknowledge financial support from the University of Chicago Booth School of Business and the Center for Applied AI.

1. Introduction

Over the last decade, the volume of data created by humans has grown at a rate of 54% per year. Ninety percent of this data is unstructured, predominantly composed of narrative text (IDC, 2023). The volume and complexity of data have made it increasingly difficult for humans to parse what is important versus what is not. In this paper, we combine the principles of information theory with recent advancements in generative AI technology to uncover information from the vast sea of textual data. Specifically, we develop a method to identify textual information that is "new," in the sense that it has not been communicated previously.

While the challenge of identifying information is central to many domains and, therefore, is relevant to nearly all aspects of human life, our study focuses on uncovering new information in the context of financial reporting. We do so for three primary reasons. First, both academics and practitioners have long been concerned about the rapid growth in the volume and complexity of corporate disclosures. Second, identifying new information in disclosures is important to firms' stakeholders, e.g., the society at large, and hence has the potential to be highly impactful. Third, the setting is particularly attractive because disclosures have accurate timestamps and thus provide us with a timeline of what investors knew and when. This feature is instrumental in implementing our methodology.

In order to assess whether corporate disclosures contain information, i.e., content useful to decision-makers, one needs first to develop a way of detecting and measuring information. Within capital markets, a disclosure is informative to the extent that it updates beliefs about firms' cash flows or risks associated with these cash flows (Fama, 1970). When considering numerical disclosures (e.g., numbers in financial statements), developing such a measure is relatively straightforward. For instance, much of the prior literature quantifies the news in earnings based on the surprise: the difference between realized and expected earnings (Ball and Brown, 1968). Unfortunately, adopting this approach within the domain of narrative

disclosure is not straightforward because one must form an expectation over all realizations of narrative disclosure, which quickly becomes intractable. In this paper, we present one solution to this challenge by implementing a theoretically founded measure of narrative "news" that relies on recent technological advances in language modeling. Since the narrative portion of financial disclosures tends to dwarf the numerical portion, a measure that uncovers information contained in narrative disclosure will likely be of considerable value to market participants.

Developing such a measure, however, presents a tall challenge. Take, for example, the average 10-K containing over 60,000 words. Unveiling its information requires a comprehensive analysis of the words, sentences, and their broader context both within the document and within the history of disclosures that came before it. Facing this challenge, prior work generally relies on dictionaries of words relevant to a single construct of interest – e.g., litigation risk or sentiment. Researchers thus measure disclosure volume along a given factor using the count of words in the dictionary (Loughran and McDonald, 2016). While being simple, this approach has many limitations. For example, it relies on the researcher's subjective vocabulary of relevant terms rather than learning them directly from data. Further, the meaning of a word or phrase typically depends on the context in which it is used—something that dictionaries and word counts cannot capture. Finding a word that belongs to a dictionary does not say whether the information has been communicated earlier. Given these limitations, a fundamentally different approach is needed to detect new information communicated to investors.

Our paper relies on a foundational idea from the information theory literature pioneered by Shannon (1948). When a sender transmits a message to a receiver, the information content of that message depends not only on the message itself but also on the receiver's expectations. To be informative, a disclosure should update investors' beliefs about the firm. In this regard, the key insight proposed by Shannon (1948) when quantifying new information is that observing unlikely content is more informative than observing content that the receiver

believes will occur with high probability. Shannon's analysis implies that one can calculate a measure of new information in a text as the sum of the log probabilities of each token (word), conditional on the preceding context (prior tokens). Notably, this measure is remarkably flexible in that it allows for capturing information at the word, sentence, paragraph, section, and document levels. It thus enables us to pinpoint the location in a potentially very lengthy text where significant amounts of new information reside.

Calculating the probability that a word will appear in a text requires an estimate of conditional probability over the words that could occur conditional on prior words. Notably, large language models (LLMs) have been developed specifically for this task and demonstrated the ability to do so very accurately (Radford et al., 2019). As such, we begin by training an LLM from scratch on a cross-section of firms' narrative disclosures.[1]. We then fine-tune the general model on each firm's individual time series of disclosures to yield a firm-specific LLM for each firm in the sample. Finally, we then apply this firm-specific LLM to firms' disclosures out-of-sample to detect new informational content in these disclosures.

Our sample draws on SEC disclosures on EDGAR, including Forms 10-K, 8-K, and 10-Q, filed by companies during the period 1996-2022. Our training sample includes all documents disclosed before 2007, whereas our inference (out-of-sample) group includes disclosures from 2007 and thereafter. Our data cover 2,500 firms, nearly 570,000 filings, and almost 10 billion words. For each form-type, we compute an overall measure of information and also more granular information based on the location within the filing. We find that for 10-K and 8-K filings, a significant amount of new textual information tends to be embedded in exhibits rather than in the main filing. Further, across all filing types, information tends to be high at the beginning of the filing and spike towards the end. The results point to where investors should focus their attention when seeking new information within lengthy text documents.

Research on capital markets in accounting and finance typically classifies news as infor-

[1]We eliminate numerical tables from the training data in order to focus the analysis on narrative content.

mative if it provides decision-useful information to investors (e.g., Ball and Brown, 1968; Fama et al., 1969; Bernard and Thomas, 1989; Tetlock et al., 2008). It is important to note that we do not directly identify *decision-relevant* news, but rather measures the extent to which content is novel relative to the firm's prior disclosures. Ultimately, whether a piece of new content is decision-relevant depends on the decision at hand (Demski, 1973; Gjesdal, 1981). For example, what is relevant to an ESG activist may not be relevant to a value investor, and thus, uncovering "relevant news" would necessarily be limiting (as it would require narrowing down the objective). Given the broad objective of our study, the "new" information that we uncover may not be relevant for some uses but relevant for others.[2]

Having computed a measure of information in each disclosure by a given firm based on the corresponding firm-specific LLM, we move to our empirical analyses. In our first set of tests, we use a short-window daily event study to assess the extent to which our measure captures information within the framework of Fama (1970). If prices impound all publicly available information, and our measure captures this information, then we expect the measure to explain short window price reactions. Consistent with our measure capturing decision-relevant information, we find that our measure explains the majority of the market reaction. For example, the event-day increase in trading volume associated with a current report (Form 8-K) is 58.9% relative to the sample mean. After controlling for our measure, we find that current reports in the lowest decile of information are associated with an 18.6% increase in trading volume, while those in the highest decile of information are associated with a 99.4% increase. These results are consistent with the notion that narrative content is responsible for the majority of the information in the financial disclosures we consider.

[2]There are legal and economic reasons to expect that our measure will capture the news that is of broad relevance to capital markets. EDGAR system is the primary mechanism through which entities satisfy their disclosure requirements under the Securities Act of 1933. The act had two primary objectives: (i) to require that investors receive financial and other significant information concerning securities being offered for public sale; and (ii) to prohibit deceit, misrepresentations, and other fraud in the sale of securities (SEC, 2019). Thus, by conditioning on a set of disclosures filed on EDGAR, our measure is likely to identify decision-relevant information.

In our second set of tests, we showcase our measure of novel content by analyzing the sentiment of a given disclosure. A large literature in accounting and finance shows that markets react to disclosure sentiment (see Loughran and McDonald (2016) for a review). Our insight is that the sentiment should only explain market reactions if it relates to new information. For example, it is hardly a surprise that a company uses a positive tone, in general, when presenting operating performance. Similarly, if a disclosure contains negative content that recaps what was already disclosed in a prior period, there should not be a market reaction. We thus predict an interaction between the sentiment of a filing and information in the filing. One can think of this interaction as an information-weighted measure of sentiment, as we discuss in more detail further, or as a way to uncover the "signed" textual information.

Consistent with prior literature (e.g., Loughran and McDonald (2011)), we find that the market reaction to negative sentiment alone is modest. It roughly constitutes 13.8% of the average event-day absolute return. However, this reaction is amplified by an order of magnitude when disclosure has high textual information content. Specifically, our measure of information-weighted sentiment is approximately 231% of the average event-day absolute return. These results are consistent with our measure capturing value-relevant information. They also highlight the importance of (precisely) measuring both information and sentiment when processing a new disclosure.

In our third set of tests, we examine the extent to which our measure of information in narrative disclosures is predictive of future firm fundamental performance. If our measure is capturing value-relevant information in a firm's disclosure, it should predict the firm's future performance. Following prior literature, we examine performance by focusing on the return on assets (ROA). Using predictive regressions for horizons ranging from one to sixteen quarters ahead, we find both statistically and economically significant relation between the signed information and future performance. For example, examining 10-K/Q filings, moving from the lowest decile of signed information to the highest decile is associated with a 10.7% increase in the following quarter's ROA. We find that this relationship persists four years

into the future. The magnitudes are somewhat attenuated (4.25%) but remain economically significant.

Collectively, the results of our first three sets of tests are consistent with our measure capturing decision-relevant information in firms' narrative disclosures. Notably, these disclosures are available to capital market participants free-of-charge and in a timely manner via the SEC EDGAR system. As such, if prices fully reflect all publicly available information, then we expect markets to quickly impound into prices the information contained in these disclosures. On the other hand, if frictions inhibit or impede the processing of these disclosures, then we expect a moderated reaction followed by a drift in the direction of the information-weighted sentiment we measure. In our final set of tests, we explore this question.

Using predictive regressions of quarterly abnormal returns for horizons from one quarter ahead to 16 quarters ahead, we find evidence of a *statistically* significant but relatively small *economically* relation between our measure and *abnormal* returns. For example, when examining 10-K/Q filings, we find that moving from the lowest decile of information-weighted sentiment to the highest is associated with a 0.28% increase in abnormal return (on an absolute basis) over the following quarter. We find that this relationship persists four years into the future, where it is associated with an attenuated 0.148% increase in abnormal returns. These magnitudes, however, are likely economically *insignificant* after accounting for trading costs, price impact, interest paid to lenders of borrowed shares, etc.

The primary contribution of our study is the development of a method for identifying information within unstructured textual data, e.g., narrative disclosures. As such, our study and findings are of interest to a broad audience beyond accounting and finance. We show that LLMs can be used to form priors over narrative content, which can then be used to identify information in new content. Our study is the first to do this in the context of financial reporting, and it significantly adds to prior work on textual analysis in this setting.

Specifically, our study recognizes that text is multidimensional, is often unstructured, and is difficult to directly compare across documents. Our parsimonious measure of information accounts for all of these complexities.

Our work is related to recent papers which have shown the ability of LLMs to predict returns. These studies include asking an LLM whether a stock will rise or fall (Lopez-Lira and Tang, 2023), generating summaries of lengthy content (Kim et al., 2023), and extracting dense representations of content (Chen et al., 2022). A common theme across these studies is that an LLM is used to irreversibly transform a source document into another representation. As a result, it is not immediately clear why these representations predict returns. In contrast to these studies, our work preserves the source document and uses an LLM to overlay a measure of information on top of it (see, e.g., Figure 2) thereby "uncovering the information" and eliminating the possibility that part of the output is a hallucination–one of the primary barriers to using LLMs in fields such as finance and law where the veracity of information is crucial (Ji et al., 2023). In contemporaneous work, Sarkar and Vafa (2024) highlights that commercial LLMs appear to contain lookahead bias. In our paper, we train LLMs from scratch with fixed knowledge cutoffs such that all of our tests are out of sample relative to the training data.

Our paper is also related to the large literature on costly information processing. The closest paper in this area is Cohen et al. (2020), who study annual and quarterly reports. Specifically, they measure the similarity of text across these documents as a measure of informativeness. Similarity, however, does not get at "news" as much of the information that is summarized in a 10-K filing was disclosed throughout the year – and the level of informativeness should be a function of *all* information that has been released, i.e., also information in current reports and other filings designed to be more timely than quarterly and annual reports. After considering more than just annual and quarterly reports, we find that information is quickly impounded into stock prices upon disclosure, whereas Cohen et al. (2020) suggests some markets fail to react to newly realized information. An important

advancement of our study is that financial information is released frequently, and thus any measure of novel information should account for *all* information that came before it.

The remainder of our paper proceeds as follows. Section 2 lays out the theoretical foundations of our measure. Section 3 discusses our sample, data, and model training. Section 4 presents empirical tests of our measure. Section 5 concludes.

2. Theoretical Motivation

Information content of corporate disclosures and the associated market reactions are among the most fundamental questions within the accounting and finance literature. This literature goes back to seminal studies by Ball and Brown (1968), Beaver (1968), and Fama et al. (1969). The primary result from these studies is that security prices are quick to incorporate corporate news upon arrival and hence reflect publicly available information in a timely manner. For the most part, however, the studies examining information content over the past five decades focus on quantitative information, e.g., earnings or stock returns, and use the notion of *surprise*, i.e., the difference between the realization of a random variable and its expectation, to quantify the amount of new information. More formally, the surprise is defined as:

$$s_{t+1} = x_{t+1} - \mathbb{E}[\tilde{x}_{t+1}|\Omega_t]. \tag{1}$$

For example, in Ball and Brown (1968) $\tilde{x}_t$ is a firm's net income. More recent work has also used this approach in attempt to quantify narrative content, e.g., in Tetlock et al. (2008) $\tilde{x}_t$ is the fraction of negative words in firm-specific news stories, and in Loughran and McDonald (2011) it is the fraction of words in a firm's 10-K belonging to a given dictionary, e.g., litigious words.

In contrast to numeric data, measuring the amount of new information in textual communications presents a formidable challenge. First, narrative information is highly multidi-

mensional and becomes the notion of a surprise, which is challenging to implement directly. Second, calculating expected values for textual data is not a well-defined problem.

One way the literature tacked these challenges is by assessing the similarity between documents (annual reports) from period to period, e.g., based on bag of words representation. This approach has been adopted by several recent studies (e.g., Brown and Tucker, 2011; Cohen et al., 2020). Similarly, Loughran and McDonald (2011) use the prior period's 10-K as a proxy for the expected sentiment (a unidimensional construct) in the current period's 10-K.

One problem with this approach is that it does not identify the new information per se but instead suggests that it appears present in the filing (relative to the prior period). Thus, it is not particularly helpful if one is interested in identifying and processing the new information in its own right. Furthermore, the lack of similarity does not imply the presence of new information. Specifically, dropping irrelevant text from an older 10-K affects similarity in the same way as adding new information to the current 10-K. More importantly, the approach requires two consecutive 10-Ks and cannot be easily adapted to compare annual and quarterly reports due to their different nature and level of detail. It is plausible that disclosures, such as 8-Ks (press releases) and 10-Qs that precede the current 10-K, among others, had already communicated the information, which is repeated in the annual report. In this case, information is already reflected in prices and investors' expectations and hence does not meet the definition of new information.

This discussion suggests that measuring the degree of new information in corporate filings requires a fundamentally different approach. In this paper, we aim to address this challenge by combining the primitives from the information theory and recent developments in natural language modeling.

2.1. Measuring Information

In a foundational paper within the field of information theory, a branch of applied mathematics, Shannon (1948) examines a model of communication in which a sender transmits a message to a receiver. In this setting, the information content of the message is the component that was previously unknown to the receiver. For example, if the entire message was known to the receiver a priori, i.e., the message was expected to be received with probability one, then the information content is zero. Thus, the information content depends not only on the message but also on the receiver's expectations about the message–just as in Eq. 1 above. Shannon's key insight was that witnessing a realization of an unlikely event is more informative than witnessing an event that is likely, which leads to the well-known definition of information:

$$I(x) = -\log p(x) \tag{2}$$

where $p(\cdot)$ is the probability density function of x.

Shannon information, sometimes referred to as self-information, is closely linked to the notion of a surprise used in the prior literature. To illustrate this, consider a continuous random variable $\tilde{x}$ that follows a Gaussian distribution. Substituting the density of $\tilde{x}$ into Eq. 2 yields the following relation for the information contained in a given realization x:

$$I(x) \propto (x - \mathbb{E}(\tilde{x}))^2. \tag{3}$$

Hence, the main difference between Eqs. 1 and 2 is that the former signs the "surprise" in observing a realization x, whereas the latter does not. In other words, Eq. 2 does not make a statement about the sentiment of the surprise, only its magnitude, but otherwise retains its intuitive and desirable properties.[3]

[3]While this example assumes that $\tilde{x} \sim \mathcal{N}(\mu, \sigma^2)$, other tractable and intuitive forms of the information content can be derived for members of the exponential family of distributions—a broad class of distributions

Our measure of information in a given textual disclosure follows directly from Shannon's definition of information. Any textual disclosure $\tilde{\mathcal{D}}$ can be represented as a sequence of random variables (tokens), $\tilde{\mathcal{D}} = \{\tilde{\tau}_1, ..., \tilde{\tau}_n\}$, which follow a joint probability distribution $p_{\Omega_t}(\cdot)$. These could be words in a sentence or could also be sentences in a document. Then, the information in a realized disclosure $\mathcal{D}$ is given by:

$$\begin{aligned} I_t(\mathcal{D}) &= -\log p_{\Omega_t}(\tau_1, ..., \tau_n) \\ &= -\log \prod_{i=2}^{n} p_{\Omega_t}(\tau_i|\tau_{i-1}, ..., \tau_1) p_{\Omega_t}(\tau_1), \\ &= -\sum_{i=2}^{n} \log p_{\Omega_t}(\tau_i|\tau_{i-1}, ..., \tau_1) - \log p_{\Omega_t}(\tau_1), \end{aligned} \tag{4}$$

where Ω_t is a set of publicly available information and where we used the chain rule of conditional probabilities. This formulation is particularly useful as it implies that measuring the amount of new information in a text boils down to summing the log probabilities of each word (token) conditional on the preceding words. These conditional probabilities are precisely what large language models, such as GPT, are designed to model.[4]

Along these lines, we leverage the GPT architecture to empirically estimate $p_{\Omega_t}(\cdot)$ from a collection of disclosures available at time t. For historical context, it is important to note that the fields of information theory and computational linguistics have a long history of constructing such estimates, going back to Shannon (1948), who presents a series of approximations to the English language. While the literature examined various methods for modeling $p_{\Omega_t}(\cdot)$, it was not until the recent revolutionary advances in language modeling

which includes the Gaussian.

[4]The formulation above is causal in the sense that the probability of a given token depends only the context on the left side of the token. Alternatively, one can write the joint distribution based on the tokens to the right of a given token, i.e., $p(\mathcal{D}) = p(\tau_n)p(\tau_{n-1}|\tau_n) \times ... \times p(\tau_1|\tau_2, ..., \tau_n)$. Large language models (LLMs), such as BERT (Devlin et al., 2019), take advantage of this idea by using the context on both sides of a given token, which is particularly helpful if the goal is to encode information (e.g., encode a sentence in a foreign language before translating it). We do not follow this approach because we are interested in identifying new information conditional on prior information.

that researchers were able to generate believable and readable text (Radford et al., 2019). These developments are responsible for the recent success of "Generative AI" in a wide range of domains. For these reasons, we focus our attention on modeling $p_{\Omega_t}(\cdot)$ using the LLM architecture detailed in Section 3.4.

3. Data, Model Training, and Inference

3.1. Sample Construction

One challenge we face in our study is the computational cost of training a large language model. Specifically, our study requires (i) training an LLM from scratch on a cross-section of firms' narrative disclosures, (ii) fine-tuning the LLM from step (i) on each individual firm's time-series of disclosures to yield a firm-specific model for each firm in the sample, and (iii) applying the firm-specific model from step (ii) out-of-sample to measure the information in new narrative disclosures. The computation associated with executing this sequence of tasks is largely a function of two choices: the LLM size and the number of firms in the sample.

In selecting an appropriate LLM size, i.e., the number of parameters, one must consider the volume of training data, compute budget, and the capacity of the LLM to model the distributional properties of firms' disclosures (Hoffmann et al., 2023). We focus on the latter since the volume of training data and compute budget are largely fixed. Existing evidence suggests that model size has decreasing returns to scale, i.e., the benefits from using larger and larger models become relatively small. Radford et al. (2019) show that doubling the model size from 345 million parameters to 762 million reduces perplexity–a measure of how well a probability model fits a sample–by 30.3%, while further doubling the model size from 762 million parameters to 1.54 billion reduces perplexity by just 14.4%. Although larger models are capable of generating text that humans would attribute to a human author (e.g., Brown et al., 2020)), we are primarily concerned with modeling the distributional properties of firms' disclosures, of which perplexity is a direct measure. For these reasons, we train

roughly 750 million parameter models, which are similar to GPT-2 Large model (Radford et al., 2019).

The other factor influencing our computation costs is the number of firms in our sample. The larger the number of firms, the greater the number of firm-specific LLMs that need to be trained and the volume of out-of-sample disclosures that need to be analyzed. We moderate our computation costs by drawing a random sample of firms stratified by decile of market capitalization. Specifically, we consider all firms that filed either a 10-K or 10-Q on EDGAR during the year 2007. We require that firms have a valid link to Compustat and CRSP. Next, we collect firms' market value of equity as of the end of 2007 from Compustat and assign each firm to a market value decile. Finally, we randomly sample 250 firms from each decile of market value for a total sample size of 2,500 firms. Table 2, Panel A presents the effects of these filters on the firms eligible to be included in our sample. Given this sample of firms, we then collect firm fundamentals from Compustat, equity market data from CRSP, and intraday data from NYSE TAQ. We also collect firms' disclosures from EDGAR, as detailed below. Throughout our analyses, we winsorize all continuous variables by year at the 1% level.

3.2. Extracting Narrative Disclosure

Developing our measure requires a set of narrative disclosures in plain text format, i.e., a "corpus." We source these disclosures from SEC EDGAR database. Notably, EDGAR system accepts and disseminates a wide array of filings, ranging from lengthy annual reports (Form 10-K) to brief and standardized disclosures such as changes in beneficial ownership (Form 4). While one could consume all such disclosures on EDGAR–and our measure can be computed for all of them–the focus of our study is on measuring the information in narrative disclosures, i.e., those that primarily contain lengthy written-word content. As can be seen from Table 1, which presents the volume of narrative disclosure by form type (for the 20 longest form types based on word count), the three most voluminous form types are

the Current Report (Form 8-K), Quarterly Report (Form 10-Q), and Annual Report (Form 10-K), respectively.

We focus our analysis on Forms 8-K, 10-Q, and 10-K for three primary reasons. First, these three forms alone are responsible for a considerable amount of all the narrative disclosure on EDGAR, i.e., 40% of all narrative disclosure on EDGAR while the next 40% is spread across 17 form types and the remaining 20% is spread across 405 form types. Second, Forms 10-K/Q (annual and quarterly reports) serve to provide stakeholders with a standardized set of content, i.e., Items, at regular intervals, and the prior literature has found substantial evidence that this content contains information (see, e.g., Loughran and McDonald (2016)). Third, while the 10-Ks/10-Qs are prepared periodically, the purpose of Form 8-K (the current report) is to inform stakeholders of material events on an ongoing basis (without waiting for a 10-K/Q to be filed). The SEC requires that firms disclose the event within four business days of its occurrence. Table 2, Panel B presents the distribution of Forms 8-K, 10-Q, and 10-K filed by firms in our sample.

Focusing on Forms 8-K, 10-Q, and 10-K, Figure 1 presents the process of assembling the corpus used in this study graphically. As a first step, we extract the narrative content from all filings of our selected form types. Each form type consists of the main filing and frequently includes attachments. The main filing generally follows a standardized format based upon templates provided by the SEC.[5] These main filings take on one of two file types: plain text or HTML. In contrast, attachments can take on many different formats depending on the nature of the attachment, e.g., an asset purchase agreement, forecasts of future performance, etc. Similarly, attachments take on a variety of file types beyond plain text and HTML, e.g., PDFs, Excel files, etc. While our measure could be applied to narrative content extracted from any file type, we focus on plain text and HTML due to their ubiquitous nature.[6]

[5]For example, Form 10-K instructions and template: https://www.sec.gov/files/form10-k.pdf

[6]In exploring filings with a large number of attachments, which are neither plain text nor HTML, we noticed that firms embed content from other file types into HTML-based attachments in such a way that

We follow the process detailed in Appendix A to extract the narrative content from each form type. This approach leads to a corpus of plain text documents that are structurally similar regardless of the source file type, i.e., plain text or HTML, or the time period when the form was filed, e.g., prior to widespread adoption of HTML-based filings. We use this corpus to train and test all of our models. The training sample includes disclosures from 1996-2006 that amount to 22.8 GB of narrative content containing 3.67 billion words. In turn, the test sample includes disclosures from 2007-2022 with the total 40.1 GB of narrative content containing 6.20 billion words.

3.3. Tokenization

As we discussed in Section 2, large language models require converting text into tokens, which are subsequently encoded as vectors. One could define tokens based upon entire words, e.g., "the" and "context" would each be a unique token. This would be akin to memorizing a dictionary and simply looking up the meaning of a given token in the dictionary. One challenge with this approach is that it leads to a very large vocabulary, perhaps unnecessarily large since many words are constructed by combining sub-pieces together where each sub-piece may share meaning across the vocabulary, e.g., the prefix "un" in words such as unpresumptuous and unselfish.[7]

Following prior literature, we use Byte-Pair Encoding (BPE) (Sennrich et al., 2016), which leverages commonality across words to build a vocabulary of a more manageable size.

our process extracts the content. For example, firms assemble HTML versions of slide decks, i.e., an HTML file that presents images of each slide in sequence. Notably, accessibility devices such as screen readers may not be able to read narrative content embedded in images of slides. Presumably, to address this accessibility hurdle, firms embed the text from the images into the HTML file such that it is not visible to the eye but is visible to machines such as screen readers. Given the manner in which we extract content from filings, we also collect such embedded content.

[7] At the opposite end of the spectrum, one can define tokens based on the 26 characters in the English alphabet, plus the 10 Arabic numerals, some punctuation characters, say . and !, and the white-space character " ". This would lead to a set of 39 unique tokens. This set of tokens would be referred to as the "vocabulary" of our model. With just these 39 unique tokens much of English text could be modeled, doing so however would require learning context specific representations for every single token, i.e., the meaning of "t" when used in the sequence " the " versus " context."

This approach involves finding a smaller number of unique sub-words in the corpus, along with the individual characters (or more generally "symbols,") that can be used to form the words in the target vocabulary. The algorithm begins by merging the pairs of adjacent symbols which are most commonly found next to one another and repeats this process until the vocabulary reaches a target size.

We train a tokenizer from scratch using *only* the "in-sample" disclosures within our corpus, i.e., those filings from 1996 through 2006. The inputs to the training process are largely the same as those in Radford et al. (2019) with the exception that we explicitly include tokens for the Arabic numerals $0, 1, ..., 9$. This results in a tokenizer with a vocabulary size of 50,270: 50,257 tokens learned via the training process, 10 tokens Arabic numerals, and three special tokens for beginning of sentence, end of sentence, and unknown (word that is not in the vocabulary).

3.4. Model Training

In this section, we detail our approach to empirically estimating investors' prior beliefs about a disclosure conditional on the information available to investors at time t, i.e., $p_{\Omega_t}(\mathcal{D})$. Recall that the definition of our measure of information (Eq. 4) is:

$$I_t(\mathcal{D}) = -\log \prod_{i=1}^{n} p_{\Omega_t}(\tau_i | \tau_{i-1}, ..., \tau_1),$$

Let a collection of disclosures be indexed by $J = 1, 2, ..., m$. Then, the information set at time t can be approximated as $\widehat{\Omega}_t = \{\mathcal{D}_j | j \in J\}$, where each disclosure $\mathcal{D}_j$ is a collection of tokens $\tau_{j,1}, ..., \tau_{j,n_j}$. Let $\widehat{p}_{\Omega_t}(\cdot; \theta)$ be a neural network parameterized by θ. We obtain our empirical estimate of investors' prior beliefs by choosing θ such that

$$\theta := \arg\min_{\theta} \sum_{j=1}^{m} \sum_{i=1}^{n_j} -\log \widehat{p}_{\Omega_t}(\tau_{j,i} | \tau_{j,i-1}, ..., \tau_{j,i-k}; \theta) \tag{5}$$

where $\widehat{p}_{\Omega_t}(\cdot;\theta)$ is a multi-layer transformer neural network architecture similar to "GPT-2" (Radford et al., 2019) with a model size of 774 million parameters, and k is the size of the context window which we set to 1,024. This objective function is the standard causal language modeling objective, i.e., the context window consists only of tokens on the left side of τ_i, and is equivalent to maximizing the log-likelihood. We minimize the objective using stochastic gradient descent with a batch size of 64.[8] Each batch consists of randomly sampled sequences of 512 contiguous tokens. For pre-training, described in Section 3.4.1, we use a variable learning rate schedule which ramps from zero to a maximum learning rate of 2e-4 over the first 1000 batches and then follows a cosine decay to zero over 20 epochs. For fine-tuning, described in Section 3.4.2, we use a constant learning rate of 5e-5.

While we use the same *architecture* as Radford et al. (2019), **we train a new model from scratch** such that the model has *only* been conditioned on the historical information contained in Ω_t. This differs sharply from other studies which fine-tune models pre-trained on the entire corpus, e.g., FinBERT (Araci, 2019). Such distinction is critical when detecting new information as one must ensure that the model does not condition its predictions on the information revealed in the future.

3.4.1. Cross-sectional Model Pre-Training

We begin by pre-training a single "cross-sectional" LLM on the narrative content extracted from the disclosures of all firms in our sample filed on EDGAR from 1996 through the end of 2006. The training data from this time period totals 22.8 GB of narrative content and more than 3.6 billion words. Notably, the volume of training data is of the same order of magnitude as that in Radford et al. (2019) upon which our model architecture is based. This would suggest that it should be feasible to pre-train similarly sized models from scratch

[8]Note, however, that the memory constraints of the hardware used for training require an actual batch size of eight. We thus accumulate gradients for eight steps before performing a backward pass, yielding an effective batch size of 64.

using our corpus.

While Radford et al. (2019) do not disclose for how long GPT-2 was pre-trained, similar studies suggest pre-training on the order of 10-100 epochs.[9] For example, BERT was pre-trained on a corpus of 3.3 billion words for 40 epochs (Devlin et al., 2019), and the original GPT model was pre-trained for 100 epochs (Radford et al., 2018). We randomly initialize the weights θ and pre-train our cross-sectional model for a total of 10 epochs. At this point, the objective evaluated on an out-of-sample set of data is still decreasing with further training but only marginally so. For this reason, we stop pre-training of the cross-sectional model at this stage.

The primary benefit of using a decoder architecture such as GPT is the ability to generate text by sampling from the prior modeled by the LLM. In Appendix B, we exploit this feature of our LLM to explore the extent to which the training procedure has resulted in an LLM that reasonably models the distributional properties of narrative content. Specifically, we prompt the model with text that is commonly found at the start of the business description in Form 10-K. Table B1 presents this prompt.[10] Table B2 presents four examples of text generated by our cross-sectional LLM. We note that the model is capable of generating reasonably coherent text spanning various industries and interleaving numbers and dates as a reader might expect. We view this as additional evidence of a successful training process.

3.4.2. Firm-Specific Fine-Tuning

Our goal is to construct a firm-specific estimate of investors' prior beliefs for use in computing our measure of information. We obtain such an estimate by fine-tuning the pre-trained cross-sectional model on the time series of a given firm's narrative disclosures. Let a set of firms be indexed by $F = 1, 2, ..., p$ and $J_{f,t}$ be the set of indices associated with

[9] These models are trained iteratively using small samples of data, i.e., batches, sampled from the larger dataset. An epoch is defined as one loop through the totality of the training data.

[10] Formatting such as white space and indentation are preserved in our training data. Thus, when prompting the model we intentionally use similar formatting.

disclosures produced by firm f before time t. Then the firm-specific information set at time t for firm f can be approximated by $\widehat{\Omega}_{f,t} = \{\mathcal{D}_j | j \in J_{f,t}\}$. We then obtain p firm-specific estimates of investors' prior beliefs by choosing $\theta_{f,t}$ to minimize Eq. 5 computed over just the disclosures indexed by $J_{f,t}$.

In contrast to pre-training where the neural network parameters were initialized randomly, we begin this fine-tuning procedure by initializing the neural network weights to the optimized weights θ from the pre-training procedure. We adopt the early-stopping method to determine when to stop the fine-tuning. We evaluate the objective on a validation sample at the end of each epoch and stop training when this quantity ceases to decrease for three epochs. Carrying out this procedure culminates in an estimated firm-specific prior belief $\widehat{p}_{\Omega_{f,t}}(\cdot; \theta_f)$.

As we did with our cross-sectional LLM, we prompt these firm-specific LLMs and generate text by sampling from them. Table B3 presents examples for Apple, Cummins, Pfizer, and Weyco Group (a microcap). The diversity of the generated text is striking, with the firm-specific LLMs referencing relevant products, previous disclosures, terms of agreements with suppliers, and financial performance. We manually inspect these samples relative to the firms' actual disclosures. The text generated for Cummins refers to a particular footnote of a previous disclosure. Examining Cummins's time-series of disclosure we find that this is an accurate reference. For Pfizer, we note that the text references two drugs called Alond and Exubera. The former does not appear to exist howver the latter was indeed a Pfizer drug which was withdrawn from the market due to low sales. Collectively, we view the richness of these prompts as further evidence that our training and fine-tuning processes result in LLMs that are reasonable models of the distributional properties of narrative disclosure.

4. Empirical Tests

4.1. Measuring Information

Given a disclosure $\mathcal{D}$ produced by a firm f, we apply the firm-specific prior to quantify the information of the i-th token:

$$I(\tau_i) = -\log \widehat{p}_{\Omega_{f,t}}(\tau_i|\tau_{i-1}, ..., \tau_{i-k}; \theta_f). \tag{6}$$

This approach allows us to identify the location of comparatively more informative tokens with a high degree of precision, as we illustrate next. To see this, consider the announcement on January 17, 2007, disclosed in Apple's 8-K filing, where the company first mentions the introduction of its iPhone. As illustrated in Figure 2, the sentences shaded in green highlight the degree of "news" relative to the estimated Apple-specific prior. While the disclosure contains numerous pieces of information, such as record revenue, the most informative sentence–according to the estimated prior–is the sentence in which Apple mentions launching the iPhone and Apple TV.

As a next step, we provide descriptive evidence on the location of information in 10-K, 10-Q, and 8-K filings. For each filing, we separately consider the body of the filing, i.e., the portion of the filing that generally follows a standardized format prescribed by the SEC, and exhibits. Specifically, we compute the token-by-token information using the respective prior and normalize the position of each token in the filing such that the first token corresponds to a position marked by 0%, and the last token corresponds to 100%. Then we compute the geometric mean within each percentage increment, i.e., $[0\%, 1\%), [1\%, 2\%), ..., [99\%, 100\%]$. If a given filing has multiple exhibits, then we combine the exhibits by computing the arithmetic mean for each percentage increment across exhibits weighted by the length of each exhibit. Finally, we average across all filings.

Figure 3. Panel A presents the results. For 10-K and 8-K filings, we find that exhibits

tend to be more informative than the main filing, while the opposite is true for 10-Q filings. Across all filing types, information tends to spike toward the end of the filing. We explore this phenomenon by examining a random sample of filings and find that, for 10-K/Q filings, the spike at the end appears to be driven by footnotes. For 8-Ks, this spike is most pronounced in the exhibits; we find that this is often due to firms providing boilerplate information up front, e.g., caveats to forward-looking statements, followed by more detailed information, e.g., discussion of new business strategies. Figure 3. Panel B presents the same statistics partitioned by whether the average information in a filing is above or below the median. The results for 8-K filings are quite similar to those in Panel A. However, for 10-Ks, it is noteworthy that firms tend to discuss new information early on.

Table 3 presents the distribution of our information measure separately for the set of main filings and exhibits in our sample. For a given main filing, we compute the filing's information as the geometric mean of the token-by-token measurements. For a set of exhibits attached to a given filing, we compute the exhibit's information as the geometric mean of the token-by-token measurements and then compute the arithmetic mean across all exhibits weighted by the length of each exhibit. We find that the most informative narrative content tends to be in the exhibits attached to current reports.

4.2. Market Reaction

In our first set of tests, we use a short-window daily event study to examine whether the quantity of textual information that we measure can explain the market reaction to disclosures. Specifically, if our measure indeed captures new information, then we expect to see a more pronounced market reaction around the disclosure date when our measure suggests a higher information content.

Figure 4 plots the average daily values of *AbsoluteReturn* and *Volume* in the [-10, 10] day window around the filing date separately for filings that are above and below median *Info* – the amount of new information in the textual disclosures over the entire document. Panel A

presents results for annual and quarterly reports, Panel B is based on current reports (8-Ks), and Panel C is based on the narrative content of the Management's Discussion and Analysis. Across all panels, there is a distinct increase in the absolute return and volume on the filing date. More importantly, the increase appears markedly greater for the above median *Info* sub-sample.

We also expect that the market reaction on the disclosure day will increase with our measure. To investigate this, Figure 5 plots the average value of *AbsoluteReturn* and *Volume* for $t = 0$ as a function of *Info*. We find that both measures of market reaction tend to be increasing functions of *Info* with the exception of the most extreme deciles. We note that this non-monotonicity largely disappears for MD&A disclosures, which are less likely to discuss "news" that is economically less relevant (e.g., a change in language related to some of the footnotes).

4.2.1. Baseline Daily Event Study of Market Reaction

To test whether there is a statistically significant relation between *Info* and the market reaction, we estimate the following regression pooling across all firm-dates in the [-10, 10] day window around the filing date:

$$y_{f,t} = \beta_1 Event[0] + \beta_2 Event[0] \times Info_{f,j} + \boldsymbol{\gamma}\mathbf{X}_{f,t} + \delta_t + \lambda_{f,t} + \epsilon_{f,t} \tag{7}$$

where $y_{f,t}$ is either *AbsoluteReturn* or *Volume*, for firm f on date t. *Event*[0] is an indicator variable equal to one on the first date the market could react to the filing and zero otherwise.[11] $Info_{f,t}$ is our measure of the information contained in the j-th disclosure filed by firm f. We also present a benchmark model where we proxy for information content with document *Length* (Bonsall et al., 2017), which is the count of characters in the narrative content in the

[11] This is the filing date for filings accepted either prior to markets opening or intraday and the date of the following trading session whereas for filings accepted after markets close.

filing deciled and scaled to range from zero to one. $\mathbf{X}_{f,t}$ is a vector of control variables for firm f on date t (Defined in Table 4 and enumerated in Table 5). δ_t is a date fixed effect. $\lambda_{f,t}$ is a firm fixed effect for our 10-K/Q sample and a firm-year-quarter fixed effect for our 8-K sample. β_1 and β_2 are our coefficients of interest. β_1 represents the average market reaction to filings in the lowest decile of *Info* while β_2 represents the incremental market reaction associated with moving from the lowest to highest decile of *Info*. Throughout our analyses, we estimate regressions using OLS and report standard errors clustered by firm and independently by date, thus allowing for arbitrary correlation across time within a given firm and across firms within a given date.

Table 4 presents descriptive statistics for the variables used in our daily event study partitioned by above and below median *Info*. Table 4 suggests that firms with high *Info* disclosures are not systematically different, i.e., of the 17 variables considered none are significantly different between groups at the one percent level. Notably, firms in the high *Info* group tend to be slightly more levered (*Leverage* of 0.585 versus 0.564 and significant at the five percent level).

Table 5 presents results from estimating Eq. 7. Panel A (Panel B) presents results for our sample annual and quarterly (current) reports. Across all measures and form types, we find evidence of statistically and economically significant market reactions on the disclosure event day. For example, estimates of β_1 for *AbsoluteReturn* range from 0.465 to 1.190 with t-stats ranging from 12.17 to 18.08. These estimates represent an economically significant 22.8% to 58.6% increase in *AbsoluteReturn* relative to the sample mean. In contrast, examining *Length* as a measure of information, we find inconsistent results (t-stats range from -1.853 to 4.283). We also find moving from the lowest to the highest decile of *Info* measure is associated with a highly statistically and economically significant market reaction. For example, columns (3) and (6) in Panel B indicate the corresponding 71.4% and 143% increases in *AbsoluteReturn* and *Volume*, respectively. Further, we find that estimates of β_1 shrink considerably, e.g., in columns (3) and (6), they represent just 23.0% and 33.0% increases, respectively, versus

58.6% and 104% in columns (1) and (4), respectively.

One can interpret the magnitude of β_2 relative to β_1 as the proportion of the market reaction explained by the information in narrative content. For example, if our measure was perfect and the information in narrative content was responsible for 90% of the market reaction to the announcement, then we would expect β_2 to be roughly nine times larger than β_1. For an imperfect measure of textual information content, however, β_1 is expected to absorb a portion of the market reaction attributable to narrative content. In this sense, the magnitude of β_2 relative to β_1 puts a lower bound on the portion of the market reaction attributable to information in narrative content. Across three of four specifications (Columns (3) and (6) in Panels A and B), we find that β_2 is two to four times larger than β_1. These results are consistent with the notion that the market reaction is *primarily* explained by the information provided in narrative content.

4.3. Informative vs. Uninformative Sentiment and Signed Information

In our second set of tests, we examine whether the market reaction to the sentiment of the filing depends on the disclosed textual information content. A large literature in finance and accounting studies how markets respond to disclosures sentiment (see, e.g., Loughran and McDonald, 2016, for review). This literature generally finds that greater negative sentiment explains stock price movements. However, the sentiment of disclosure will only explain returns to the extent it is associated with the revelation of information. For example, the negative sentiment of MD&A is unlikely to trigger market reactions if the same information had previously been communicated in the prior press release. We thus examine the extent to which there is an interaction between sentiment and information when explaining stock returns upon the announcement.

We measure sentiment using the negative and positive dictionaries of Loughran and McDonald (2011). For each word in a disclosure, we sign the sentiment as negative one, one, or zero based on whether the word occurs in the negative dictionary, positive dictionary, or

neither, respectively. Following the terminology of Loughran and McDonald (2011), we refer to this signed sentiment as *FinNeg*. We then interact *FinNeg* with the information measured in Eq. 6. Given this token-by-token interaction between sentiment and information, we follow the same general approach used to calculate *Info*, i.e., we take the geometric mean over the disclosure, sort the measure into deciles, and scale it to range from minus one to one. We refer to the resulting variable *Info* × *FinNeg* as signed information or information-weighted sentiment.

Figure 6 plots the average daily values of *AbnormalReturn* and *Volume* in the [-10, 10] day window around the filing date separately for filings that are above and below median *Info* × *FinNeg*.[12] Panel A presents results for annual and quarterly reports, Panel B presents them for current reports, and Panel C presents the results based on the Management's Discussion and Analysis. Across all panels and measures, there is a distinct spread in the abnormal return on the filing date, with below (above) median *Info* × *FinNeg* filings experiencing negative (positive) abnormal returns.

To the extent our measure is effective at capturing the information content of narrative disclosures, we expect to observe two empirical phenomena. First, the *AbnormalReturn* associated with the filing should be a monotonically increasing function of the information-weighted sentiment, *Info* × *FinNeg*. In the case of *Volume*, we expect a convex relation. Second, we expect that the slope of the relation between *AbnormalReturn* and *Info* × *FinNeg* is greater for filings that contain more information according to our measure. In the case of *Volume*, we expect the convexity of the relation to increase. Figure 7 plots the average value of *AbnormalReturn* and *Volume* for $t = 0$ as a function of *Info* × *FinNeg*. Across all form types, we find that *AbnormalReturn* (*Volume*) is generally an increasing (convex) function of *Info* × *FinNeg* and that the slope (convexity) of this relation is steeper for more informative

[12] Due to the manner in which we construct *Info* × *FinNeg* this equates to positive and negative values of *Info* × *FinNeg*.

filings.

To test for a statistically significant relation between *AbnormalReturn* and *Info* × *FinNeg*, we estimate the following regression pooling across all firm-dates in the [-10, 10] day window around the filing date:

$$AbrnomalReturn_{f,t} = \beta_1 Event[0] + \beta_2 Event[0] \times Info_{f,t} \times FinNeg_{f,t} + \boldsymbol{\gamma}\mathbf{X}_{f,t} + \delta_t + \lambda_{f,t} + \epsilon_{f,t} \quad (8)$$

where *AbnormalReturn* as the daily residual from the expected returns model of Fama and French (2015) and Carhart (1997) estimated over the [-70,-11] day window relative to the event, expressed as a percent. *FinNeg* is the count of the number of negative words in the filing scaled by the length of the filing (Loughran and McDonald, 2011) multiplied by negative one, then deciled and scaled to range from negative one to one. As before, β_1 and β_2 are our coefficients of interest. However, β_1 now represents the average abnormal return to filings in the lowest decile of *Info* × *FinNeg* while β_2 represents the incremental abnormal return associated with moving from the lowest to highest decile of *Info* × *FinNeg*.

Table 6 Panel A (Panel B) presents results for annual and quarterly (current) reports. Across form types, the results in column (1) indicate a statistically significant market reaction, on average. However this reaction is not economically significant (only -5.4 and 7.7 bps for Forms 10-K/Q and Form 8-K, respectively). Column (2) presents results for when *FinNeg* is interacted with the announcement event. We find that filings in the lowest decile of *FinNeg* are associated with a statistically significant negative event day abnormal return of 7.6 to 10.4 basis points (t-stats range from -3.071 to -2.281). Moving from the lowest to highest decile of *FinNeg* is associated with a statistically significant 18.1 to 26.5 basis points increase in the abnormal return on the event day (t-stats range from 2.055 to 5.870). Notably, these results are similar in magnitude to the findings in Loughran and McDonald (2011) who study solely 10-K filings. Column (3) presents results for the information-weighted sentiment measure, *Info* × *FinNeg*. We find that filings in the lowest decile of *Info* × *FinNeg* are associ-

ated with a statistically *and economically* significant negative event day abnormal return of -254.9 to -156.9 basis points (t-stats range from -49.525 to -24.655). Moving from the lowest to highest decile of the information-weighted sentiment, $Info \times FinNeg$, is associated with a statistically significant and economically massive 318.8 to 524.5 basis points increase in the abnormal return on the event day (t-stats range from 10.305 to 13.393). Finally, in column (4), we include both $FinNeg$ and $Info \times FinNeg$ in the regression. We find that our inferences in terms of statistical and economic significance are generally unchanged.

Collectively, the evidence in Table 6 highlights the importance of measuring ***information*** in the textual data to understand the sentiment. Examining solely 10-K filings, Loughran and McDonald (2011) find a roughly 25 bps spread between the extreme quintiles of negative sentiment (p. 51), which we confirm. However, when we account for the interaction between information and sentiment, i.e., consider sentiment over the informative portion of the document, the effect is more than an order of magnitude greater, i.e., between 318.7 and 524.5 basis points. This is consistent with the notion that (i) our measure is capturing the *information* in narrative content rather than some other latent time-varying firm characteristic, and (ii) the sentiment of non-informative content should not influence prices.

4.4. Long-Run Fundamental Performance

The results of our earlier tests focus on the link between our measure of narrative information content and market reactions. In our third set of tests, we examine the extent to which our measure of information in textual disclosures is predictive of future firm fundamental performance. We estimate the following model:

$$y_{f,q+t} = \beta_1 Info_{f,q} \times FinNeg_{f,q} + \boldsymbol{\gamma}\mathbf{X}_{f,q} + \delta_q + +\epsilon_{f,q} \tag{9}$$

where $y_{f,q+t}$ is a measure of firm performance for firm f in quarter $q+t$. We examine the predictive ability of our measure for horizons ranging from one quarter ahead ($t = 1$), to

four years ahead ($t = 16$). We tabulate results for return on assets, *ROA* (defined in Table 7), as a measure of firm performance. However, our inferences are similar if we use other measures such as those in Cohen et al. (2020). As before, $Info_{f,q} \times FinNeg_{f,q}$ is the signed information, i.e., information-weighted sentiment, for firm f. We aggregate this measure over firm f's filings for quarter q. In the case of Form 10-K/Q filings, $Info_{f,q} \times FinNeg_{f,q}$ is the measure that we calculated in the prior subsection. In the case of Form 8-K filings, we compute the arithmetic mean of $Info_{f,q} \times FinNeg_{f,q}$ over all filings released during the interval $[q-1, q)$.[13] $\mathbf{X}_{f,t}$ is a vector of control variables for firm f in quarter q (Defined in Table 7 and enumerated in Table 8). δ_q is a time fixed effect. β_1 is our coefficient of interest and represents the incremental change in firm performance associated with moving from the lowest to the highest decile of information-weighted sentiment $Info_{f,q} \times FinNeg_{f,q}$.

Notably, Eq. 9 defines a *predictive* regression regarding the value of $y_{f,q+t}$ based solely on information available in quarter q.[14] As a result, introducing a time-invariant fixed effect into the regression is inappropriate since doing so would require foresight into future realizations. For this reason, we estimate two versions of Eq. 9: a levels specification and a differences specification. The benefit of the latter specification is that any time-invariant firm characteristics, which may be responsible for firm performance, are differenced out.

Table 8 presents results. Panels A and B (Panels C and D) present results for our "levels" and "differences" specifications, respectively, for annual and quarterly (current) reports. Across all form types and model specifications, we find evidence of a statistically and economically significant relation between *Info* × *FinNeg* and future firm performance as measured by *ROA*. For example, in Panel A, we find that moving from the lowest to highest decile of *Info* × *FinNeg* is associated with an 11% increase in *ROA* during the quarter of disclo-

[13] In untabulated analyses, we find that the choice of arithmetic mean versus a weighted combination does not change our inferences. As such, we opt to use the simplest aggregation.

[14] For our analysis of annual and quarterly reports, the Form 10-K/Q filing over which we compute information comes out some time after q but generally before $q+1$. In that sense, the data required to make our predictions were not available until some time after q but before $q+1$.

sure. We find that this relation is persistent but declining in economic magnitude over time. Specifically, one, two, three, and four years later the relation is associated with 8.3%, 6.0%, 4.6%, and 4.2% increases in *ROA*, respectively. We find a similar, albeit less economically meaningful, pattern for current reports. Specifically, in Panel C, we see that during the same quarter the relation is associated with an increase of 5.6%, while one, two, three, and four years later, it is associated with 3.3%, 2.1%, 0.9%, and 0.8% increases in *ROA*, respectively. Our results from the differences specifications (Panels B and D) suggest that this pattern cannot be attributed to an omitted time-invariant firm characteristic.

4.5. Long-Run Return Predictability

Our final set of tests examines the relation between our measure of signed information and long-run returns. We test the predictive ability of $Info \times FinNeg$ for long-run returns by re-estimating Eq. 9 using quarterly abnormal returns, *BHAR*, for $y_{f,q+t}$. Within each firm-quarter in our sample, we estimate the expected returns model of Fama and French (2015) and Carhart (1997) using daily returns data. We then define *BHAR* as the intercept from the expected returns model, i.e., the "alpha," multiplied by the number of trading days during the quarter and expressed as a percent. Notably, this approach controls for time-varying firm characteristics, e.g., risk exposure, which vary at the firm-year-quarter frequency or slower. Further, since some firm characteristics such as size and value are controlled for via factor loadings, we do not include them when re-estimating Eq. 9.

In this specification, β_1 is our coefficient of interest. It represents the incremental change in *BHAR* associated with moving from the lowest to highest decile of $Info_{f,q} \times FinNeg_{f,q}$. One can also broadly interpret β_1 as the return on a portfolio which is long firms with above median $Info_{f,q} \times FinNeg_{f,q}$ and short those below the median. However, this interpretation needs qualification due to the presence of control variables and year-fixed effects.

On the one hand, if markets are efficient in the sense that all publicly available information is fully reflected in price then we do not expect to find a statistically significant relation

between *BHAR* and signed information. On the other hand, if markets are inefficient in the sense that there is information contained in Form 10-K/Q and 8-K filings that is not fully priced, e.g., Cohen et al. (2020), then we expect to find evidence of a statistically and ***economically*** significant relation between *BHAR* and signed information. Finally, if markets are relatively efficient but there are limits to arbitrage such as trading costs in the form of spreads and price impact, then we expect to find a statistically ***but not economically*** significant relation.

Table 9 presents results. Across all form types and model specifications, we find evidence of a statistically significant association between *BHAR* and $Info_{f,q} \times FinNeg_{f,q}$, with t-statistics ranging from 3.404 to 9.799. This relation, however, is economically small with the effect running in magnitude from 35.2 down to 7.30 basis points per quarter (or 141 to 29 basis points per year). This evidence is consistent with the notion that the market impounds the information contained in narrative disclosure within a matter of days. One explanation for the differences between our results and those of Cohen et al. (2020), is that we measure information on a more frequent basis, i.e., via current reports, whereas Cohen et al. (2020) focus on quarterly information releases.

5. Conclusion

Although narrative disclosure, i.e., the written text, is a frequent form of communication by firms to outside stakeholders, most studies examining the reaction to new information have focused on quantitative information such as earnings. In this paper, we draw on the foundations of information theory and the emergence of large language models to measure investors' prior beliefs about disclosure. We then use these priors to uncover new information contained in textual data and study the value of narrative content in capital markets.

We begin by training a 774 million parameter large language model (LLM) on a sample of Form 10-K/Q and 8-K filings from 1996 through 2006. We then model firm-specific prior

beliefs by fine-tuning this model on the time series of individual firms' disclosures. In a set of descriptive analyses, we use these priors to compute our measure on a granular basis, i.e., word by word. We then examine where, within firms' filings, new textual information is comparatively more likely to occur. We find that for annual and quarterly reports, firms tend to provide the most information early on in the main filing. In contrast, for current reports (Form 8-K), we find that most of the information is contained in exhibits attached to the filing.

In the first set of tests, we examine the extent to which our measure explains market reactions to the filings posted on SEC EDGAR database. We find that our measure of information uncovered from the narrative communication explains the majority of the market reaction that takes place upon the announcement. This indicates that narrative content is a critical mechanism for communicating key information about a firm. Next, we examine whether our measure, when signed based on information sentiment, can explain the market reactions. We measure the signed information in a filing on a word-by-word basis by interacting the information content of a given word with its sentiment and subsequently aggregating words into sentences and sentences into a document. Comparing filings with low versus high information, we find the relation between sentiment and announcement abnormal returns (trading volume) is amplified when a document contains higher textual information content. We also examine the predictive ability of our signed information measure with respect to long-run firm fundamental performance. We find that the signed information is statistically and economically predictive of future firm performance vis a vis return on assets. We also find a statistically significant and economically moderate relationship between the signed information and future stock returns. The long-term predictability persists up to sixteen quarters into the future.

Lastly, we conduct a test of the efficient markets hypothesis by examining whether our measure is predictive of statistically and economically significant future abnormal returns. Consistent with market efficiency up to the limits of arbitrage, we find a statistically signif-

icant relation which is economically insignificant.

Collectively, our evidence suggests that LLMs are an effective tool for modeling investors' priors over disclosure. While using LLMs to model priors appears to be a fruitful direction, we only consider one specific type of LLM, i.e., a decoder model that follows GPT-2 architecture. Additionally, we use a comparatively small LLM, e.g., just 774 million parameters whereas GPT-3's 175 billion. More importantly, we do not roll the model forward in time, i.e., we do not revise the parameter estimates as the future unfolds. Relaxing these constraints offers a number of interesting avenues for future work.

References

Araci, D. (2019). Finbert: Financial sentiment analysis with pre-trained language models.

Ball, R. and Brown, P. (1968). An empirical evaluation of accounting income numbers. *Journal of Accounting Research*, 6(2):159–178.

Beaver, W. H. (1968). The information content of annual earnings announcements. *Journal of accounting research*, pages 67–92.

Bernard, V. L. and Thomas, J. K. (1989). Post-earnings-announcement drift: Delayed price response or risk premium? *Journal of Accounting Research*, 27:1–36.

Bonsall, S. B., Leone, A. J., Miller, B. P., and Rennekamp, K. (2017). A plain english measure of financial reporting readability. *Journal of Accounting and Economics*, 63(2):329–357.

Brown, S. V. and Tucker, J. W. (2011). Large-sample evidence on firms' year-over-year MD&A modifications. *Journal of Accounting Research*, 49(2):309–346.

Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D. M., Wu, J., Winter, C., Hesse, C., Chen, M., Sigler, E., Litwin, M., Gray, S., Chess, B., Clark, J., Berner, C., McCandlish, S., Radford, A., Sutskever, I., and Amodei, D. (2020). Language models are few-shot learners.

Carhart, M. M. (1997). On persistence in mutual fund performance. *Journal of Finance*, LII:57–82.

Chen, Y., Kelly, B. T., and Xiu, D. (2022). Expected returns and large language models. *Social Science Research Network*.

Cohen, L., Malloy, C., and Nguyen, Q. (2020). Lazy prices. *Journal of Finance*, 75:1371–1415.

Demski, J. S. (1973). The general impossibility of normative accounting standards. *The Accounting Review*, 48(4):718–723.

Devlin, J., Chang, M.-W., Lee, K., and Toutanova, K. (2019). Bert: Pre-training of deep

bidirectional transformers for language understanding. In *North American Chapter of the Association for Computational Linguistics*.

Fama, E. F. (1970). Efficient capital markets: A review of theory and empirical work. *The Journal of Finance*, 25(2):383–417.

Fama, E. F., Fisher, L., Jensen, M. C., and Roll, R. (1969). The adjustment of stock prices to new information. *International economic review*, 10(1):1–21.

Fama, E. F. and French, K. R. (2015). A five-factor asset pricing model. *Journal of Financial Economics*, 116:1–22.

Gjesdal, F. (1981). Accounting for stewardship. *Journal of Accounting Research*, 19(1):208–231.

Hoffmann, J., Borgeaud, S., Mensch, A., Buchatskaya, E., Cai, T., Rutherford, E., De, D., Casas, L., Hendricks, L. A., Welbl, J., Clark, A., Hennigan, T., Noland, E., Millican, K., Driessche, G. V. D., Damoc, B., Guy, A., Osindero, S., Simonyan, K., Elsen, E., Rae, J. W., Vinyals, O., and Sifre, L. (2023). Training compute-optimal large language models.

IDC (2023). Worldwide idc global datasphere forecast, 2023-2027.

Ji, Z., Lee, N., Frieske, R., Yu, T., Su, D., Xu, Y., Ishii, E., Bang, Y. J., Madotto, A., and Fung, P. (2023). Survey of hallucination in natural language generation. *ACM Comput. Surv.*, 55(12).

Kim, A. G., Muhn, M., and Nikolaev, V. V. (2023). Bloated disclosures: Can ChatGPT help investors process information? *SSRN Electron. J.*

Lopez-Lira, A. and Tang, Y. (2023). Can ChatGPT forecast stock price movements? return predictability and large language models. *SSRN Electron. J.*

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.

Loughran, T. and McDonald, B. (2016). Textual analysis in accounting and finance: A survey. *Journal of Accounting Research*, 54(4):1187–1230.

Radford, A., Narasimhan, K., Salimans, T., and Sutskever, I. (2018). Improving language

understanding by generative pre-training.

Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., and Sutskever, I. (2019). Language models are unsupervised multitask learners.

Sarkar, S. and Vafa, K. (2024). Lookahead bias in pretrained language models. *SSRN Electron. J.*

SEC (2019). Registration under the securities act of 1933. https://www.investor.gov/introduction-investing/investing-basics/glossary/registration-under-securities-act-1933. Accessed: 2023-11-07.

Sennrich, R., Haddow, B., and Birch, A. (2016). Neural machine translation of rare words with subword units. In Erk, K. and Smith, N. A., editors, *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1715–1725, Berlin, Germany. Association for Computational Linguistics.

Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal*, 27(3):379–423.

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

Appendices

A. Details of Narrative Content Extraction

Notably, the filings we study usually contain a mix of narrative and quantitative content such as tabular financial statements. Since our focus is on the former, we extract narrative content using one of two approaches. The particular approach depends on the file type–plain text or HTML–and leverages specific details of each in order to accurately extract narrative content.

Plain text filings and attachments–which are systematically older filings on average–often contain standardized general markup language (SGML) tags. Historically, these tags were used to identify special content, e.g., tabular content, and render it in a more human-readable format based on the capabilities of the system used to view the content, e.g., a personal computer versus a Bloomberg Terminal. We leverage these tags to identify tabular quantitative content in plain text filings and remove it.

For HTML filings and attachments, firms often use tables to structure content, e.g., present a graph on one half of the page while typesetting narrative content on the other half. This makes the task of removing tabular quantitative content comparatively more difficult. Prior literature has removed tables when the more than 25% of non-blank characters in a table are numbers (e.g., Loughran and McDonald (2011)). We take a similar approach but instead measure the number of characters–as rendered by a web browser–per HTML tag, CPT. This measure is commonly used in the computational linguistics literature to measure the text density of HTML-based content.

We select a CPT threshold of 10 to identify narrative content embedded in HTML tables, i.e., when the number of non-numeric, non-blank, and non-punctuation characters per tag in a given table exceeds 10 then we extract the narrative content from the table and ignore the table otherwise. Looking at a random sample of tables, we find that this threshold achieves

both higher precision and recall than counting the fraction of numeric content in a given table. Anecdotally, this appears to be driven by firms using HTML tables to enumerate *narrative* summaries of their performance. For example, the phrase "2022 Q2 subscriber growth was 3.354 MM versus 2.939 MM in 2021 Q1" might appear in a table as part of an enumerated list. In this case, 34% of the non-blank characters are numeric, thus the table should be removed according to rules in the prior literature. In contrast, CPT is well above the threshold of 10, thus the statement would be retained as desired for our study.

After removing tables where $CPT < 10$, we extract the text from the remaining HTML while preserving meaningful formatting such as indentation and line breaks. We find that this approach leads to a corpus of plain text documents that are structurally similar regardless of source file type or the time period when the form was filed, e.g., prior to widespread adoption of HTML-based filings.

B. Examples of Disclosure Generated by the LLMs

In this section we present examples of narrative disclosures generated by conditioning the LLMs we train on a particular prompt and then sampling from the model. First, we condition the models with a prompt that is similar to what the SEC recommends for the start of Item 1. Business Description, presented in Table B1. After conditioning we then sample from the model using beam search. Table B2 presents examples generated using our cross-sectional LLM described in Section 3.4.1. Table B3 presents examples generated using a sample of the firm-specific LLMs described in Section 3.4.2.

Table B1. Examples of Disclosures Generated by the LLMs

Item 1. Business Description
Our principal operating activities consist of*[…]*

Table B2. Examples of Disclosures Generated by the LLMs

Sample	Generated Text
1	*[…]*the design, development, manufacture, marketing and sale of digital radio systems. Our principal research and development resources are located in the United States, Belgium, Germany and Canada.*[…]*
2	*[…]*the production, storage, processing and sale of natural gas as well as associated operating, general, and administrative costs necessary to provide benefits to our customers and to operate our business, including our marketing of natural gas, storage and processing, gas gathering, and transportation services.*[…]*
3	*[…]*developing and marketing our proprietary drug delivery technologies, research and development of drug delivery technologies associated with drug compounds and other non-topharmaceutical products and the provision of related clinical, regulatory, manufacturing and other related services. For the year ended December 31, 2004, revenues represented $13.4 million or 95% of our expenses.*[…]*
4	*[…]*the operation of our network infrastructure, including those related to our satellite service; data centers and communications facilities, including those located in Europe, Latin America and Canada; sales and marketing; and administrative activities.*[…]*

Table B3. Examples of Disclosures Generated by the LLMs

Firm	Generated Text
Apple	*[…]*the design, development, marketing and sale of personal computers and related software, the activities of which are conducted throughout the Company's operating segments. The Company's products consist of the Macintosh line of computers, related peripherals and services, and the Mac OS® X operating system (the Macintosh). The Company sells its products worldwide through its online stores, its retail stores, its direct sales force, and third-party wholes*[…]*
Cummins	*[…]*the operations of our Engine Segment, which is described in Note 17 to the Consolidated Financial Statements included in our 2004 Annual Report on Form 10-K. The Engine Segment produces engines and parts for sale to customers in on-highway, commercial, institutional and mining applications. Our Power Generation Segment produces engines and parts for power generation equipment and is headquartered in Columbus, Indiana. We have manufacturing plants worldwide, including major operations*[…]*
Pfizer	*[…]*the discovery, development, manufacture, marketing and sale of prescription medicines for humans and animals as well as our over-the-counter products. We also purchase products from numerous suppliers. These purchases are generally made through various means including through our operating subsidiaries. Wherever possible, commercial relationships are maintained with our suppliers. Where possible, these relationships can be terminated early with 90 days' notice. We believe that we have good relationships with our suppliers. We made significant progress during 1999 in the development of our pharmaceutical products, manufacturing processes and commercialized products. We have successfully enhanced these systems and are implementing significant changes in our internal systems and processes including: • Alond, a treatment for schizophrenia, was launched in September 1999 to increase capacity and efficiency in our manufacturing operations • Exubera, our inhaled insulin medication for chronic obstructive*[…]*
Weyco Group	*[…]*the manufacture and sale of men's casual, dress and fashion footwear. The majority of foreign-sourced purchases are denominated in U.S. dollars. The Company also has a small number of retail stores in Europe and, accordingly, the Company's results are primarily impacted by the economic conditions and the retail environment in the United States. We have historically generated adequate cash flow from operations to meet working capital requirements. In 2003, cash flows from operations was $7.5 million, as compared to $6.7 million last year.*[…]*

Figure 1. Model Training and Inference

This figure presents the data partitions used for model training and inference. The dotted line encompasses the data used for training the cross-sectional model described in Section 3.4.1. The dashed line encompasses the data used for fine-tuning the cross-sectional model on one particular firm's disclosures, as described in Section 3.4.2. Firm-specific models are applied to their respective disclosures from 2007 through 2022 to generate out-of-sample estimates of the information contained in each disclosure.

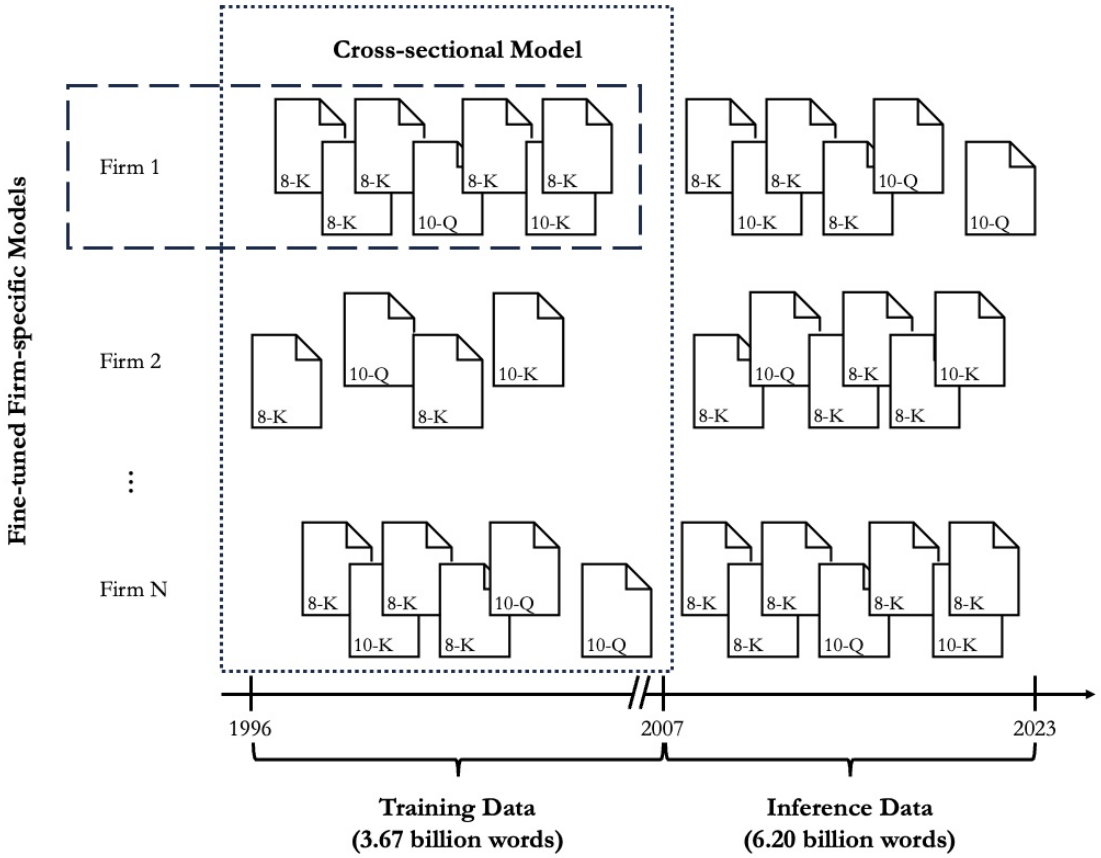


Figure 2. Example of Model Inference for First Mention of Apple iPhone

This figure presents inference results from a model fine-tuned on Apple's disclosures from 1996 through 2006. Each sentence is highlighted in green based on the geometric mean of the model's token-by-token measurement of information, with darker green representing more information. The disclosure was filed with the SEC via EDGAR on January 17, 2007.

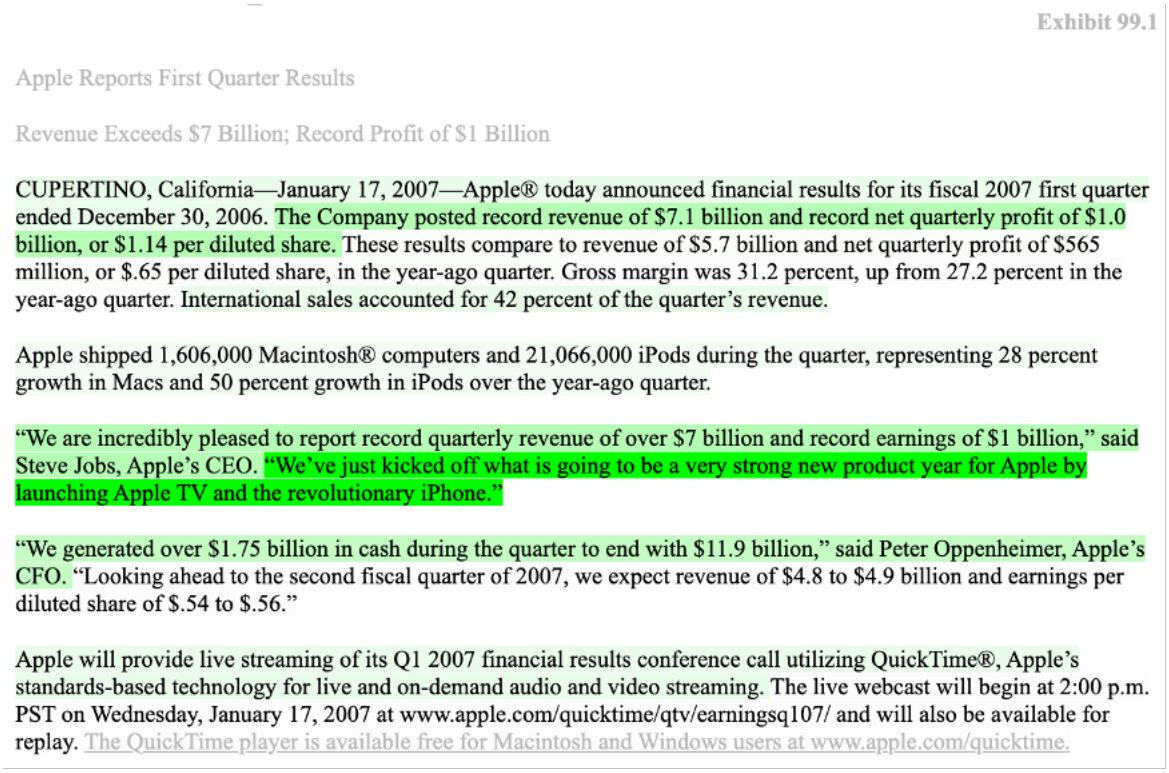


Exhibit 99.1

Apple Reports First Quarter Results

Revenue Exceeds $7 Billion; Record Profit of $1 Billion

CUPERTINO, California—January 17, 2007—Apple® today announced financial results for its fiscal 2007 first quarter ended December 30, 2006. The Company posted record revenue of $7.1 billion and record net quarterly profit of $1.0 billion, or $1.14 per diluted share. These results compare to revenue of $5.7 billion and net quarterly profit of $565 million, or $.65 per diluted share, in the year-ago quarter. Gross margin was 31.2 percent, up from 27.2 percent in the year-ago quarter. International sales accounted for 42 percent of the quarter's revenue.

Apple shipped 1,606,000 Macintosh® computers and 21,066,000 iPods during the quarter, representing 28 percent growth in Macs and 50 percent growth in iPods over the year-ago quarter.

"We are incredibly pleased to report record quarterly revenue of over $7 billion and record earnings of $1 billion," said Steve Jobs, Apple's CEO. "We've just kicked off what is going to be a very strong new product year for Apple by launching Apple TV and the revolutionary iPhone."

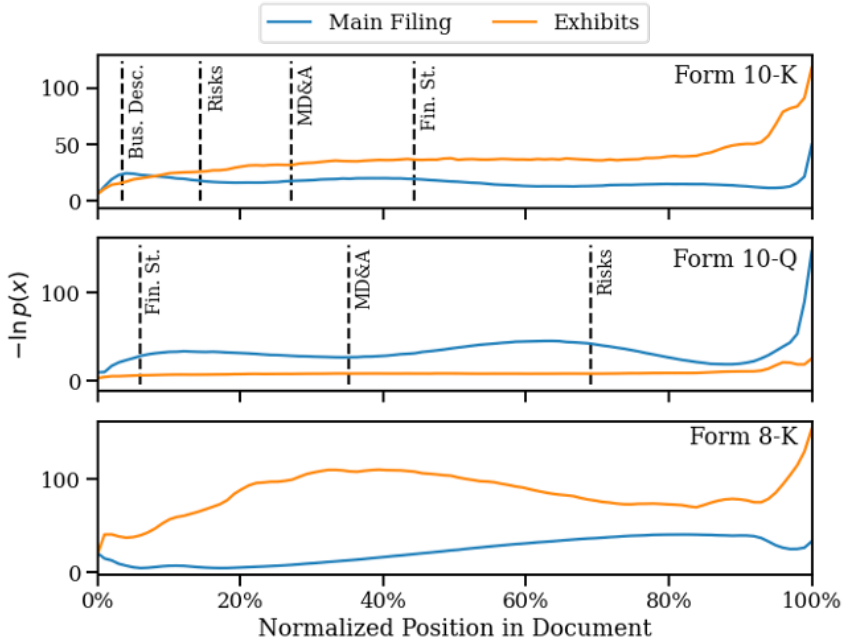
"We generated over $1.75 billion in cash during the quarter to end with $11.9 billion," said Peter Oppenheimer, Apple's CFO. "Looking ahead to the second fiscal quarter of 2007, we expect revenue of $4.8 to $4.9 billion and earnings per diluted share of $.54 to $.56."

Apple will provide live streaming of its Q1 2007 financial results conference call utilizing QuickTime®, Apple's standards-based technology for live and on-demand audio and video streaming. The live webcast will begin at 2:00 p.m. PST on Wednesday, January 17, 2007 at www.apple.com/quicktime/qtv/earningsq107/ and will also be available for replay. The QuickTime player is available free for Macintosh and Windows users at www.apple.com/quicktime.

Figure 3. Location of Information within Filings

This figure plots the average information as a function of the normalized position in the document. Dashed vertical lines represent the average starting positions of common Form 10-K and 10-Q items. Panel A presents averages separately for the main filing and exhibits of Forms 10-K, 10-Q, and 8-K. Panel B presents averages separately for filings with above and below median information. Sample of 40,043,600 form-attachment-positions.

Panel A. Main Filing vs. Exhibits



Panel B. By Volume of Information

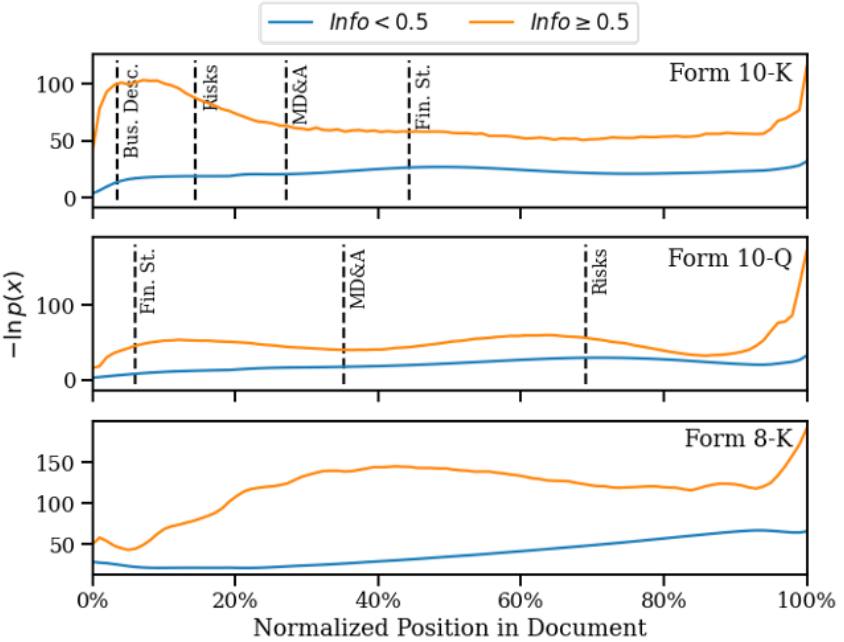
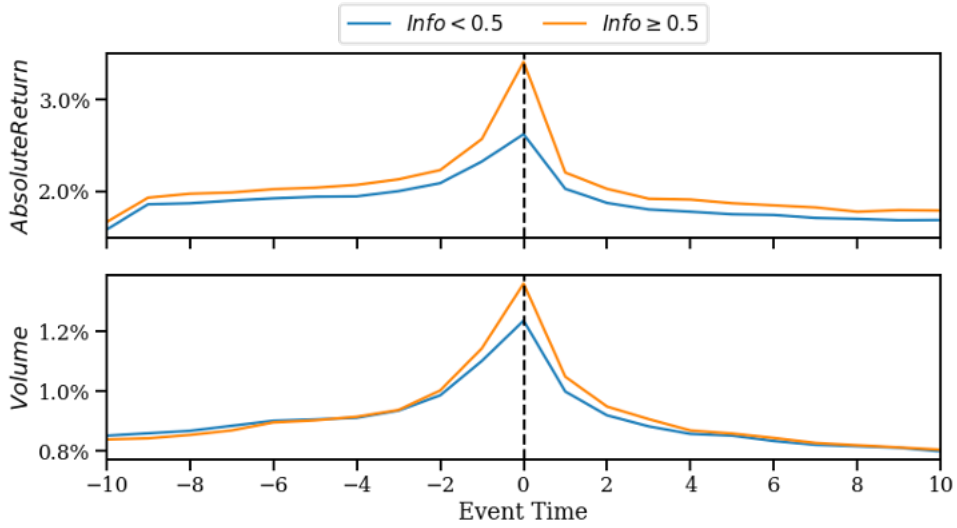


Figure 4. Market Reaction Event Study

This figure plots the average value of daily absolute returns and trading volume in the [-10, 10] window around the form filing date separately for filings with above and below median *Info*. Day 0 represents the first day markets were open following acceptance of the filing on EDGAR. Specifically, Day 0 for filings accepted either prior to markets opening or intraday is the filing date, whereas Day 0 for filings accepted after markets close is the date of the following trading session. Panel A presents results for a sample of 1,856,705 form-days in the [–10,+10] day window around 90,437 Form 10-K/Q filings. Panel B presents results for a sample of 4,984,954 form-days in the [–10,+10] day window around 245,993 current report filings. Panel C presents results for a sample of 1,734,976 form-days in the [–10,+10] day window around 84,484 Management's Discussion and Analysis (MD&A) items extracted from Form 10-K/Q.

Panel A. Annual and Quarterly Reports (Form 10-K/Q)



Panel B. Current Reports (Form 8-K)

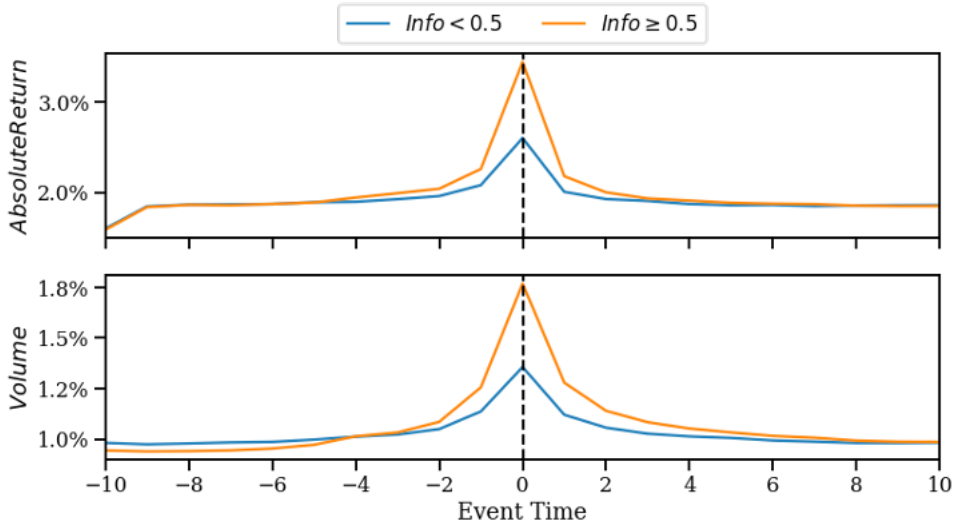


Figure 4. Market Reaction Event Study (Cont'd)

Panel C. Management's Discussion and Analysis (Extracted from Form 10-K/Q)

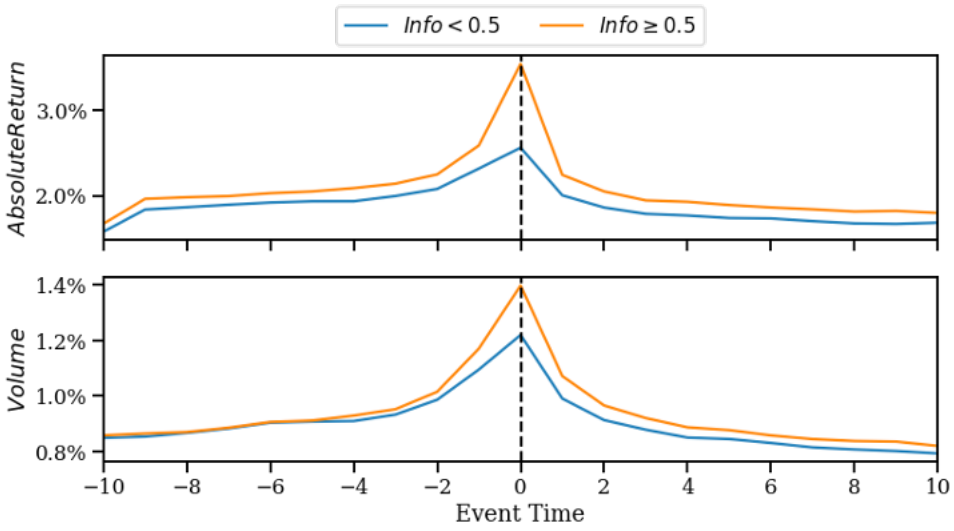
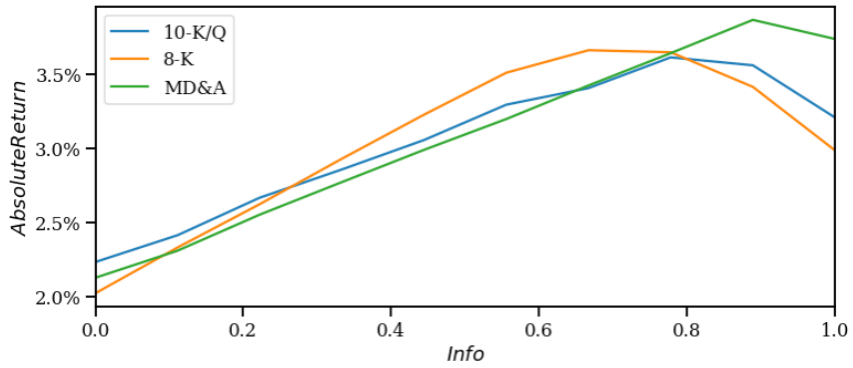


Figure 5. Market Reaction and Information

This figure plots the average market reaction to filings as a function of *Info*. The market reaction is measured on the first day markets were open following acceptance of the filing on EDGAR. Specifically, the market reaction to filings accepted either prior to markets opening or intraday is measured on the same day, whereas the reaction to filings accepted after markets close is measured the following trading session. Panel A (Panel B) presents the average value of *AbsoluteReturn* (*Volume*) by decile of *Info*. Sample of 90,473 Form 10-K/Q filings, 245,993 Form 8-K filings, and 84,484 Management's Discussion and Analysis (MD&A) items extracted from Form 10-K/Q.

Panel A. AbsoluteReturn



Panel B. By Volume

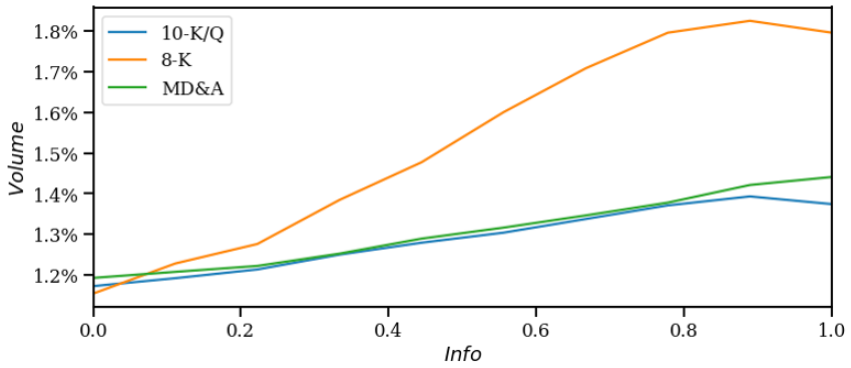
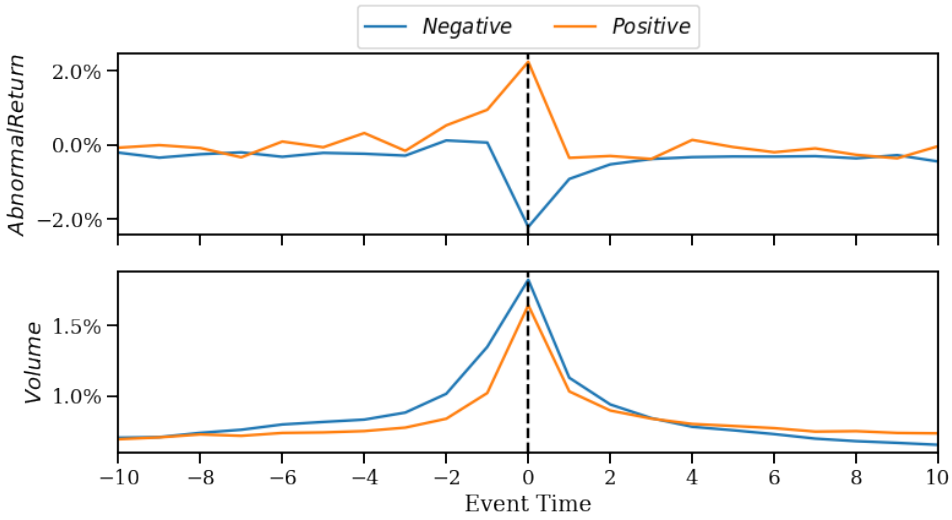


Figure 6. Market Reaction to Signed Information Event Study

This figure plots the average value of daily abnormal returns and trading volume in the [-10, 10] window around the form filing date separately for filings with above and below median *Info*×*FinNeg*. Day 0 represents the first day markets were open following acceptance of the filing on EDGAR. Specifically, Day 0 for filings accepted either prior to markets opening or intraday is the filing date, whereas Day 0 for filings accepted after markets close is the date of the following trading session. Panel A presents results for a sample of 4,263,714 form-days in the [–10,+10] day window around 203,034 Form 10-K/Q filings. Panel B presents results for a sample of 11,445,924 form-days in the [–10,+10] day window around 545,044 current report filings. Panel C presents results for a sample of 4,166,169 form-days in the [–10,+10] day window around 198,389 Management's Discussion and Analysis (MD&A) items extracted from Form 10-K/Q.

Panel A. Annual and Quarterly Reports (Form 10-K/Q)



Panel B. Current Reports (Form 8-K)

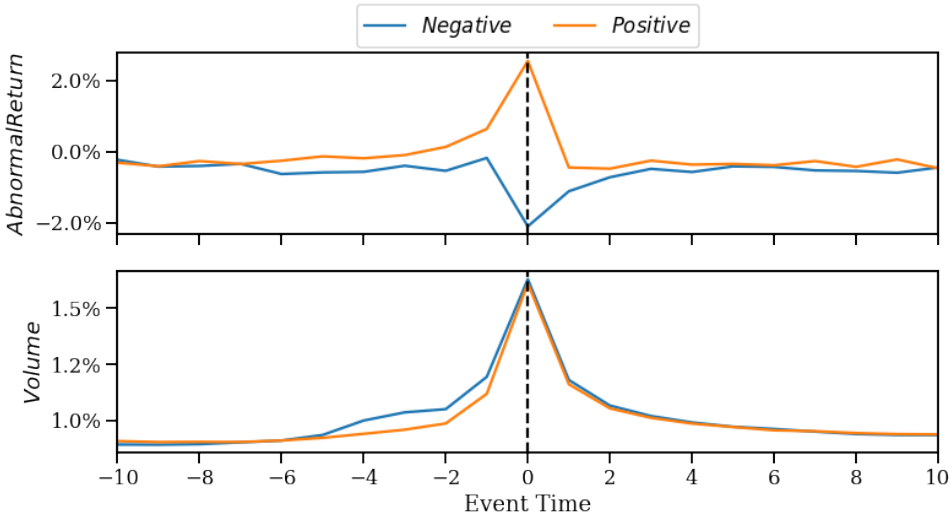


Figure 6. Market Reaction to Signed Information Event Study (Cont'd)

Panel C. Management's Discussion and Analysis (Extracted from Form 10-K/Q)

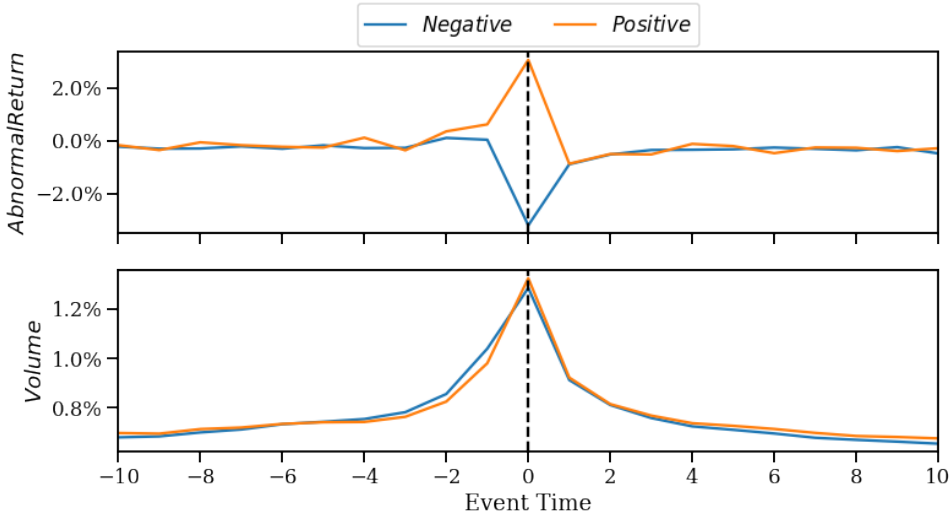
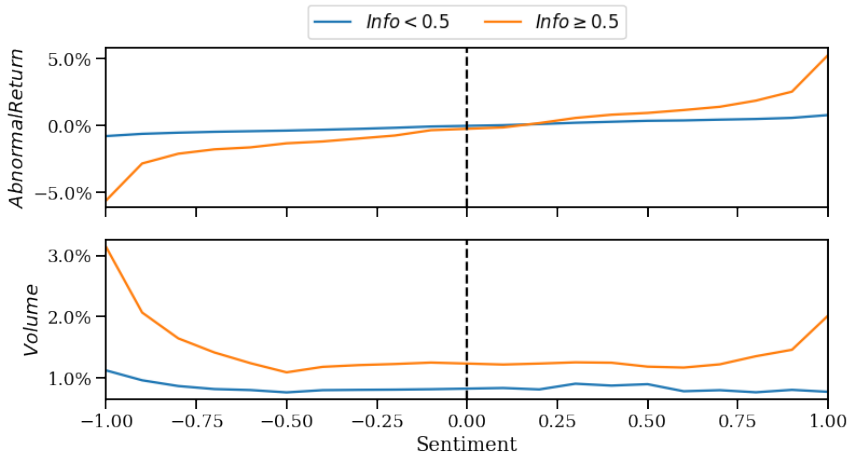


Figure 7. Market Reaction and Signed Information

This figure plots the average market reaction to filings as a function of *Info* × *FinNeg*. The market reaction is measured on the first day markets were open following acceptance of the filing on EDGAR. Specifically, the market reaction to filings accepted either prior to markets opening or intraday is measured on the same day, whereas the reaction to filings accepted after markets close is measured the following trading session. Panel A (Panel B) presents the average value of *AbsoluteReturn* (*Volume*) by decile of Info. Sample of 90,473 Form 10-K/Q filings, 245,993 Form 8-K filings, and 84,484 Management's Discussion and Analysis (MD&A) items extracted from Form 10-K/Q.

Panel A. Annual and Quarterly Reports (Form 10-K/Q)



Panel B. Current Reports (Form 8-K)

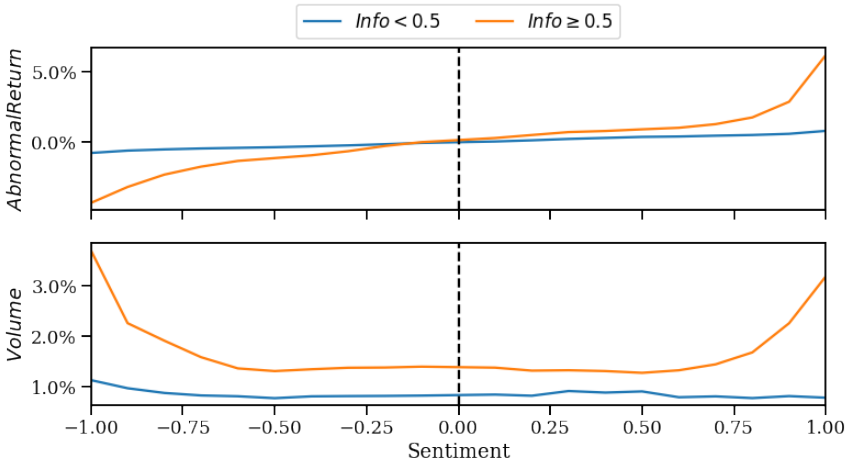


Figure 7. Market Reaction and Signed Information (Cont'd)

Panel C. Management's Discussion and Analysis (Extracted from Form 10-K/Q)

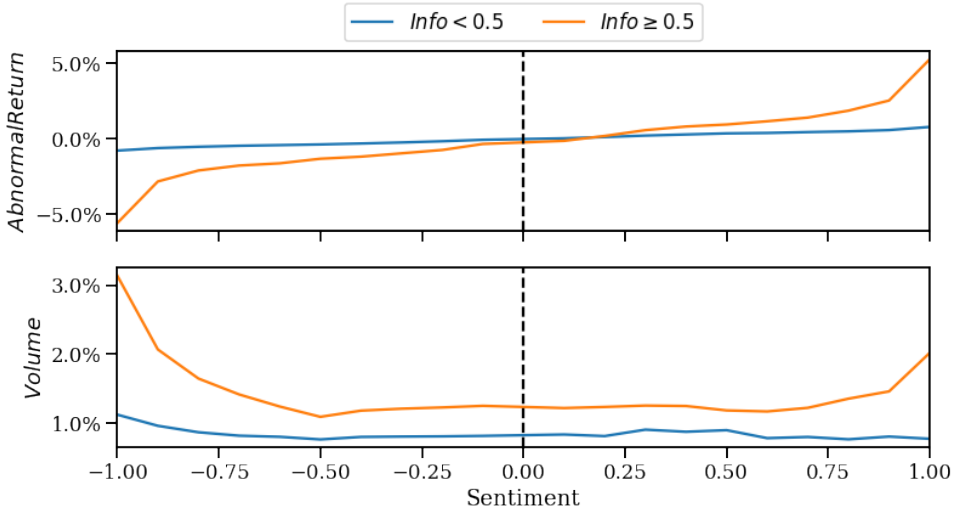


Table 1. **Volume of Narrative Disclosure on SEC EDGAR**

This table presents the volume of narrative disclosure filed on the Securities and Exchange Commission's Electronic Data Gathering and Retrieval (SEC EDGAR) system from 1996 through 2022 for the top 20 most voluminous form types.

Form Type	N-words (millions)	Cumulative Percent of All Disclosure
8-K	8,223	14%
10-Q	7,979	28%
10-K	6,675	40%
424B2	3,457	46%
424B3	3,045	51%
DEF 14A	2,918	57%
6-K	1,877	60%
S-4/A	1,612	63%
424B5	1,431	65%
20-F	1,419	68%
S-4	1,255	70%
S-3	1,007	72%
PRE 14A	648	73%
424B4	639	74%
S-3/A	605	75%
S-1/A	581	76%
DEFM14A	556	77%
FWP	541	78%
10-K405	527	79%
S-8	517	80%

Table 2. **Sample Construction**

This table presents the impact of various filters and random sampling methods on our sample construction. Final sample consists of 2,500 firms stratified by firm-size decile.

Panel A. Selection of Firms in Sample

Criteria	N-firms
Firms that filed a 10-K/Q on EDGAR in 2007	8,996
Requiring link to Compustat	4,374
Stratified sampling by firm size decile	2,500
Decile 0: [$0, $37]	250
Decile 1: ($37, $82]	250
Decile 2: ($82, $144]	250
Decile 3: ($144, $246]	250
Decile 4: ($246, $402]	250
Decile 5: ($402, $667]	250
Decile 6: ($667, $1,146]	250
Decile 7: ($1,146, $2,128]	250
Decile 8: ($2,128, $5,192]	250
Decile 9: ($5,192, >$5,192)	250

Panel B. Volume of Filings in Sample

Final Sample of Filings	Training: 1996-2006	Inference: 2007-2022
# of Form 10-K	18,906	22,778
# of Form 10-Q	71,520	67,659
# of Form 8-K	142,542	245,993
GB of narrative content	22.8	40.1
# of words (billions)	3.67	6.20

Table 3. Measure of Information

This table presents descriptive statistics for our measure of information. For each firm we train a firm-specific large language model (LLM) using filings from 1996 through 2006. We then apply the firm-specific models out-of-sample, i.e., on filings from 2007 onwards, to measure the information in new filings. Specifically, for each new filing we use the respective firm-specific LLM to measure the information token-by-token according to Eq. 6. We define the average information in a filing as the geometric mean of the set of token-by-token measurements. Sample of 22,778 Form 10-K, 67,659 Form 10-Q, and 245,993 Form 8-K filings submitted to EDGAR from 2007 through 2022.

	Distribution of Information					
	Annual Report (Form 10-K)		Quarterly Report (Form 10-Q)		Current Report (Form 8-K)	
	Main	Exhibits	Main	Exhibits	Main	Exhibits
Mean	14.60	9.54	16.83	5.19	7.15	35.64
Percentiles						
1-st	2.60	1.03	2.31	1.00	1.05	1.18
5-th	3.21	1.10	3.09	1.00	1.11	1.55
25-th	4.55	1.33	4.99	1.06	1.39	3.99
50-th	6.10	2.29	7.11	1.16	2.21	10.45
75-th	8.96	6.46	10.47	2.06	6.53	27.15
95-th	15.61	20.33	17.39	8.18	18.36	63.68
99-th	209.92	97.04	212.10	69.62	62.55	406.41

Table 4. Daily Event Study Descriptive Statistics

This table presents the distribution of variables used in our daily event study tests of market reaction partitioned by group and results for tests of differences in means and medians across groups. *AbsoluteReturn* is the absolute value of the daily residual from the expected returns model of Fama and French (2015) and Carhart (1997) estimated over the [-70,-11] window relative to the event, expressed as a percent. *Volume* is the ratio of daily volume of shares traded to shares outstanding, expressed as a percent. *AbnormalReturn* is the daily residual from the expected returns model of Fama and French (2015) and Carhart (1997) estimated over the [-70,-11] window relative to the event, expressed as a percent. *AdExp* is the ratio of advertising expense to sales. *CapEx* is the ratio of capital expenditures to total assets. *CapIntens* is the ratio of property, plant, and equipment to total assets. *CorpAcq* is an indicator variable equal to one if an acquisition accounts for at least 20% of sales. *Financing* is of the sum of equity and debt issuances over the fiscal quarter scaled by total assets. *IdioVol* is the standard deviation of the residual from a Fama and French (2015) and Carhart (1997) six-factor expected returns model using the daily returns over the prior fiscal quarter expressed as a percent. *Leverage* is the natural logarithm of the sum of long-term debt and debt in current liabilities scaled by total assets. *MTB* is the ratio of market to book value of equity. R&D is the ratio of research and development expense to sales. *QtrlyReturn* is the buy and hold return over the fiscal quarter expressed as a percent. *ShortTermReturn* is the buy and hold return during the window [t-5, t-1] for each date t expressed as a percent. *ShortTermVolat* is the daily stock return volatility during the window [t-5, t-1] for each date t expressed as a percent. Size is the natural logarithm of total assets. *SpecItems* is special items scaled market value of equity. ***, **, and * denote statistical significance at the 0.01, 0.05, and 0.10 levels (two–tail) based on standard errors clustered by firm and date, respectively. Low (High) Info Group sample of 4,130,980 (3,992,388) firm-days in the [–10,+10] day window around 171,116 (165,314) disclosures.

	Low Info Group ($Info < 0.5$)		High Info Group ($Info \geq 0.5$)	
Variable	Mean	Median	Mean	Median
AbsoluteReturn	2.031	1.141	2.090	1.157
Volume	1.107	0.645	1.215	0.655
AbnormalReturn	-0.007	-0.049	0.000	-0.050
AdExp	0.093	0.003	0.084	0.004
CapEx	0.026	0.014	0.026	0.013
CapIntens	0.256	0.165	0.243	0.152
CorpAcq	0.180	0.000	0.191	0.000
Financing	0.084	0.010	0.088	0.012
IdioVol	2.321	1.784	2.328	1.789
Leverage	0.564	0.548	0.585**	0.570**
MTB	1.914	1.451	1.850	1.400
QtrlyReturn	3.479	2.100	3.145**	1.821
R&D	1.575	0.000	1.754	0.000
ShortTermReturn	0.261	0.116	0.301	0.111
ShortTermVolat	0.168	0.039	0.191	0.040
Size	7.082	7.156	7.182*	7.270*
SpecItems	-0.455	0.000	-0.887	0.000

Table 5. Market Reaction

This table presents the market reaction as a function of measures of information. Panel A presents results for annual and quarterly reports on Forms 10-K/Q. Panel B presents results for current reports on Form 8-K. Events are defined as days where a filing is posted to EDGAR. *Event*[0] is first day that the market can react to a filing posted to EDGAR. *Length* is the length of the filing deciled and scaled to range from zero to one. *Info* is the information in the filing. Specifically, for each token we use a firm-specific LLM to measure the token's information according to Eq. 6. We take the geometric mean over the set of each token's information then decile and scale the result to range from zero to one. Daily Controls includes *ShortTermReturn* and *ShortTermVolat*. Quarterly Controls includes *AdExp*, *CapEx*, *CapIntens*, *CorpAcq*, *Financing*, *IdioVol*, *Leverage*, *MTB*, *QtrlyReturn*, *R&D*, *Size*, and *SpecItems*. Firm-Year-Qtr fixed effects subsume all variables calculated at the quarterly level. *t*-statistics appear in parentheses and are based on standard errors clustered by firm and date. ***, **, and * denote statistical significance at the 0.01, 0.05, and 0.10 levels (two–tail), respectively. Panel A (Panel B) is based on a sample of 2,204,716 (5,918,652) firm-days in the [–10,+10] day window around 90,437 (245,993) Form 10-K/Q (Form 8-K) filings.

Panel A. Annual and Quarterly Reports (Forms 10-K/Q)

Independent	Dependent Variable: *AbsoluteReturn*			Dependent Variable: *Volume*		
Variable	(1)	(2)	(3)	(4)	(5)	(6)
Event[0]	0.989***	1.062***	0.511***	0.392***	0.316***	0.279***
	(13.183)	(18.075)	(11.823)	(13.436)	(6.626)	(7.839)
Event[0]×*Length*		-0.156*			0.155**	
		(-1.853)			(2.319)	
Event[0]×*Info*			1.151***			0.271***
			(19.897)			(4.295)
Daily Controls	Y	Y	Y	Y	Y	Y
Quarterly Controls	Y	Y	Y	Y	Y	Y
Date FEs	Y	Y	Y	Y	Y	Y
Firm FEs	Y	Y	Y	Y	Y	Y
N-obs	2,204,716	2,204,716	2,204,716	2,204,716	2,204,716	2,204,716
N-firms	2,460	2,460	2,460	2,460	2,460	2,460

Panel B. Current Reports (Form 8-K)

Independent	Dependent Variable: *AbsoluteReturn*			Dependent Variable: *Volume*		
Variable	(1)	(2)	(3)	(4)	(5)	(6)
Event[0]	1.190***	1.147***	0.465***	0.684***	0.658***	0.216***
	(12.172)	(12.604)	(17.357)	(10.873)	(10.623)	(3.064)
Event[0]×*Length*		0.246***			0.153	
		(4.283)			(1.519)	
Event[0]×*Info*			1.450***			0.939***
			(13.597)			(8.343)
Daily Controls	Y	Y	Y	Y	Y	Y
Date FEs	Y	Y	Y	Y	Y	Y
Firm-Year-Qtr FEs	Y	Y	Y	Y	Y	Y
N-obs	5,918,652	5,918,652	5,918,652	5,918,652	5,918,652	5,918,652
N-firms	2,499	2,499	2,499	2,499	2,499	2,499

Table 6. Cross-sectional Variation in Market Reaction

This table presents tests for cross-sectional variation in the market reaction. Panel A presents results for annual and quarterly reports on Forms 10-K/Q. Panel B presents results for current reports on Form 8-K. Events are defined as days where a filing is posted to EDGAR. *FinNeg* is the count of the number of negative words in the filing scaled by the length of the filing (Loughran and McDonald, 2011) multiplied by negative one, then deciled and scaled to range from negative one to one. *Info* × *FinNeg* is the interaction between information and sentiment. Specifically, for each token we use a firm-specific LLM to measure the token's information according to Eq. 6 and sign the information based on the Fin-Neg dictionary (Loughran and McDonald, 2011). We take the geometric mean over the set of signed information then decile and scale the result to range from negative one to one. All other variables previously defined. Firm-Year-Qtr fixed effects subsume all variables calculated at the quarterly level. *t*-statistics appear in parentheses and are based on standard errors clustered by firm and date. ***, **, and * denote statistical significance at the 0.01, 0.05, and 0.10 levels (two–tail), respectively. Panel A (Panel B) is based on a sample of 2,204,716 (5,918,652) firm-days in the [–10,+10] day window around 90,437 (245,993) Form 10-K/Q (Form 8-K) filings.

Panel A. Annual and Quarterly Reports (Forms 10-K/Q)

Independent Variable	Dependent Variable: *AbnormalReturn*			
	(1)	(2)	(3)	(4)
Event[0]	-0.054**	-0.104***	-1.569***	-1.631***
	(-2.528)	(-3.071)	(-24.655)	(-23.080)
Event[0]×*FinNeg*		0.181**		0.230***
		(2.055)		(2.690)
Event[0]×*Info*×*FinNeg*			3.188***	3.187***
			(13.393)	(13.390)
Daily Controls	Y	Y	Y	Y
Quarterly Controls	Y	Y	Y	Y
Date FEs	Y	Y	Y	Y
Firm FEs	Y	Y	Y	Y
N-obs	2,204,716	2,204,716	2,204,716	2,204,716
N-firms	2,460	2,460	2,460	2,460

Panel B. Current Reports (Form 8-K)

Independent Variable	Dependent Variable: *AbnormalReturn*			
	(1)	(2)	(3)	(4)
Event[0]	0.077***	-0.076**	-2.549***	-2.652***
	(5.706)	(-2.281)	(-49.525)	(-43.561)
Event[0]×*FinNeg*		0.265***		0.179***
		(5.870)		(4.144)
Event[0]×*Info*×*FinNeg*			5.245***	5.244***
			(10.305)	(10.306)
Daily Controls	Y	Y	Y	Y
Date FEs	Y	Y	Y	Y
Firm-Year-Qtr FEs	Y	Y	Y	Y
N-obs	5,918,652	5,918,652	5,918,652	5,918,652
N-firms	2,499	2,499	2,499	2,499

Table 7. Descriptive Statistics for Long-Run Performance

This table presents the distribution of variables used in our tests of long-run performance. *ROA* is income before extraordinary items scaled by total assets at the beginning of the quarter. *BHAR* is the intercept from the daily expected returns model of Fama and French (2015) and Carhart (1997) estimated over the quarter and multiplied by the number of trading days in the quarter, expressed as a percent. All other variables previously defined. Sample of 82,650 firm-quarters.

	Mean	Std	25-th	50-th	75-th
ROA	1.931	5.356	0.031	0.260	1.220
BHAR	0.014	0.202	-0.090	0.007	0.107
AdExp	0.049	0.115	0.000	0.004	0.044
CapEx	0.025	0.032	0.005	0.014	0.031
CapIntens	0.242	0.234	0.064	0.156	0.351
CorpAcq	0.187	0.390	0.000	0.000	0.000
Financing	0.076	0.153	0.001	0.010	0.072
IdioVol	2.213	1.545	1.186	1.772	2.709
Leverage	0.543	0.259	0.352	0.532	0.707
MTB	1.868	1.258	1.089	1.448	2.137
QtrlyReturn	3.005	22.865	-9.403	2.076	13.684
R&D	0.118	0.512	0.000	0.000	0.057
Size	6.958	2.036	5.499	7.014	8.372
SpecItems	-0.005	0.024	-0.002	0.000	0.000

Table 8. Out-Of-Sample Long-Run Fundamentals

This table presents out-of-sample predictions of future firm performance. Panel A (C) presents results from regressing levels of *ROA* on levels of *Info* × *FinNeg* and controls for Form 10-K/Q (Form 8-K) filings. Panel B (D) presents results from regressing first differences of *ROA* on first differences of *Info* × *FinNeg* and controls for Form 10-K/Q (Form 8-K) filings. Controls includes the set of variables in Quarterly Controls defined in Table 5. *t*-statistics appear in parentheses and are based on standard errors clustered by firm and quarter. ***, **, and * denote statistical significance at the 0.01, 0.05, and 0.10 levels (two–tail), respectively. Column (1) in Panels A and B (C and D) based on a sample of 82,650 (81,420) firm-quarter observations. Requirement of future measurements of *ROA* leads to a loss of observations in columns (2) through (11).

Panel A. Annual and Quarterly Reports (Forms 10-K/Q) Levels Specification

	Dependent Variable:										
	ROA_{q}	ROA_{q+1}	ROA_{q+2}	ROA_{q+3}	ROA_{q+4}	ROA_{q+5}	ROA_{q+6}	ROA_{q+7}	ROA_{q+8}	ROA_{q+12}	ROA_{q+16}
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Info×*FinNeg*	0.207***	0.178***	0.173***	0.188***	0.161***	0.133***	0.120***	0.144***	0.116***	0.089***	0.082***
	(15.746)	(13.302)	(11.794)	(12.738)	(10.540)	(9.957)	(9.367)	(11.127)	(8.056)	(5.498)	(5.663)
Controls	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Year FE	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Observations	82,650	82,091	81,753	81,453	81,151	80,995	80,847	80,675	78,871	71,169	63,918
Firms	2,327	2,318	2,315	2,316	2,314	2,312	2,313	2,311	2,265	2,086	1,933
F	1639.0***	1598.0***	1510.0***	1529.8***	1495.0***	1421.9***	1345.8***	1331.9***	1221.9***	908.0***	653.0***
R^2	0.205	0.202	0.194	0.196	0.193	0.186	0.178	0.177	0.168	0.142	0.117
Within R^2	0.187	0.159	0.182	0.189	0.164	0.179	0.169	0.170	0.157	0.140	0.110

Panel B. Annual and Quarterly Reports (Forms 10-K/Q) Differences Specification

	Dependent Variable:										
	ΔROA_{q}	ΔROA_{q+1}	ΔROA_{q+2}	ΔROA_{q+3}	ΔROA_{q+4}	ΔROA_{q+5}	ΔROA_{q+6}	ΔROA_{q+7}	ΔROA_{q+8}	ΔROA_{q+12}	ΔROA_{q+16}
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Δ*Info*×*FinNeg*	0.038**	0.069***	0.055***	0.019*	0.065***	0.045***	0.035***	0.023*	0.055***	0.034***	0.027***
	(2.436)	(4.334)	(4.003)	(1.766)	(4.938)	(3.778)	(2.623)	(1.941)	(5.955)	(4.114)	(3.552)
Controls	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Year FE	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Observations	82,650	82,475	82,396	82,331	82,268	82,203	82,146	82,072	80,303	72,561	65,209
Firms	2,327	2,318	2,315	2,316	2,314	2,312	2,313	2,311	2,265	2,086	1,933
F	2430.0***	175.0***	135.9***	127.7***	119.2***	98.4***	83.7***	96.4***	77.8***	55.0***	21.6***
R^2	0.277	0.269	0.210	0.198	0.185	0.153	0.131	0.150	0.124	0.098	0.043
Within R^2	0.281	0.149	0.164	0.172	0.159	0.143	0.141	0.154	0.127	0.088	0.041

Table 8. Out-Of-Sample Long-Run Fundamentals (Cont'd)

Panel C. Current Reports (Form 8-K) Levels Specification

	Dependent Variable:										
	ROA_{q}	ROA_{q+1}	ROA_{q+2}	ROA_{q+3}	ROA_{q+4}	ROA_{q+5}	ROA_{q+6}	ROA_{q+7}	ROA_{q+8}	ROA_{q+12}	ROA_{q+16}
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Info×*FinNeg*	0.109***	0.117***	0.086***	0.071***	0.063***	0.059***	0.044***	0.039***	0.040***	0.018*	0.016*
	(11.220)	(12.042)	(9.930)	(8.799)	(8.442)	(8.265)	(5.822)	(4.748)	(4.940)	(1.942)	(1.657)
Controls	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Year FE	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Observations	81,420	80,906	80,568	80,267	79,976	79,824	79,675	79,535	79,372	71,552	64,234
Firms	2,287	2,280	2,278	2,277	2,276	2,274	2,267	2,262	2,263	2,087	1,935
F	1563.4***	1564.4***	1464.1***	1432.8***	1411.7***	1357.4***	1296.4***	1255.0***	1193.2***	868.5***	609.7***
R^2	0.200	0.201	0.191	0.188	0.187	0.181	0.175	0.170	0.164	0.136	0.110
Within R^2	0.174	0.135	0.181	0.163	0.157	0.184	0.163	0.166	0.141	0.113	0.104

Panel D. Current Reports (Form 8-K) Differences Specification

	Dependent Variable:										
	ΔROA_{q}	ΔROA_{q+1}	ΔROA_{q+2}	ΔROA_{q+3}	ΔROA_{q+4}	ΔROA_{q+5}	ΔROA_{q+6}	ΔROA_{q+7}	ΔROA_{q+8}	ΔROA_{q+12}	ΔROA_{q+16}
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Δ*Info*×*FinNeg*	0.008***	0.024***	0.012***	0.009***	0.008***	0.013***	0.005*	0.003	0.006***	0.006***	0.005*
	(4.376)	(11.142)	(5.052)	(4.351)	(3.257)	(7.053)	(1.713)	(1.382)	(4.181)	(5.483)	(1.665)
Controls	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Year FE	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Observations	81,420	80,906	80,568	80,267	79,976	79,824	79,675	79,535	79,372	71,552	64,234
Firms	2,287	2,280	2,278	2,277	2,276	2,274	2,267	2,262	2,263	2,087	1,935
F	1582.6***	201.1***	390.3***	533.9***	220.0***	231.9***	373.2***	470.8***	272.0***	281.7***	274.4***
R^2	0.202	0.313	0.593	0.796	0.345	0.364	0.574	0.715	0.427	0.487	0.526
Within R^2	0.206	0.409	0.636	0.842	0.359	0.306	0.467	0.720	0.409	0.383	0.405

Table 9. Out-Of-Sample Long-Run Returns

This table presents out-of-sample predictions of future returns. Panel A (B) presents results from regressing *BHAR* on *Info* × *FinNeg* and controls for Form 10-K/Q (Form 8-K) filings. Controls includes *AdExp*, *CapEx*, *CapIntens*, *CorpAcq*, *Financing*, *Leverage*, *R&D*, and *SpecItems*. Other firm characteristics such as size and value, as well as quarterly time-varying risk exposure, are implicitly controlled for via the estimation of *BHAR* (see Table 7 for definition). *t*-statistics appear in parentheses and are based on standard errors clustered by firm and quarter. ***, **, and * denote statistical significance at the 0.01, 0.05, and 0.10 levels (two–tail), respectively. Column (1) in Panels A and B (C and D) based on a sample of 82,650 (83,768) firm-quarter observations. Requirement of future measurements of *BHAR* leads to a loss of observations in columns (2) through (11).

Panel A. Annual and Quarterly Reports (Forms 10-K/Q)

	Dependent Variable:										
	$BHAR_{q}$	$BHAR_{q+1}$	$BHAR_{q+2}$	$BHAR_{q+3}$	$BHAR_{q+4}$	$BHAR_{q+5}$	$BHAR_{q+6}$	$BHAR_{q+7}$	$BHAR_{q+8}$	$BHAR_{q+12}$	$BHAR_{q+16}$
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Info×*FinNeg*	0.280***	0.352***	0.291***	0.212***	0.232***	0.322***	0.219***	0.189***	0.187***	0.170***	0.148***
	(3.913)	(5.781)	(9.799)	(7.327)	(4.540)	(7.278)	(9.115)	(6.709)	(5.267)	(4.173)	(3.561)
Controls	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Year FE	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Observations	82,650	82,454	82,352	82,268	82,191	82,106	82,029	81,937	80,159	72,403	65,059
Firms	2,327	2,325	2,325	2,325	2,323	2,321	2,322	2,320	2,274	2,096	1,943
F	23.1***	34.5***	25.6***	21.7***	26.0***	42.0***	29.9***	24.0***	21.1***	17.0***	12.9***
R^2	0.025	0.038	0.028	0.024	0.028	0.046	0.033	0.026	0.024	0.021	0.018
Within R^2	0.017	0.023	0.019	0.016	0.018	0.037	0.026	0.020	0.016	0.010	0.007

Panel B. Current Reports (Form 8-K)

	Dependent Variable:										
	$BHAR_{q}$	$BHAR_{q+1}$	$BHAR_{q+2}$	$BHAR_{q+3}$	$BHAR_{q+4}$	$BHAR_{q+5}$	$BHAR_{q+6}$	$BHAR_{q+7}$	$BHAR_{q+8}$	$BHAR_{q+12}$	$BHAR_{q+16}$
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Info×*FinNeg*	0.215***	0.230***	0.280***	0.182***	0.173***	0.169***	0.114***	0.137***	0.073***	0.112***	0.087***
	(3.904)	(4.498)	(6.713)	(8.702)	(4.434)	(4.643)	(4.579)	(5.648)	(2.818)	(3.821)	(3.404)
Controls	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Year FE	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Observations	83,768	83,602	83,494	83,399	83,311	83,240	83,162	83,092	83,002	74,976	67,415
Firms	2,296	2,294	2,293	2,293	2,291	2,290	2,287	2,287	2,284	2,104	1,951
F	19.2***	26.2***	26.8***	25.9***	22.2***	28.4***	25.7***	24.5***	18.6***	14.8***	10.8***
R^2	0.021	0.028	0.029	0.028	0.024	0.031	0.028	0.026	0.020	0.018	0.014
Within R^2	0.018	0.020	0.020	0.020	0.018	0.024	0.023	0.021	0.017	0.010	0.006