Measuring Information Quality by Topic Attention Divergence: Evidence from Earnings Calls*

Zicheng Xiao Tengfei Zhang

First written: February 2024

Updated: September 2024

Abstract

Leveraging computational linguistics and 20 million turns of dialogues from earnings conference calls from 2006-2022, we introduce a novel measure that quantifies the disparity in narrative focus between managers' disclosures and analysts' questions during these calls, denoted as Topic Attention Divergence (TAD). A higher level of TAD indicates a higher level of firm-investor asymmetry and lower information quality. Our results confirm that higher TAD in earnings conference calls inversely (positively) predict firms' future stock liquidity (cost of equity capital). Further, the predictive power of TAD is more pronounced in firms with higher information processing costs characterized by smaller firm size, lower levels of analyst coverage, and lower institutional ownership. A long-short portfolio sorted by TAD earns an annual 8.18% risk-adjusted return and cannot be explained by existing factor models.

Keywords: *Information Quality, Earnings Conference Call, Cost of Capital, Artificial Intelligence, Machine Learning*

JEL classifications : C82, D8, G1

---

*Zicheng Xiao: Sam M. Walton College of Business, University of Arkansas. Email: ZXiao@walton.uark.edu.
Tengfei Zhang: Rutgers School of Business-Camden, Rutgers University. Email: tengfei.zhang@rutgers.edu.

1. Introduction

Corporate disclosures, including voluntary communication during earnings conference calls, are among the most important sources of information for investors (Healy and Palepu, 2001; Matsumoto et al., 2011; Brown et al., 2019). Managers can also choose to meet privately with investors. These meetings are highly valued by investors (Brown et al., 2019). While it is generally not possible to examine the effects of these sorts of voluntary disclosures due to data limitations, recent advances in computational linguistics can be used to examine managers' transparency during conference calls that proxy the information quality of a firm's information characteristics. In this study, we exploit the earnings conference call as a laboratory to examine whether a firm discusses information that investors want by comparing the similarity of the topics discussed in the presentation section, which is driven by the firm's managers, and the question and answer (Q&A) section, which is driven by analysts.

I propose a novel measure, Topic Attention Divergence (TAD), which directly measures information quality by quantifying the firm's propensity to disclose information that investors prioritize in earnings calls. TAD provides a novel perspective to understand the content of firm-investor interactions in corporate disclosure, thereby underscoring the underlying information asymmetry. This study is inspired by Lambert et al. (2007) and Ng (2011), who posit that higher information quality (i.e., more precise signals) can reduce market risk and, consequently, the cost of capital in the traditional asset pricing model.[1]

To illustrate our approach, consider Apple Inc.'s 2006 Q3 conference call as an example. In the presentation section of the call, CFO Peter Oppenheimer and CEO Tim Cook spent 10.9% and 0% of their time discussing cash flow and industry trends, respectively. In contrast, analysts asked no questions pertaining to cash flow, but directed 5% of their questions toward industry

[1] Following Ng (2011), I define information quality as a characteristic of a firm's information that affects the degree of uncertainty over the firm's value and/or adverse selection in trading the firm's stock (Healy and Palepu, 2001; Verrecchia, 2001; Easley et al., 2002; Easley and O'Hara, 2004).

trends. Furthermore, about 20% of the topics that analysts inquired about were not covered in the presentation section. This disparity in topic attention during the conference call is the essence of what we term as TAD. We employ TAD as a proxy for the inverse quality of information disclosed by firms in their earnings conference calls and examine the links between TAD and capital markets.[2]

This variance of disparity in topic attention (topic percentage) between presentation sessions and Q&A sessions during the conference call is termed Topic Attention Divergence (TAD). We employ TAD as a proxy for the inverse quality of information disclosed by firms in their earnings conference calls and we examine capital markets implications of TAD. Specifically, this study addresses three major questions: 1) Does higher TAD (lower information quality) indicate a higher firm's illiquidity (liquidity risk)? 2) Is this information quality priced in asset prices? and 3) How does information processing costs impact the effect of TAD on asset prices?

To examine the relationship between topic attention and future returns, quantifying narrative attention posed a challenge. By leveraging advancements in Natural Language Processing (NLP) techniques, specifically Latent Dirichlet Allocation (LDA) and GPT-text embedding, we have categorized over 20 million dialogues from earnings calls spanning 17 years and encompassing 5,092 firms into distinct thematic topics. This categorization includes both presentation segments and Q&A sessions. The analysis revealed that the top five topics during presentations were "Guidance," "Growth," "Investment," "Cash Flow," and "Urgency & Emphasis" with percentages ranging from 19.43%-7.32%. Conversely, in Q&A segments, "Guidance" and "Urgent & Emphasis" remained prevalent, alongside "Investment," "Growth,"

[2] In the Apple F3Q06 Earnings Conference Call, specifically during the presentation segment, CFO Peter Oppenheimer dedicated approximately 18 minutes to outlining Apple Inc.'s key financial indicators. He stated, "Operating margin for the quarter was 13% and net income was $472 million, or $0.54 per diluted share on a GAAP basis. Excluding the impacts of $37 million in non-cash, stock-based compensation expense, operating margin was 13.8%, net income was $499 million, and diluted earnings per share were $0.57." Utilizing LDA generated topic dictionaries, this statement can be effectively categorized into relevant topics. The analysis reveals that out of 46 words, five bigrams pertain to "revenue and earnings." These include three distinct bigrams: "operating margin," "net income," and "earnings per share." This enables the generation of topic-specific embedded vectors for each identified topic within the sentence.

and “Geopolitical Risk” indicating a shift toward more strategic concerns. This disparity highlights the divergence between managers' disclosures and analysts' demands for information. Investors predominantly sought insights into the company’s operations, strategic decisions, and publicly available information highlighting their pursuit of deeper operational understanding and response to macroeconomic conditions or potential risks. Their active engagement, evident through probing questions and strategic suggestions, underscores the dynamic nature of investor-manager interactions during earnings calls.

With respect to the first goal, we seek to investigate the implications of TAD and examine its role in signaling low information quality or high information risk. The literature links high information quality to improved stock liquidity.[3] Thus, if TAD reflects lower information quality, a negative correlation between TAD and stock liquidity should be observed.

When stock liquidity decreases, the impact on firms varies. Those with lower information quality face more pronounced market withdrawals. This is primarily due to a decrease in the willingness of investors to engage in stocks that carry higher uncertainty of future cash flows and risk of adverse selection. Further, adverse selection makes market makers less inclined to offer liquidity for such stocks that could further reduce the demand for these stocks and lead to poorer performance. Conversely, an increase in stock liquidity attracts more investors and market makers enhancing the demand and liquidity for stocks marked by higher uncertainty and risk. This situation can affect the bid-ask spread reflecting changes in liquidity. Therefore, the returns of stocks with lower information quality are more affected by fluctuations in stock liquidity illustrating how information quality is linked to liquidity risk.

[3] Bushman and Indjejikian (1995) argue that high quality disclosures influence the actions of uninformed liquidity traders, thereby attracting more liquidity. Leuz and Verrecchia (2000), as well as Shroff et al. (2013), demonstrate that enhanced disclosures, as evidenced by the voluntary IFRS adoption in Germany and the 2005 Securities Offering Reform, lead to increased liquidity. Ng (2011) documents a positive relationship between information quality and stock liquidity. Kim (2014) observes that firms' deliberate efforts to boost transparency, such as opting to be listed on Euronext's 'special segments,' result in liquidity surges. Furthermore, Balakrishnan et al. (2014) provide evidence that voluntary disclosure directly contributes to liquidity enhancement. Bushee et al. (2018) use Amihud’s (2002) illiquidity as the measure of information asymmetry around earnings conference calls.

I find that there is a significantly positive relation between TAD and liquidity risk (*illiquidity*). Liquidity risk is measured by bid-ask spread (Corwin and Schultz, 2012) and the Amihud (2002) illiquidity measure. In addition, an alternative information risk (i.e., the inverse of information quality) measure, such as absolute forecast errors in earnings forecasts and the dispersion of analysts' forecasts, are also found to be positively correlated with illiquidity. These observed linkages, indicating that lower information quality is related to greater liquidity risk, align with the evidence reported by Ng (2011). These outcomes imply that an increase in TAD during earnings announcements is associated with lower stock liquidity validating the conjecture that TAD inversely reflects information quality.

In the second stage of our analysis, we examine whether lower information quality (higher information risk), identified by TAD, is priced. The traditional asset pricing literature often posits that investors are compensated solely for holding market risks assuming uniform information access across all investors. However, this assumption becomes complex when information asymmetry exists among investors. To navigate this complexity, we utilize both portfolio and regression analyses to study the impact of TAD on stock returns. Our panel regression results reveal that firms with high TAD typically exhibit higher future stock returns. Moreover, the impact of TAD on pricing is more significant for small stocks or firms with high information processing costs. A trading strategy that involves taking a long position in a high-TAD value-weighted quintile portfolio and a short position in a low-TAD quintile value-weighted portfolio yields an 8.18% risk-adjusted annual return, which increases to 12.68% when this long-short strategy is applied to the smallest size quintile. These results are robust using the Fama and MacBeth (1973) method that also controls for other well-known pricing factors. Furthermore, the pricing relevance of TAD remains robust even after excluding inactive or penny stocks, using different topic dictionaries, controlling for alternative variables of information risks, and employing several alternative measures of TAD.

Finally, we provide further evidence to enhance the understanding of the pricing impact of TAD. Considering that the post-earnings announcement is a well-documented return

anomaly, and the potential relationship of this anomaly with TAD, we have verified that the pricing of TAD is not driven by post-earnings announcement drift. This distinction ensures that TAD's pricing impact is distinct from well-known market anomalies. Additionally, we find that firms with high TAD are often undervalued resulting in them earning significantly positive alpha during the post-call periods. Furthermore, our research reveals that companies facing higher information processing costs, characterized by smaller market capitalization, lower levels of institutional ownership, and reduced analysts' coverage, exhibit a more significant impact from TAD.

The contributions of this paper are multifaceted. Our study enhances literature by utilizing advanced computational linguistics tools, such as Latent Dirichlet Allocation (LDA) and Large Language Models (LLM), to mitigate the information advantages of corporate insiders over outsiders. This is unlike traditional quantity-based measures (Easley et al., 1996; Easley et al., 2002; Choi et al., 2014) or price-based measures (Roll, 1988; Moorck et al., 2000; Durnev et al., 2004; Ferreira et al, 2007; Dang et al., 2015; Yang et al., 2020 - hereafter YZZ) that indirectly measure information quality of a firm. Our research fills this gap using machine learning techniques (e.g., LDA, FinBERT, and GPT text embedding) and proposes a direct measure regarding information quality by comparing the similarity of the topics discussed in the presentation section, which is driven by the firm's managers, and the question and answer (Q&A) section that is driven by analysts. One of our measure's key contributions is the real-time preservation of earnings call context with NLP techniques. Building on this, our findings reveal that higher TAD is associated with greater liquidity risk consistent with the findings of the previous literature that information quality is negatively correlated with liquidity risk (Ng, 2011; Lambert et al., 2010) supporting our conjecture that TAD represents information quality.

In addition, this paper contributes to the broader objective of enhancing our understanding of the benefits of corporate disclosure, including voluntary communications, that are among the most crucial sources of information for investors (Healy and Palepu, 2001; Brown et al., 2019). Studies by Frankel et al. (1999), Matsumoto et al. (2011), Hollander et al.

(2010), and Mayew and Venkatachalam (2012) have demonstrated that the Q&A segments of these calls yield significant insights. This paper finds that the divergent topic attention of market participants is important and incorporates information predictive of future asset prices, thereby complementing previous work in the corporate disclosure literature.

Moreover, this work contributes to the broader objective of enhancing our understanding of the mechanisms underlining the relationship between information quality and the cost of capital (Botosan, 1997; Easley et al., 2002; 2004; Ecker, 2013; Francis et al. 2005; Core et al. 2008; Akins et al. 2012). Our investigation supports the previous finding that poor information quality (higher TAD) is associated with equity premiums (Easley et al., 2004; Brevik and D'Addona, 2010; Epstein and Schneider, 2008; AI H., 2010, Hughes et al., 2017). Lower information quality predicts higher costs of capital due to greater uncertainty and adverse selection issues. The newly constructed proxy TAD, as inverse information quality, aligns with and expands upon the existing knowledge regarding the link between information quality and market outcomes documented in the previous literature.

Finally, this paper further contributes to the literature on information processing costs by revealing that firms with higher information processing costs, characterized by low institutional ownership, low analyst coverage, and smaller firm size, exhibit a more pronounced relationship between TAD and the equity premium. This finding confirms the role of institutional investors and financial analysts in enhancing information transparency and reducing information processing costs as suggested by Healy and Palepu (2001), Beyer et al. (2010), and Frankel et al. (2006). This study demonstrates that higher institutional ownership can mitigate the proportion of informed trades and enhance liquidity by improving the quality of information or reducing information asymmetry. This insight enriches our understanding as to how various factors influence the effects of TAD on market outcomes underscoring the importance of institutional mechanisms in facilitating more efficient information dissemination and processing in the financial markets.

This paper is organized as follows. Section 2 describes the methodology. Section 3 details the dataset used in this research. Section 4 presents the empirical results, while Section 5 provides the robustness check. Section 6 reports our conclusions.

2. Methodology

In this paper, we create two distinct methods to construct similarities between the presentation and discussion in the earnings call at the topic level. Our model for the structure of topics in corporate voluntary disclosure follows the Latent Dirichlet Allocation (LDA) topic modeling approach of Blei et al. (2003). We provide a brief summary of the model and its estimation and refer interested readers to the original paper for additional detail. These two similarities measures are computed at the sentence level by comparing the scaled topic frequency between managers' disclosures and discussions. To obtain this scaled topic frequency from each segment of the earnings call, we take the average value of all of the sentence's topic frequency in each segment (Presentation and Q&A) separately. Section 2.1 measures the attention of topics. Section 2.2 presents the model selection. Section 2.3 quantifies the topic attention divergence, while Section 2.4 quantifies the topic sentiment divergence. Figure IA1 illustrates the flowchart of the construction of TAD.

2.1 Measuring Attention of Topics: LDA

Given the complex nature of unprocessed transcripts from human dialogues, Latent Dirichlet Allocation (LDA) seeks to distill a comprehensible thematic structure from the intricate details of historical earnings discussions rendering them accessible for straightforward understanding and examination. This summarization is achieved through the application of defined parametric constraints (i.e., a multinomial distribution of term frequencies guided by Dirichlet priors) and the establishment of a structured framework to predict term frequencies. In particular, LDA posits that the term counts for any given document $\boldsymbol{t}$, represented by the V-dimensional vector $\mathbf{w}_t$, following a multinomial distribution.

$$w_t \sim Mult(\Phi' \theta_t, N_t) \quad (1)$$

where N_t is the total number of terms in document t and governs the scale of the multinomial distribution. In other words, expected term counts are summarized by a comparatively low dimension set of parameters, $\boldsymbol{\theta_t}$ and $\boldsymbol{\Phi} = [\boldsymbol{\Phi_1}, \ldots \boldsymbol{\Phi_K}]$. The $\boldsymbol{k^{th}}$ "topic" in the text is defined by the V-dimensional parameter vector $\boldsymbol{\phi_k}$, where $\boldsymbol{\phi_{k,v}} \geq 0$ for all v and $\sum_v \phi_{k,v} = 1$. That is, a topic issue is a probability distribution over terms. The set of terms that take especially high probability in $\boldsymbol{\phi_k}$ convey the thematic content of the topic. The model's dimension reduction is achieved by setting the total number of topics $\boldsymbol{K}$ much smaller than the size of the vocabulary.

While the topics, $\boldsymbol{\phi_k}$, outline the overarching themes within the entire corpus, LDA conceptualizes each individual document as a blend of these topics. The unique parameter vector for each document, $\theta_t = (\theta_{t,1}, \ldots, \theta_{t,K})$, functions as a probability vector ensuring that $\boldsymbol{\theta_{t,k}}$ is non-negative for all $\boldsymbol{k}$ and the sum over all $\boldsymbol{k}$ of $\boldsymbol{\theta_{t,k}}$ equals one. Essentially, $\boldsymbol{\theta_t}$ delineates the distribution of a document's focus among the various topics. Through LDA, a factor structure is incorporated where the topics ($\boldsymbol{\phi_k}$) act as universal factors and $\boldsymbol{\theta_t}$ delineates the document-specific affinity toward these factors.

From a theoretical standpoint, one might lean toward employing a maximum likelihood estimation to deduce $\boldsymbol{\phi}$ and $\boldsymbol{\theta}$ as per the given equation. Nevertheless, the computational intensity required for most textual analyses renders this method often unfeasible. To circumvent this limitation, we adopt Bayesian techniques, in line with Bybee et al., (2023), specifically employing the collapsed Gibbs sampling approach as introduced by Xiao and Stibor (2010).

2.2 Topic Model Selection

In this study, our primary analysis centers around the examination of two-word ("bi-grams") and three-word ("tri-grams") phrases commonly found in earnings call transcripts. These n-grams, or sequences of multiple words, are organized into ten main topics that either

highlight essential aspects of a company's fundamental operations or attract considerable attention from those involved in the markets. Latent Dirichlet Allocation (LDA) serves as the unsupervised machine learning method employed here requiring only two inputs from the researcher: the dataset of term counts and the number of topics denoted by *K*. To determine the optimal value of *K*, we adopt a data-centric strategy estimating various models with *K* values ranging from two to 100 in increments of five. The model that demonstrates the highest Bayes factor is selected for further analysis. For added robustness, we also utilize ten-fold cross-validation in the model selection. Specifically, we divide the sample of earnings conference calls into 10 equally sized groups, or "folds," of documents. Each *K*-topic model is then re-estimated 10 times on the dataset, each time omitting one of the 10 validation samples, and its performance is assessed on the omitted sample. The overall effectiveness of each model is determined by averaging the log likelihood values across the 10 validation samples.

Insert Figure 1

Figure 1 displays the goodness-fit across various proposed models. Both the Bayes factor and the cross-validation methods suggest that a topic range of 35-50 nearly maximizes our criteria for model selection. Further manual examination of estimates for different values of K supports the choice of K = 37 as reasonable. Models with a reduced number of topics, like 10, tend to amalgamate disparate themes into single topics leading to a lack of clarity. Conversely, opting for a model with an excessive number of topics, such as 60, yields a similar level of clarity as the 37-topic model, but includes numerous highly specific topics that reflect singular occurrences. Notably, models around the K = 37-mark exhibit very alike topic structures.

Insert Figure 2

Figure 2 displays a relatively low correlation across topics generated by the model in the earnings conference calls, providing further visual evidence that the 37-topic model achieves a balance between interpretability and parsimony. For additional illustrative support, this demonstrates that the 37-topic model effectively balances interpretability and parsimony.

Internet Appendix C provides the word lists of the 37-topic model. Figure IA2 exhibits examples of word clouds of topic attentions separated by presentation sections and Q&A sections in earnings conference calls.

Additionally, to avoid the looking-ahead bias, we turn to an Online LDA (oLDA) estimator from Hoffman et al. (2010) that constructs topic model estimates in an entirely backward-looking fashion. oLDA has the advantage of processing documents in sequential batches. Therefore, time t topic attention estimates are based only on data from periods one through t. As a result, time t topic attention is a valid input for forming out-of-sample forecasts of t+1 outcomes without the look-ahead bias. Please refer to Hoffman et al., (2010) for additional details.

2.3 Quantifying Topic Attention

Topic attention, defined as a scaled topic frequency, is estimated at the paragraph level from voluntary corporate disclosures. We aggregate and compute the average the topic attention across all paragraphs for each manager's disclosure and discussion of earnings conference call separately. We quantify the earnings call text corpus to topics based on two approaches.

The first is "bag-of-words" form, represented numerically as the document-term matrix, **w**. This is a $T \times V$ matrix where indices correspond to the list of T different documents in the corpus, and column indices correspond to the vocabulary of V unique terms in the corpus (Manning et al., 2008; Hassan et al., 2019). The second approach, "Doc2Vec," involves text embedding and topic searching using the GPT text embedding model, specifically the "text-embedding-ada-002" or "Ada v2" version with cosine similarity to determine semantic similarities.[4] This involves comparing the embeddings of the model generated dictionaries with the text embeddings from the context of earnings conference calls.

2.3.1 Primary Approach: Bag-of-Words

To describe the measure construction procedure formally, we define the following set of terms. For each firm *i* and quarter *t*, let the list of bigrams in the transcripts note as $\{B_{i,t}\}$, and *K* the total number of subtopics. For the "bag-of-words" approach, we define the function to count word frequency as $\mathcal{F}: B_{i,t,k} \rightarrow \mathbb{R}_k$, where $\mathbb{R}_k$ denotes the same set of synonyms for the k^{th} topic, $k \in [1,K]$. Objectively, we seek to quantify the attention of the market participants by K topics for each firm *i*, quarter *t*. Specifically, we construct this measure of overall topic frequency by first defining a training library of model generated topics, denoted as $\{\mathbb{R}_k\}$, $k \in [1, K]$.[5] Each training library is the set of all adjacent two-word combinations ("bigrams") contained in the respective k^{th} topic. We then similarly decompose each conference call transcript of firm *i* and quarter *t* into a list of bigrams, $b = \{1, \cdots, B_{i,t}\}$.[6] We then count the number of occurrences of the bigrams indicating discussion of a given kth topic in a transcript's segment, $B_{i,t,s}$ and divide by the total number of bigrams in the segment B_{its}, , s$\in [p, q]$ where p stands for presentation, and q stands for Q&A:

$$A_{itsk} = \frac{\sum_{b_k}^{B_{itk}} \mathbb{I}[b_k \in \mathbb{R}_k]}{B_{its}} \times \frac{\mathcal{F}_{b_k,\mathbb{R}_k}}{B_{\mathbb{R}_k}} \qquad (2\text{-}1)$$

In this context, the indicator function is represented as $\mathbb{I}[\bullet]$, while $\mathbb{R}_k$ signifies a set of synonyms associated with the k^{th} topic. The first couple of terms in the numerator serve to quantify the number of bigrams linked to the discussion topic *k*. In our core model, each bigram is also assigned a weight that mirrors the intensity of its association with a particular topic. Here, $\mathcal{F}_{b_k,\mathbb{R}_k}$ stands for the frequency of the bigram in the k^{th} topic and $B_{\mathbb{R}_k}$ denotes the bigrams in the transcript's total count of bigrams in the topic *k* library. The first summation, $\sum_{b_k}^{B_{itsk}}(\mathbb{I}[b_k \in \mathbb{R}_k])$, counts the total number of bigrams in topic *k* in the segment, $B_{i,t,s}$. The

[5] Internet Appendix B. Dataset Construction Details.

first summation is scaled by the total number of bigrams in each sentence. The second summation sums all fractions of the topic k bigrams relative to the total bigrams in each segment within an earnings call transcript. Consequently, we obtain the topic attention of each segment as shown in Equation (2-1) by using the "Bag-of-Words" approach.

This specification adheres closely to the most canonical weighting scheme used in the automated text classification literature, where the two terms $\mathbb{I}[b_k \in \mathbb{R}_k] \times \mathcal{F}_{b_k,\mathbb{R}_k}$ are commonly referred to as the bigram's inverse document frequency interacted with its term frequency (Spark Jones, 1972; Salton and McGill, 1983; Salton and Buckley, 1988, Hassan et al., 2019). When there are more than two training libraries, the former generalizes to the more familiar form: log (# of training libraries/# of training libraries in which the bigram occurs). Essentially, this method (2-1) is a straight-forward implementation of a standard text-classification algorithm augmented by conditioning upon the proximity to a synonym for 37 different topics in earnings calls and normalizes for transcript length. In robustness checks reported below, we experiment with a number of plausible variations of (2-1). Across all these variations, we generally find this conventional approach yields the most consistent results.

2.3.2 Construct Topic Attention Divergence (TAD)

To streamline the process of the earnings call transcript analysis, we partition the transcript into two sections: Presentation, which exclusively is management's synopsis of the firm's quarter, and the Questions and Answers (Q&A) section of the conference call where management answers analysts' questions. This segregation is contributed for each firm *i* and quarter *t* resulting in a one-by-K matrix in which K is the number of topics. To remove the impact of topic selecting words bias, we standardize the topic frequency of the presentation and Q&A section and denote as $\widetilde{A_{Q_{it}}}, \widetilde{A_{P_{it}}}$.

To evaluate the TAD, we employ two distinct methods. The first method calculates the attention divergence as the Frobenius norm (Euclidean norm) of the discrepancy between two

matrices $A_{P_{it}}$and $A_{Q_{it}}$. To remove the impact of topic selecting words bias, we standardize the attention divergence sentiment of the presentation and Q&A section and denote as $\widetilde{A_{Q_{it}}}, \widetilde{A_{P_{it}}}$.

The method assesses topic attention divergence by utilizing the cosine similarity score. In this case, divergence is computed as:

$$TAD_{it} = 1 - \frac{\langle \widetilde{A_{P_{it}}}, \widetilde{A_{Q_{it}}} \rangle}{\left\| \widetilde{A_{P_{it}}} \right\| \times \left\| \widetilde{A_{Q_{it}}} \right\|} \tag{2-3}$$

Here, the numerator $\langle \widetilde{A_{P_{it}}}, \widetilde{A_{Q_{it}}} \rangle$ stands for the dot product of matrices $A_{P_{it}}$ and $A_{Q_{it}}$, and the denominator $\left\| A_{P_{it}} \right\|$ and $\left\| A_{Q_{it}} \right\|$ represents the Frobenius norms of the corresponding matrices. The subtraction of this fraction from one yields the divergence of attention measure.

2.4 Measuring Sentiment of Market Participants

Following Tetlock's (2007) work, there has been a significant increase in research using various textual data within the fields of Finance and Accounting. In this paper, we also apply two methods to measure the sentiment (or tone) of market participants: 1) Bag-of-words and 2) FinBert. The leading method for assessing sentiment in this research area is the "bag-of-words" approach. This technique involves counting the occurrences of words using specific dictionaries catered toward Finance and Accounting, such as those developed by Loughran and McDonald (2011), commonly referred to as LM dictionaries.

2.4.1 Measuring Sentiment of Topics

In this study, we primarily employ the sentiment of FinBERT and use the Loughran-McDonald Dictionary approach as a robustness check. FinBERT, developed by Huang et al. (2023), is a refined adaptation of the substantial language model BERT for sentiment measurement at the sentence level.[7] BERT, originally developed by Google (Devlin et al.,

[7] FinBERT is a pre-trained NLP model to analyze the sentiment of financial text. It is built by further training the BERT language model in the finance domain using a large financial corpus and thereby fine-tuning it for financial

2018), enhances classification tasks by incorporating the context of surrounding words and sentences, both preceding and succeeding the focus word.

This method facilitates a comprehensive understanding of the context, given its ability to incorporate fully the words appearing before and after the focal word. Huang et al. (2023) present a comparative accuracy analysis of several classification models and FinBERT outperforms BERT and others methods used in previous studies. A substantial body of the literature provides evidence supporting FinBERT's superior performance relative to alternative sentiment classification approaches.

To describe our measurement procedure better, we denote the sentiment function as $\mathcal{S}: P(B_{i,t,n,k}) \rightarrow \{-1, 0, 1\}$. Given a sentence c, let p = P(c) represent the output of the FinBERT model where $P: B_{i,t,n,k} \rightarrow \{p_1, p_2, p_3\}$, such that p_i corresponds to the probability that the sentiment of sentence c falls into category i. We define a function $\mathcal{S}: p_i \rightarrow \{-1,0,1\}$ as follows:

$$\mathcal{S}: = \begin{cases} -1, & \text{if } p_i = max(p) \wedge i = 1, \\ 0, & \text{if } p_i = max(p) \wedge i = 2, \\ 1, & \text{if } p_i = max(p) \wedge i = 3 \end{cases} \tag{2-5}$$

For topic classification, we follow Hassan et al. (2019) and use the Bag-of-Words (BoW) approach counting the frequency of words for each paragraph, which represents the attention that managers and analysts place on various topics during the earnings conference call.

$$ASentiment_{itk} = \frac{1}{N_{it}} \bullet \sum_{n=1}^{N_{i,t}} \frac{\sum_{b_k}^{B_{itnk}} (\mathbb{I}[b_k \in \mathbb{R}_k] \times \mathcal{S}(c_{i,t,n})) \times \frac{\mathcal{F}_{b_k,\mathbb{R}_k}}{B_{\mathbb{R}_k}}}{B_{itn}} \tag{2-6}$$

In this context, the indicator function is represented as $\mathbb{I}[\bullet]$, while $\mathbb{R}_k$ signifies a consistent set of synonyms associated with each topic k. The first couple of terms in the numerator serve to quantify the number of bigrams linked to the discussion topic k. In our core model, each bigram

sentiment classification. Financial PhraseBank by Malo et al. (2014) is used for fine-tuning. For more details, please see LINK.

is also assigned a weight that mirrors the intensity of its association with a particular topic. Here, $\mathcal{F}_{b_k,\mathbb{R}_k}$ stands for the frequency of the bigram in topic k, and $B_{\mathbb{R}_k}$ denotes the bigrams in the transcript total count of bigrams in the topic k library. The first summation, $\sum_{b_k}^{B_{itnk}}(\mathbb{1}[b_k \in \mathbb{R}_k])$, counts the total number of bigrams in topic k in the sentence, $B_{i,t,n}$. The first summation is scaled by the total number of bigrams in each sentence. The second summation sums all fractions of the topic k bigrams relative to the total bigrams in each sentence within an earnings call transcript. Consequently, the comprehensive measure of the portion of the conversation devoted to sentiment related to each topic k is the weighted total of the bigrams associated with the corresponding topic used in combination with synonyms for different topics.

2.4.2 Construct Topic Sentiment Divergence (TSD)

In contrast to the method for constructing TAD, topic sentiment divergence (*TSD*) is measured by evaluating the sentiment difference across various managers and analysts by separating the Q&A section of earnings calls into two parts: analysts' and managers' transcripts. Each part is analyzed for each firm i and quarter t resulting in a one-by-K matrix, where K indicates the number of topics. The average sentiment by topics of analysts (managers) in the Q&A section is denoted as $\widetilde{S_{A_{it}}}(\widetilde{S_{M_{it}}})$.

To determine the sentiment discrepancy by topic between managers and analysts, two distinct methods are employed. The first method calculates the attention-weighted topic sentiment divergence as one minus the cosine similarity between the sentiment vectors of the two groups: analysts $(S_{A_{it}})$and managers $(S_{M_{it}})$. This topic sentiment divergence can be mathematically represented as:

$$TSD_{it} = 1 - \frac{\langle \widetilde{S_{A_{it}}}, \widetilde{S_{M_{it}}} \rangle}{\left\| \widetilde{S_{A_{it}}} \right\| \times \left\| \widetilde{S_{M_{it}}} \right\|} \tag{2-6}$$

3. Data

In this section, we describe the process of generating the dispersion of text from the raw text data. We begin with the data description, followed by the transformer-based word embedding model to generate the dictionary, the methodology to construct the SUE measure, and the calculation of prior dispersions of analyst and indirect dispersion proxies, such as trading volume and bid-ask spread.

3.1 Datasets

This research employs a meticulously crafted dataset derived from quarterly earnings call transcripts from January 2006-December 2022. These transcripts were manually collected from Seeking Alpha, a well-respected platform known for its insights into the finance industry and market trends. A detailed account of the construction, merging, filtration, and text processing procedures implemented during data collection is available in Internet Appendix B. Upon completing the dataset, the divergence of opinion was analyzed yielding 124,606 observations involving 5,429 unique permnos. After excluding firms with only four quarters or fewer, the dataset was refined to 110,882 rows with 5,092 unique permnos spanning from the first quarter of 2006 to the fourth quarter of 2022. Internet Appendix C provides the details of the data construction.

For this empirical analysis, data were assimilated from various sources. Analyst earnings forecasts, focusing on Earnings Per Share (*EPS*), were extracted from the I/B/E/S unadjusted detailed history file. The CRSP daily and monthly files were utilized to collect information on stock prices, bid-ask spreads, returns, trading volume, and the number of outstanding shares. Accounting data were sourced from Compustat. To ensure consistency, all continuous variables were winsorized at the 1% and 99% levels. To merge the Transcripts, CRSP, and IBES datasets seamlessly, the identifier crosswalk provided with the Transcripts dataset was employed, along with the CRSP to IBES linking table constructed using the script

developed by Qingyi (Freda) Song Drechsler. [8] Classic earnings surprises were computed according to the method outlined by Livnat and Mendenhall (2006).

3.2. Controlling for Other Variables Can Explain Information Quality

Models that focus on disclosure usually define information quality as the accuracy of a signal regarding a firm's value with signals that have higher accuracy (i.e., lower variance) being considered to be of higher quality (Verrecchia, 2001). Thus, to ensure that empirical proxies accurately represent the conceptual understanding of information quality, we have selected metrics that gauge the fidelity of the earnings signals. In particular, we utilize Absolute Forecast Errors (*AFE*) and Analyst Forecast Dispersion (*DISP*) as proxies for assessing information quality (see Appendix A for a detailed description of these proxies).

4. Empirical Results

4.1. Descriptive Statistics About Textual Topics Analysis

Table 1 presents the descriptive statistics of the scaled topic attention frequencies in our entire sample, over 200,000 earnings calls. On average, managers allocate 19.43%, 9.24%, 8.97%, 8.71%, and 7.32% of their time to the top five topics during the presentations, namely guidance, growth, investment, cash flow, and urgency, respectively. In contrast, during the Q&A sessions, analysts focus predominantly on guidance, urgency, investment, growth, and geo-political risk, dedicating 21.53%, 9.19%, 7.5%, 6.94%, and an unspecified percentage of their time to these topics, respectively. The least frequently discussed topics in the presentations are liquidity risk, impairment, market trend, credit risk, and cybersecurity risk with standardized frequencies of 0.2%, 0.13%, 0.1%, 0.09%, and 0.02%, respectively. Alternatively, the topics least enquired about by analysts include climate risk, market trend, AI, credit risk, and cybersecurity risk with corresponding frequencies of 0.2%, 0.1%, 0.1%, 0.06%, and 0.02%,

[8] https://www.fredasongdrechsler.com/full-python-code/iclink.

respectively. This comparison of the average importance frequencies between the presentation and Q&A sections underscores the disparities in the focus on various topics between firms and investors.

Insert Table I

Figure IA3 presents the time-series trends of TAD by industry. The TAD index shows a significant increase at the end of 2010. Specifically, the Finance and Mining industries indicate a sudden increase in TAD compared to the industries of Public Administration and Agriculture. This might imply that the Dodd-Frank Act enforced in 2010 has a more pronounced effect on the disclosure quality of specific industries.

4.2. Data and Descriptive Statistic

Panel A of Table II, reports the earnings conference call as hand collected on Seeking Alpha from January 2006-December 2022 outlining the distribution of natural firm characteristics, proxies of information quality, and attributes of the earnings conference calls. This sample comprises all CRSP-listed common stocks that have records of earnings conference calls. The *Ret*[1,*X*] is the abnormal returns adjusted using the Fama-French (1993) three-factor model from *t*+1 to *t*+*X* trading days post earnings call. Idiosyncratic volatility (*IVOL*) is defined as the standard deviation of residuals from a regression of daily stock returns in the preceding *X* trading days (*t*-*X* to *t*-1) on the Fama and French (1993) factors. *SUE* measures deviations in the reported earnings from expectations scaled by the average price over one month. *Complexity* is calculated by measuring firm complexity following the method proposed by Loughran and McDonald (2024). β_{MKT} is the regression coefficient of the past year's daily returns on market returns. *Momentum* is calculated as the buy-and-hold returns from month *m*-12 to *m*-1 prior to the earnings call. *DISP* represents the standard deviation of analyst forecasts scaled by the stock price of the previous month. *AFE* is the absolute forecast errors in the previous month of the earnings call.

Insert Table II

The main sample is comprised of 110,882 earnings conference calls. The mean of TAD_{BoW} is 0.24 with a standard deviation of 0.146. The average sentiment of managers is 0.018, which is greater than the average tone of analysts indicating that managers tend to be more optimistic in the presentation of earnings calls. The mean of *TSD* is 1.266 with a standard deviation of 0.667. The average call length is 7,287 words with a standard deviation of 2,259 words. The average ratio of uncertainty words in the presentations is 0.751%, which is much lower than the mean uncertainty of words (1.159%) in the Q&A section. Both the averages of the Fog and Smog indexes in the presentations are greater than those in the Q&A section. This implies that managers' disclosures are more ambiguous than the information in the Q&A section. Overall, the nature of the conference call shows discrepancies in different attributes between the presentation and Q&A sections. In subsequent tests, we compute alternative proxies of information quality based on the earnings variance including analysts' forecast dispersion (*DISP*) and absolute forecast error (*AFE*).

Additionally, Panel B of Table II presents the pairwise correlation matrix of TAD with numerous other variables, such as alternative measures of information quality. It reveals that TAD_{BoW} and TAD_{GPT} are highly correlated with a correlation of 64%. Therefore, in subsequent tests, we will use TAD_{BoW} as our primary independent variable (hereafter referred to as TAD). TAD exhibits a positive correlation with AFE, but negatively correlates with disagreement and uncertainty as evidenced by its relationships with AFE and DISP. Moreover, we find TAD to be negatively associated with both uncertainty and disagreement. Given that the previous literature documents greater ambiguity (*uncertainty*) in predicting a risk premium (Epstein et al., 2008), the negative association between TAD and ambiguity suggests that the predictability of TAD on future expected returns does not likely originate from uncertainty or ambiguity. This helps to rule out the potential endogeneity of TAD in relation to future stock prices.

4.3. Information Quality with Stock Liquidity

In this subsection, we examine the relation between different proxies of information quality on average daily stock liquidity. We computed two proxies to measure stock liquidity. The first measure of liquidity we use is the average daily bid-ask spread (*BAspread*). Bid-ask spread is an estimator from daily high and low prices following Corwin and Schultz (2012). The second measure we use is the Amihud (2002) illiquidity measure (*AMIHUD*) defined as the average ratio of the daily absolute return to the $dollar trading volume. For each firm, we compute these measures daily, and then average them across different windows around the earnings call dates.

4.3.1 Examine the Effect of TAD with Stock Liquidity.

The concept of TAD is pivotal in understanding the relationship between a firm's information risk and stock liquidity. TAD measures the disparity between the information demanded by market participants and what is supplied by firms. In line with the market microstructure theory, as posited by Kyle (1985) and Glosten and Milgrom (1985), information asymmetry, often heightened by TAD, can lead to lower liquidity. This reduction occurs as uninformed investors engage in “price protection” to mitigate adverse selection risk as noted by Welker (1995). Disclosures that effectively narrow the information gap should, theoretically, enhance stock liquidity (Leuz and Verrecchia, 2000). However, despite clear accounting disclosures, the absence of critical private information can reflect the level of information asymmetry negatively affecting uninformed investor behavior. This aligns with the findings of Bushman and Indjejikian (1995), Leuz and Verrecchia (2000), Shroff et al. (2013), Balakrishnan et al. (2014), Ng (2011), and Goldstein, Yang (2017), who observed a positive correlation between information quality and stock liquidity. Further, Bushee et al. (2018) use Amihud’s (2002) illiquidity as the measure of information asymmetry around earnings conference calls and document a negative relationship between linguistic complexity and information asymmetry. Therefore, we hypothesize that a higher TAD, indicative of increased

information risk or lower information quality, correlates with wider bid-ask spreads and decreased stock liquidity as it deters uninformed investors and discourages market makers from providing liquidity due to adverse selection concerns by investors and market makers (Easley et al., 2004).

To examine the relation between illiquidity and TAD, we regress stock illiquidity following earnings calls on TAD and the control variables. We controlled for different measures of information risk including analysts forecast dispersion, analyst forecast errors, readability, uncertainty, topic sentiment divergence (*TSD*), earnings surprises, and retail investors' disagreement. Additionally, we considered market dynamics like volatility in the one to five trading days' range prior to the earnings calls, momentum, and specific characteristics of the firms including standard unexpected earnings (*SUE*), size (logarithmically expressed), analyst coverage (number of analysts in the preceding quarter), the market-to-book ratio (*MTB*), leverage, return on assets (*ROA*), and sales growth. The characteristics of the earnings calls, like their duration, the diversity of topics discussed, and the tone of both managers and analysts, were also taken into account along with firm and year-quarter fixed effects. To address potential serial correlation in the error term, standard errors are clustered by firm (Petersen, 2009). The regression is described below:

$$Illiqduity_{i,t+1\rightarrow t+n} = \beta_0 + \beta_1 TAD_{i,t} + Controls + Fixed\ Effects + \varepsilon_{it} \quad (4\text{-}1)$$

Panel A of Table III reports a consistently significant positive correlation between the TAD and the bid-ask spread across all five columns with statistical significance at the 1% level. This implies that the average bid-ask spread tends to be higher following the earnings calls with high TAD. These results suggest a correlation between TAD and reduced stock liquidity, likely due to the increased risks associated with information asymmetry.

Insert Table III

In Panel B of Table III, we further investigate the stock liquidity effect of TAD using the Amihud (2002) illiquidity measure (*Amihud*) defined as the ratio of the daily absolute return to the dollar trading volume. We conduct a similar set of regressions with *Amihud* as the

dependent variable by following Equation (4-1). The findings reveal a distinct correlation between the level of TAD and stock liquidity. That is, higher TAD is positively associated with illiquidity. For parsimony, the estimated coefficients on the other information risk measures are not tabulated, but each is positive and significant as expected.[9]

Overall, the evidence in Table III indicates that controlling for other forms of information risk and uncertainty measurements, higher TAD disclosures are associated with a negative impact on future stock liquidity validating the effectiveness of TAD as a proxy for low information quality.[10]

4.3.2 Comparing TAD with Alternative Proxies of Inverse Information Quality with Stock Illiquidity

In this subsection, we explore various proxies of inverse information quality and their relationship with stock illiquidity during the post-earnings call period. Following Ng (2011), we focus on earnings as a crucial source of information for investors. Verrecchia (2001) proposes a theory suggesting that investor uncertainty regarding firm value and adverse selection among investors is heightened when information precision is low (i.e., higher variance), ceteris paribus. Consequently, this paper employs two alternative measures to capture the precision of earnings signals. These measures share a common characteristic. They quantify the second moment associated with an earnings signal. Specifically, we construct two proxies, analyst forecast dispersion and analysts' forecast errors, to assess the inverse quality of information (or information risk). The data used to calculate analysts' forecast dispersion and analysts' forecast errors are sourced from the CRSP daily and I/B/E/S detailed actual database. Analysts' Forecast Dispersion (*DISP*) quantifies the degree of variance in analyst

[9] The other information risk measures include analyst forecast dispersion and analyst mean errors. Both show a positive correlation with illiquidity, consistent with Ng (2011).

[10] In supplemental (untabulated) tests, I found that disclosures with high TAD (and the alternative measure, TAD_{GPT}) exhibit a strong positive relationship with the average bid-ask spread and the Amihud (2002) illiquidity ratio over different periods including intraday and future intervals of one week, two weeks, one month, and one quarter.

forecasts scaled by the mean stock price of the previous month. Absolute Forecast Errors (*AFE*) represent the absolute value of the discrepancy between announced earnings and the mean forecasted earnings scaled by the mean stock price of the previous month. To compute analysts' forecast dispersion, we stipulate that at least three analysts must cover the firm. In Table IV, the control variables encompass turnover, prior month's return, volatility, sales growth, analysts' coverage, the market-to-book ratio, leverage, profitability, market capitalization, and institutional ownership. *Turnover* is defined as the trading volume divided by the number of outstanding shares. Prior return captures the stock return from the preceding month. *Volatility* is determined by the standard deviation of daily stock returns over one week prior to the earnings calls. *Size* is denoted by the log of market capitalization. The book-to-market ratio (*MTB*) reflects the equity value relative to its book value. *Sales growth* is measured by the change in sales. For a detailed explanation of these terms, please refer to Appendix A.

$$Illiqduity_{i,t+1\to t+n} = \beta_1 \times TAD_{i,t} + \beta_2 \times Other\ Information\ Risk_{i,t} + Controls + Fixed\ Effects + \varepsilon_{it+1\to t+n} \quad (4\text{-}2a)$$

$$Illiqduity_{i,t+1\to t+n} = \beta_1 \times Information\ Risk_{i,t} + Controls + Fixed\ Effects + \varepsilon_{it+1\to t+n} \quad (4\text{-}2b)$$

Table IV illustrates the results of regressions exploring the cross-sectional effects of information risk (poor information quality) as indicated by absolute forecast error (*AFE*), analysts' forecast dispersion (*DISP*), and TAD on stock illiquidity. We expect the regression to follow specification Equation (4-2a) and exhibit a positive correlation between *illiquidity*, AFE, DISP, and TAD. Specifically, the first columns of Panels A and B present the outcomes when AFE, DISP, and TAD are analyzed together. All of the coefficients for these information risk factors on illiquidity are positive and statistically significant at the 1% level. Notably, in Panel A, the coefficient for AFE is 0.0789 (t=15.32), while in Panel B, it is 0.0772 (t=11.64) suggesting that a one standard deviation increase in AFE leads to a 7.89% (Panel A) and 7.72% (Panel B) increase in the bid-ask spread (Amihud Illiquidity). From Specifications (2)-(4) in

both panels, with the regression following Equation (4-2b), a one standard deviation increase in TAD corresponds to a 1.20% (Panel A) and 2.15% (Panel B) increase in the bid-ask spread (*Amihud*) with t-values of 5.99 and 8.21, respectively. For a one standard deviation increase in AFE, the spread rises by 9.45% (Panel A) and 7.72% (Panel B) with t-values of 19.35 and 13.16, respectively. Finally, for a one standard deviation increase in DISP, the spread grows by 7.83% (Panel A) and 4.12% (Panel B) with t-values of 16.11 and 6.82, respectively. These positive and significant coefficients align with expectations (Lambert et al., 2007; Ng, 2011). Intriguingly, after adjusting for *AFE* and *DISP*, the impact of TAD on *illiquidity* increases by 9.10% (Panel A) and 5.2% (Panel B), respectively.

Table IV

In summary, the results presented in Table IV demonstrate that various proxies for inverse information quality (*information risk*) are negatively associated with stock liquidity. Furthermore, TAD remains statistically significant even after controlling for alternative proxies of poor information quality. This suggests that TAD incorporates textual information beyond the precision of quantitative earnings information.

4.4. TAD and Stock Returns

In this section, we employ several steps to test the pricing of TAD. First, we look at the distribution of stock returns across portfolios of stocks single sorted by TAD and double-sorted by market capitalization (*Size*) and then by TAD. In addition, we test whether TAD survives in spanning tests using the Fama and French (1993) risk factors. Finally, we conduct a variety of robustness tests on the pricing of TAD including alternative TAD based on a text embedding approach.

4.4.1 Portfolio Approach

To assess our hypothesis that the price-based information risk, as indicated by TAD, influences future stock returns, we have constructed both value-weighted (shown in Panels A and B) and equally-weighted (depicted in Panels C and D) monthly portfolios sorted by TAD in Table V. The long-short portfolio is constructed by dynamically sorting the TAD of all firms within the last quarter and rebalancing the portfolio on a daily basis. The stock returns initiate at one trading day following the earnings call at day t=0 (See IB.3 for the details to identify the current day of the earnings call). Panel A presents the average monthly returns (*Ret*) and the Fama-French-Carhart four-factor risk-adjusted portfolio alphas (Ret_{Adj}) for portfolios organized based on TAD. Panel B details the Ret_{Adj} for portfolios double-sorted monthly, initially by market capitalization as of the previous June (*Size*) and subsequently by the prior quarter's TAD. This panel also includes future values of *Ret* spanning various timeframes, ranging from intraday to 26 weeks, and highlights the disparities in *Ret* and Ret_{Adj} between the top and bottom portfolios underscored by the Newey-West (1987) standard errors presented in brackets. This analysis spans from January 2006-December 2022.

Insert Table V

In a similar vein, Panel C illustrates the monthly excess returns (*Ret*) and risk-adjusted portfolio alphas (Ret_{Adj}) single-sorted by TAD. Alternatively, Panel D displays the risk-adjusted alphas (Ret_{Adj}) that are value-weighted and double-sorted by market capitalization and then by TAD. Furthermore, it reports the differences in *Ret* and Ret_{Adj} between the highest and lowest quintile portfolios complete with Newey-West (1987) standard errors enclosed in parentheses. The period covered in this analysis extends from January 2006-December 2022.

The results of Panels A and C indicate a monotonic increasing relation between TAD and future stock returns. The two panels report that from Quintile 1 to Quintile 5, the relation is monotonically increasing. The difference between the high and low TAD quintile value weighted portfolios is both positive and significant at the 5% level, and the difference between the high and low quintile equally-weighted portfolios are both positive and significant at the

1% level. Most importantly, the return spreads between the high and low TAD quintile portfolios are significant economically. A trading strategy combining a long position in a high TAD equally-weighted quintile portfolio with a short position in a low TAD quintile equally-weighted portfolio generate a 8.189% annualized excess return and a 8.589% annualized (Fama-French-Carhart) risk-adjusted return. The trading strategy combining a long position in a high TAD value weighted quintile portfolio with a short position in a low TAD quintile portfolio generate a 6.638% annualized excess return and a 8.175% annualized (Fama-French-Carhart) risk-adjusted return.

There might be a concern that the positive risk premium for TAD is simply a manifestation of the return effect related to firm size. To address this potential concern, we employ double-sorted portfolio returns in Panels B and D to provide robust evidence that the positive relation between TAD and future stock returns is not fully driven by market capitalization. The difference in risk-adjusted returns between the high and low TAD quintile portfolios is statistically significant in three of five *Size* quintiles for both the equally-weighted and value weighted double-sorted portfolios. The long high-TAD and short low-TAD trading strategy applied in the lowest size quintile portfolio yields 12.95% and 11.35% annualized risk-adjusted alphas from value-weighted and equally-weighted portfolios, respectively. Overall, the portfolio returns provide the first evidence that the text-based information risk provided by TAD positively affects future stock returns. The positive relationship of TAD and future expected returns support our hypothesis that TAD is positively associated with the cost of capital.

4.4.2. Spanning Tests Using Fama-French Factors

To better account for differences in the riskiness of information risk-based portfolios, we examine the risk-adjusted performance of TAD-based trading strategies using various unconditional factor models. These extended factor models allow me to control for the effects of additional factors and allow the factor sensitivity to vary over time.

Insert Table VI

The unconditional factor model estimates are reported in Table VI. The unconditional factor models contain some combination of factors of the market (*RMRF*) factor, the size (*SMB*) factor, the value (*HML*) factor, the momentum (*UMD*) factor, the profitability (*RMW*) factor, the investment (*CMA*) factor, and the liquidity (*LIQ*) factor where the estimation period is from January 2006-December 2022. We find that the performance of TAD-based portfolios remain economically significantly event when we include a large number of factors in the risk adjustment models. For example, in Specifications (10)-(12), the monthly seven-factor alpha (Newey-West Standard errors) estimates for Long, Short, and Long-Short portfolios are 1.064% (0.178), 0.509% (0.131), and 0.555% (0.225), respectively. The Long-Short alpha estimate translates into an annual risk-adjusted performance of about 6.86%, statistically at 5% level, and the strategy does not rely on the ability to take a short position.

Across the various factor models, the two main factors that are important in explaining the Long-Short portfolio return are the market risk (*RMRF*) factor and the size (*SMB*) factor. In the previous section, we check the alphas' pattern of the double-sorted portfolios on *Size* and TAD to ensure that the TAD prediction model does not repackage the size effect.

In sum, we find that the Long-Short portfolios sorted by TAD cannot be spanned by the Fama and French (1993) risk factors suggesting that TAD captures additional risk information beyond the existing risk factors.

4.4.3. Fama-MacBeth Cross-Sectional Regression Estimates

In this section, we follow the Fama and MacBeth (1973) method with cross-sectional return determinants including market beta, market capitalization, and the market-to-book ratio as controls. In addition, following Brennan et al. (2012), we include idiosyncratic volatility, liquidity, information quality, past stock returns, and textual characteristics of the earnings conference call. For each month, we run cross-sectional regression stock excess returns on the return determinants as follows.

$$R_{i[t+1,t+n]} = \beta_0 + \beta_1 TAD_{i,t} + \beta_2 \beta_{MKT,it} + \beta_3 Size_{it} + \beta_4 MTB_{it} + \beta_5 SUE_{it} + \beta_6 DISP_{it} + \beta_7 ROA_{it} + \beta_8 Leverage_{it} + \beta_9 Length_{it} + \beta_{10} MOM_{it} + \beta_{11} Tone_{Mgrs,it} + \beta_{12} Tone_{Alyts,it} + \beta_{13} FOG_{it} + \beta_{14} I.O._{it} + \beta_{15} TSD_{it} + \beta_{16} Illiquidity_{it} + \varepsilon_{it+1} \quad (4\text{-}3)$$

where $R_{i,t+1}$ is the quarterly excess stock return for firm *i* from one day following the earnings conference call time *t*+1 to 60 trading days post the call, *t+60*, and TAD is the topic attention divergence that measures the different narrative focus between the managers and analysts in the earnings conference call at time *t*. $\beta_{MKT,it}$ is the market beta computed from the past one year's daily stock returns on the Fama-French-Carhart model. *Size* is the market capitalization, *MTB* is the market-to-book ratio, and *SUE* is the standard unexpected earnings. Profitability is measured as return on assets (*ROA*). Additional factors include analysts' coverage (*Coverage*), analysts forecast dispersion (*DISP*), financial leverage (*Leverage*), and call length (*Length*). Momentum is the past 11 months stock returns from 12 months to one month prior to each firms' earnings call (*MOM*). Tone of managers ($Tone_{Mgrs}$), Tone of Analysts ($Tone_{Alyts}$), Readability (*FOG*), Institutional Ownership (*I. O.*), and topic sentiment divergence (*TSD*) measured by the aggregated topic level sentiment variance in the Q&A section of the earnings call, are other factors contributing to information quality.

$$\frac{R_{i[t+1,t+m]}}{\mathrm{m}} = \beta_0 + \beta_1 TAD_{i,t} + \beta_2 \beta_{MKT,it} + \beta_3 Size_{it} + \beta_4 MTB_{it} + \beta_5 SUE_{it} + \beta_6 DISP_{it} + \beta_7 ROA_{it} + \beta_8 Leverage_{it} + \beta_9 Length_{it} + \beta_{10} MOM_{it} + \beta_{11} Tone_{Mgrs,it} + \beta_{12} Tone_{Alyts,it} + \beta_{13} FOG_{it} + \beta_{14} I.O._{it} + \beta_{15} TSD_{it} + \beta_{16} Illiquidity_{it} + YearQ\ FE + \varepsilon_{it+1} \quad (4\text{-}4)$$

Insert Table VII

Panel A of Table VII reveals several notable findings. Our hypothesis states that uninformed investors demand a risk premium for trading stocks with informed investors prior

to earnings conference call. As such, stocks with high information risk measured by TAD should compensate for uninformed investors' potential trading losses. We predict higher TAD is positively associated with future cross-sectional stock returns. In our empirical test, the coefficients of TAD are all positive with t-statistics varying from 2.44 to 3.38 compared with the results of the portfolio approach.

In addition, we find consistent signals and significance levels for the coefficients of other conventional pricing factors except for *Size*, *Illiquidity*, and *MTB*. For example, *MOM*, β_{MKT}, *IVOL*, *FOG,* and $Ret_{[-5,-1]}$ have a negative association with quarterly excess stock returns. *ROA*, *Tone*, and *SUE* are positively associated with quarterly excess stock returns.

In Panel B of Table VII, aligning with the methodology of Yang et al. (2020), the relationship between the TAD index and long-term average expected stock returns is examined through panel regressions with using a quarter fixed effect. The standard errors clustered are adjusted for heteroskedasticity and clustered at both the firm and quarterly levels. The findings indicate that the TAD index possesses predictive power for future cumulative stock returns over a long horizon. However, subsequent tests on long-term weekly average returns reveal that TAD's predictive strength primarily emanates from the short-term as evidenced by a diminishing slope coefficient of TAD over time. Despite this, TAD's coefficients on future returns consistently remain positive up to 52 weeks after its construction. When juxtaposed with *SUE*, which exerts a significantly stronger effect in the immediate aftermath of the earnings calls, TAD's impact, although initially smaller by a factor of approximately 5.5 times, surpasses that of *SUE* by 20% in terms of the coefficient magnitude over a 52-week period. This suggests TAD's greater efficacy in forecasting long-term asset performance.

Panel B utilizes a regression model with quarterly fixed effects, and reports robust standard errors adjusted for heteroskedasticity and clustered at both the firm and quarterly levels.

Insert Figure 3

Figure 3 provides a visual illustration of the impact of the Tone at the Day of Announcement (*TAD*) and Topic Sentiment Divergence (*TSD*) on the variability of post-earnings announcement stock returns showcasing contrasting patterns between TAD and TSD in predicting future stock returns. Specifically, Figure 3a depicts the coefficients of regressing cumulative abnormal returns on TAD following the earnings conference call, highlighting a discernible trend of increasing returns following the earnings call that demonstrates the increasingly predictive patterns of TAD in cross-sectional regressions over a 60-day trading period. In terms of economics significance, a one standard deviation increase in TAD is associated with a 0.19% (0.45%) higher abnormal return on the current day (over a quarter) indicating the cross-sectional stock returns remain to increase by 0.26% over the following one quarter after earnings calls. Figure 3b complements this by charting the influence of TSD on returns presenting an impounded negative impact on returns, rather than a gradual decrease. This pattern appears to be stable and not reverse as the post-earnings announcement period progresses. Both subfigures are grounded in the coefficients from panel regressions that incorporate a robust set of control variables alongside firm fixed effects and year-quarter fixed effects. The standard errors of these regression models are clustered by firm and year-quarter ensuring that the observed relationships are not artifacts of firm-specific volatility or seasonal reporting patterns. Collectively, the evidence underscores the significance of divergence of attention and tone on stock returns following the earnings announcement, respectively. The comparison between these two patterns implies topic attention (narratives) is more nuanced than sentiment and it's harder for investors to process.

4.5. Cross-Sectional Test by Information Processing Costs

In this subsection, we conduct a cross-section analysis to investigate the impact of TAD on future stock returns across firms with varying levels of information processing costs. Drawing on the methodology of Kim and Liu (2018), firms are classified into high and low information processing cost categories based on the median values of analysts' coverage,

institutional ownership, and firm size. This classification is informed by the challenge of directly quantifying information quality, a task that previous studies have approached through theoretical or experimental frameworks.

$$Ret_{i[t+1,t+n]} = \beta_1 \times TAD_{i,t} + \beta_2 \times Other\ Information\ Risk_{i,t} + Controls + Firm\ FE + YearQ\ FE + \varepsilon_{it} \quad (4\text{-}5)$$

Insert Table VIII

My analysis reveals a significant relationship between TAD and future cross-section risk adjusted returns ($Ret_{i[t+1,\ t+60]}$) and is particularly pronounced in firms characterized by higher information processing costs. Specifically, Table VIII divides our sample based on institutional ownership (Specifications 1 and 2), analysts' coverage (Specifications 3 and 4), and market capitalization (Specification 5 and 6) using the median as a benchmark for classification.

For firms with lower institutional ownership (Specification 1), we observe the robust positive predictive power of TAD on future stock returns. A one standard deviation increase in TAD is associated with a 0.428% increase in accumulative risk-adjusted returns over one quarter. However, this effect is not significant in the high institutional ownership group (Specification 2). A similar pattern emerges when categorizing firms by analysts' coverage where lower coverage firms exhibit a stronger predictive relationship between TAD and future asset returns. Furthermore, when classifying stocks by market capitalization, we find that TAD's predictive power is markedly stronger for smaller firms.

These findings support the conjecture that TAD serves as an indicator of a firm's information quality. In environments where information asymmetry is more pronounced, signals by lower levels of institutional ownership, fewer financial analysts covering the firm, and smaller firm size, TAD becomes a critical predictor of future stock performance. This suggests that TAD captures aspects of the information environment that are particularly relevant for firms facing higher barriers to information processing and dissemination.

5. Robustness Tests

5.1. Topic Sensitivity Test

In this subsection, we examine how each of the individual topics (listed in Internet Appendix) influences our results. Specifically, we report the scaled frequency of each topic. We also examine whether the findings from Table VII of the paper are dependent upon any specific topic by running panel regressions controlling for firm characteristics and alternative anomalies that could impact the future asset prices.

$$\begin{aligned} R_{i[t+1,t+60]} = {} & \beta_1 TAD_{i,t} + \beta_2 \beta_{MKT,it} + \beta_3 Size_{it} + \beta_4 MTB_{it} + \beta_5 SUE_{it} \\ & + \beta_6 DISP_{it} + \beta_7 ROA_{it} + \beta_8 Leverage_{it} + \beta_9 Length_{it} \\ & + \beta_{10} MOM_{it} + \beta_{11} Tone_{Mgrs,it} + \beta_{12} Tone_{Alyts,it} + \beta_{13} FOG_{it} \\ & + \beta_{14} I.O._{it} + \beta_{15} TSD_{it} + \beta_{16} Illiquidity_{it} + Firm\ FE + YearQ\ FE \\ & + \varepsilon_{it+1} \end{aligned} \tag{5-1}$$

Insert Figure 3

Figure 4 presents the topic sensitivity analysis. The dependent variable is the Fama-French three factor model adjusted cumulative returns from the one following day *t* to 60 days (trading days) post earnings calls. Table I presents the topic frequencies of the three most prevalent topics in the presentation of earnings conference calls: “guidance” (19.43% of all topics), “Growth” (9.24%), and “Investment” (8.97%). The three most prevalent topics in Q&A section are “Guidance” (21.53%), “Urgency & Emphasis” (9.19%), and “Investment” (7.5%). Figure 3 illustrates the importance of each topic in predicting future stock returns. It shows that topics such as "guidance," "dividends," and "impairment" are the most significant in impacting TAD’s predictability of future stock returns.

It indicates the predictability of topics do not perfectly align with the frequency of topics in the earnings call. Additionally, we observe that after excluding specific topics from TAD, the findings from Table VII are still robust. The predictability can even increase after excluding the topics “cash flow” and “sovereign risk.” Overall, we conclude that our findings

from Table VII are not sensitive to the inclusion of any specific topics in TAD, and the predictability of TAD could be optimized by adjusting the weight of each topic depending on the informativeness of each topic.

Figure 4 displays the results of the topic sensitivity analysis. The dependent variable is the Fama-French (1993) three-factor model adjusted cumulative returns from the one day following time *t* to 60 trading days post-earnings calls. Table we identifies the most frequent topics discussed during earnings conference calls, with "Guidance" accounting for 19.43% of all topics, followed by "Growth" (9.24%) and "Investment" (8.97%). Conversely, in the Q&A sections, "Guidance" remains the most prevalent topic at 21.53%, with "Urgency & Emphasis" (9.19%), and "Investment" (7.5%) also featuring prominently. Figure 4 highlights significant discrepancies in topics such as "Guidance," "Dividends," and "Impairment" between the managers' disclosures and investors' attention, markedly influencing the predictability of TAD in forecasting future stock returns. This observation suggests that the predictability of topics does not perfectly align with their frequency in earnings calls. Moreover, after excluding specific topics from *TAD*, the robustness of the findings presented in Table VII remains unaffected. Notably, the predictive capability of TAD may even improve upon the exclusion of topics like "Cash Flow" and "Sovereign Risk."

In conclusion, the results from Figure 4 demonstrate that even the inclusion of any specific topic in TAD does not impact the statistical significance of the results as shown in Table VII. However, optimizing the weight assigned to each topic based on its informativeness could enhance the predictive accuracy of TAD.

6. Conclusion

This paper undertakes a critical examination of the link between information quality and capital markets, unravelling the complex dynamics of corporate disclosure and information asymmetry. Through the innovative measure of Topic Attention Divergence (TAD), it offers a

fresh perspective as to how discrepancies in topic attention during earnings calls can proxy for information quality. This direct approach to measuring information quality starkly contrasts with the indirect methodologies traditionally adopted in the field, marking a leap forward in the analysis of corporate communications and information asymmetry.

The research findings highlight that a high TAD, indicative of lower information quality, forecasts increased liquidity risk and equity premiums. These findings lend empirical support to the theoretical model proposed by Easley et al. (2004), which posits that information quality inversely impacts the cost of capital through heightened uncertainty and adverse selection issues. The pronounced effect of this relationship in firms burdened with higher information processing costs underscores the indispensable role of institutional investors and financial analysts in fostering information transparency and mitigating asymmetry.

By exploring whether information quality (or *information risk*) is priced in asset pricing, this paper broadens the discourse on asset pricing and market efficiency. It leverages a text-based measure to ascertain corporate insiders' propensity to meet the informational demands of outsiders. The study's nuanced approach reveals that higher TAD not only predicts higher liquidity risk and equity premiums, but also signifies the critical interplay between information quality and the function of financial markets.

In summary, this research significantly enriches the existing body of literature by elucidating the nuances of corporate disclosure practices and their implications for stock liquidity, information asymmetry, and asset pricing. It emphasizes the crucial need for effective communication between firms and investors and heralds the pivotal role of technology, particularly advancements in AI and data analytics, in enhancing market efficiency. As it lays the groundwork for future exploration, this paper underscores the transformative potential of technological advancements in reshaping our understanding of financial markets and corporate behavior, paving the way for more informed and efficient market practices.

References

AI, H. (2010). Information quality and long-run risk: Asset pricing implications. *Journal of Finance 65*(4), 1333-1367. https://doi.org/10.1111/j.1540-6261.2010.01572.x

Akins, B.K., Ng, J., and Verdi, R.S. (2012). Investor competition over information and the pricing of information asymmetry. *The Accounting Review 87*(1), 35-58. http://www.jstor.org/stable/41408060

Amihud, Y. (2002). Illiquidity and stock returns: Cross-section and time-series effects. *Journal of Financial Markets 5*(1), 31-56. https://doi.org/10.1016/S1386-4181(01)00024-6

Ang, A., Hodrick, R.J., Xing, Y., and Zhang, X. (2006). The cross-section of volatility and expected returns. *Journal of Finance* 61(1), 259-299. https://doi.org/10.1111/j.1540-6261.2006.00836.x

Balakrishnan, K., Billings, M.B., Kelly, B., and Ljungqvist, A. (2014). Shaping liquidity: On the causal effects of voluntary disclosure. *Journal of Finance 69*(5), 2237-2278. https://doi.org/10.1111/jofi.12180

Beyer, A. and Guttman, I. (2010). Voluntary disclosure, disclosure bias and real effects. *Journal of Accounting Research*. *50*(5), 1142-1177. https://doi.org/10.1111/j.1475-679X.2012.00459.x

Blei, D.M., Ng, A.Y., and Jordan, M.I. (2003). Latent dirichlet allocation. *Journal of Machine Learning Research 3*, 993-1022. https://doi.org/https://dl.acm.org/doi/10.5555/944919.944937

Botosan, C.A. (1997). Disclosure level and the cost of equity capital. *The Accounting Review 72*(3), 323-349. http://www.jstor.org/stable/248475

Brennan, M.J., Chordia, T., Subrahmanyam, A., and Tong, Q. (2012). Sell-order liquidity and the cross-section of expected stock returns. *Journal of Financial Economics 105*(3), 523-541. https://doi.org/10.1016/j.jfineco.2012.04.006

Brevik, F. and D'Addona, S. (2010). Information quality and stock returns revisited. *Journal of Financial and Quantitative Analysis 45*(6), 1419-1446. http://www.jstor.org/stable/27919572

Brown, L.D., Call, A.C., Clement, M.B., and Sharp, N.Y. (2019). Managing the narrative: Investor relations officers and corporate disclosure. *Journal of Accounting and Economics 67*(1), 58-79. https://doi.org/https://doi.org/10.1016/j.jacceco.2018.08.014

Bushee, B.J., Gow, I.D., and Taylor, D.J. (2018). Linguistic complexity in firm disclosures: Obfuscation or information? *Journal of Accounting Research 56*(1), 85-121. https://doi.org/10.1111/1475-679X.12179

Bushman, R.M. and Indjejikian, R.J. (1995). Voluntary disclosures and the trading behavior of corporate insiders. *Journal of Accounting Research 33*(2), 293-316. https://doi.org/10.2307/2491489

Bybee, L., Kelly, B.T., Manela, A., and Xiu, D. (2023). Business news and business cycles. *Journal of Finance,* forthcoming. http://dx.doi.org/10.2139/ssrn.3446225

Choi, J., Jin, L., and Yan, H. (2016). Informed trading and the cost of capital. *NBER Working Paper*, *18680*.

Cookson, J.A. and Niessner, M. (2020). Why don't we agree? Evidence from a social network of investors. Journal of Finance, 75(1), 173-228. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2754086

Core, J.E., Guay, W.R., and Verdi, R. (2008). Is accruals quality a priced risk factor? *Journal of Accounting and Economics* 46(1), 2-22. https://doi.org/10.1016/j.jacceco.2007.08.001

Corwin, S.A. and Schultz, P. (2012). A simple way to estimate bid-ask spreads from daily high and low prices. *Journal of Finance 67*(2), 719-760. https://doi.org/10.1111/j.1540-6261.2012.01729.x

Dang, T.L., Moshirian, F., and Zhang, B. (2015). Commonality in news around the world. *Journal of Financial Economics*, *116*(1), 82-110. https://doi.org/10.1016/j.jfineco.2014.11.007

Daniel, K., Grinblatt, M., Titman, S., and Wermers, R. (1997). Measuring mutual fund performance with characteristic-based benchmarks. *Journal of Finance*, 52(3), 1035-1058.

Devlin, J., Chang, M.W., Lee, K., and Toutanova, K. (2018). BERT: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.

Durnev, A., Morck, R., and Yeung, B. (2004). Value-enhancing capital budgeting and firm-specific stock return variation. *Journal of Finance* 59(1), 65-105. https://doi.org/10.1111/j.1540-6261.2004.00627.x

Easley, D., Hvidkjaer, S., and O'Hara, M. (2002). Is information risk a determinant of asset returns? *Journal of Finance 57*(5), 2185-2221. https://doi.org/10.1111/1540-6261.00493

Easley, D., Kiefer, N.M., O'Hara, M., & Paperman, J.B. (1996). Liquidity, information, and infrequently traded stocks. *Journal of Finance 51*(4), 1405-1436. https://doi.org/10.1111/j.1540-6261.1996.tb04074.x

Easley, D. and O'Hara, M. (2004). Information and the cost of capital. *Journal of Finance* 59(4), 1553-1583. https://doi.org/10.1111/j.1540-6261.2004.00672.x

Ecker, F. (2013). Information precision and long-run performance of initial public offerings. *Contemporary Accounting Research*, 31(3), 876-910. https://doi.org/10.1111/1911-3846.12043

Epstein, L.G. and Schneider, M. (2008). Ambiguity, information quality, and asset pricing. *Journal of Finance 63*(1), 197-228. https://doi.org/10.1111/j.1540-6261.2008.01314.x

Fama, E.F. and French, K.R. (1993). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics* 33(1), 3-56. https://doi.org/10.1016/0304-405X(93)90023-5

Fama, E.F. and MacBeth, J.D. (1973). Risk, return, and equilibrium: Empirical tests. *Journal of Political Economy 81*(3), 607-636. http://www.jstor.org/stable/1831028

Ferreira, M.A. and Laux, P.A. (2007). Corporate governance, idiosyncratic risk, and information flow. *Journal of Finance 62*(2), 951-989. https://doi.org/10.1111/j.1540-6261.2007.01228.x

Francis, J., Nanda, D., and Olsson, P. (2008). Voluntary disclosure, earnings quality, and cost of capital. *Journal of Accounting Research* 46(1), 53-99. https://doi.org/10.1111/j.1475-679X.2008.00267.x

Frankel, R., Johnson, M., and Skinner, D.J. (1999). An empirical examination of conference calls as a voluntary disclosure medium. *Journal of Accounting Research 37*(1), 133-150. https://doi.org/10.2307/2491400

Frankel, R., Kothari, S.P., and Weber, J. (2006). Determinants of the informativeness of analyst research. *Journal of Accounting and Economics*. *41*(1-2), 29-54. https://doi.org/https://doi.org/10.1016/j.jacceco.2005.10.004

Glosten, L.R. and Milgrom, P.R. (1985). Bid, ask and transaction prices in a specialist market with heterogeneously informed traders. *Journal of Financial Economics 14*(1), 71-100. https://doi.org/10.1016/0304-405X(85)90044-3

Goldstein, I. and Yang, L. (2017). Information disclosure in financial markets. *Annual Review of Financial Economics* 9, 101-125. https://doi.org/10.1146/annurev-financial-110716-032355

Healy, P.M. and Palepu, K.G. (2001). Information asymmetry, corporate disclosure, and the capital markets: A review of the empirical disclosure literature. *Journal of Accounting and Economics 31*(1-3), 405-440. https://doi.org/10.1016/S0165-4101(01)00018-0

Hoffman, M.D., Blei, D.M., and Bach, F. (2010). Online learning for Latent Dirichlet Allocation. In Proceedings of the 23rd International Conference on Neural Information

Processing Systems - Volume 1 (pp. 856-864). Published December 6, 2010. https://dl.acm.org/doi/10.5555/2997189.2997285

Hollander, S., Pronk, M., and Roelofsen, E. (2010). Does silence speak? An empirical analysis of disclosure choices during conference calls. *Journal of Accounting Research 48*(3), 531-563. https://doi.org/10.1111/j.1475-679X.2010.00365.x

Huang, A.H., Wang, H., and Yang, Y. (2022). FinBERT: A large language model for extracting information from financial text. *Contemporary Accounting Research* 40(2), 806-841. https://doi.org/10.1111/1911-3846.12832

Hughes, J.S., Liu, J., and Liu, J. (2007). Information asymmetry, diversification, and cost of capital. *The Accounting Review 82*(3), 705-729. https://doi.org/10.2308/accr.2007.82.3.705

Kim, A. (2014). The value of firms' voluntary commitment to improve transparency: The case of special segments on Euronext. *Journal of Corporate Finance 25*, 342-359. https://doi.org/10.1016/j.jcorpfin.2013.12.012

Kim, J.-B., Li, B., and Liu, Z. (2018). Information-processing costs and breadth of ownership. *Contemporary Accounting Research* 36(4), 2408-2436. https://doi.org/10.1111/1911-3846.12451

Kyle, A.S. (1985). Continuous auctions and insider trading. *Econometrica* 53(6), 1315-1335. https://doi.org/10.2307/1913210

Lambert, R., Leuz, C., and Verrecchia, R.E. (2007). Accounting information, disclosure, and the cost of capital. *Journal of Accounting Research 45*(2), 385-420. http://www.jstor.org/stable/4622036

Lambert, R.A., Leuz, C., and Verrecchia, R.E. (2012). Information asymmetry, information precision, and the cost of capital. *Review of Finance* 16(1), 1-29, https://doi.org/10.1093/rof/rfr014

Leuz, C. and Verrecchia, R.E. (2000). The economic consequences of increased disclosure. *Journal of Accounting Research 38*, 91-124. https://doi.org/10.2307/2672910

Livnat, J. and Mendenhall, R.R. (2006). Comparing the post-earnings announcement drift for surprises calculated from analyst and time series forecasts. *Journal of Accounting Research*, 44(1), 177-205. https://doi.org/10.1111/j.1475-679X.2006.00196.x

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

Loughran, T. and McDonald, B. (2024). Measuring firm complexity. *Journal of Financial and Quantitative Analysis* 1-28. https://doi:10.1017/S0022109023000716

Manning, C.D., Raghavan, P., and Schütze, H. (2008). Introduction to Information Retrieval. Cambridge University Press. ISBN 9781139472104.

Matsumoto, D., Pronk, M., and Roelofsen, E. (2011). What makes conference calls useful? The information content of managers' presentations and analysts' discussion sessions. *The Accounting Review 86*(4), 1383-1414. https://www.jstor.org/stable/23045606

Mayew, W.J. and Venkatachalam, M. (2012). The power of voice: Managerial affective states and future firm performance. *Journal of Finance 67*(1), 1-43. https://doi.org/10.1111/j.1540-6261.2011.01705.x

Morck, R., Yeung, B., and Yu, W. (2000). The information content of stock markets: Why do emerging markets have synchronous stock price movements? *Journal of Financial Economics* 58(1-2), 215-260. https://doi.org/10.1016/S0304-405X(00)00071-4

Newey, W.K. and West, K.D. (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica* 55(3), 703-708. https://doi.org/10.2307/1913610

Ng, J. (2011). The effect of information quality on liquidity risk. *Journal of Accounting and Economics 52*(2-3), 126-143. https://doi.org/10.1016/j.jacceco.2011.03.004

Roll, R. (1988). R^2. *Journal of Finance 43*(3), 541-566. https://doi.org/10.1111/j.1540-6261.1988.tb04591.x

Salton, G., & McGill, M. J. (1983). Introduction to modern information retrieval (2nd ed.). McGraw-Hill.

Shroff, N., Sun, A.X., White, H.D., and Zhang, W. (2013). Voluntary disclosure and information asymmetry: Evidence from the 2005 securities offering reform. *Journal of Accounting Research*,51(5), 1299-1345. https://doi.org/10.1111/1475-679X.12022

Tetlock, P. C. (2007). Giving content to investor sentiment: The role of media in the stock market. The Journal of Finance, 62(3), 1139-1168. https://doi.org/10.1111/j.1540-6261.2007.01232.x

Verrecchia, R.E. (2001). Essays on disclosure. *Journal of Accounting and Economics* 32(1-3), 97-180. https://doi.org/10.1016/S0165-4101(01)00025-8

Welker, M. (1995). Disclosure policy, information asymmetry, and liquidity in equity markets. *Contemporary Accounting Research* 11(2), 801-827. https://doi.org/10.1111/j.1911-3846.1995.tb00467.x

Xiao, H. and Stibor, T. (2010). Efficient collapsed Gibbs sampling for latent dirichlet allocation. Proceedings of 2nd Asian Conference on Machine Learning, Proceedings of Machine Learning Research, *13*, 63-78. Retrieved from https://proceedings.mlr.press/v13/xiao10a.html.

Yang, Y.C., Zhang, B., and Zhang, C. (2020). Is information risk priced? Evidence from abnormal idiosyncratic volatility. *Journal of Financial Economics 135*(2), 528-554. https://doi.org/10.1016/j.jfineco.2019.06.013

Appendix

A.1 Outcome Variables

- *Bid-Ask Spread (BAspread,* Table III, Panel A) – the daily average bid-ask spread is estimated from daily high and low prices by following Corwin and Schultz (2012). It computed the *BAspread* across different windows. If an earnings call ends before the pre-market trading period, the intraday *t* is as same as the earnings call day. Otherwise, day *t* is the following trading day of the earnings call day on *IBES*. The dependent variable, daily average BAspread, ranges from varying windows including one week (from day [*t+1*] to [*t+5*]), two weeks (from day [*t+1*] to [*t+10*]), one month (from day [*t+1*] to [*t+20*]), and one quarter (from day [*t+1*] to [*t+60*]) after the earnings call. (Source: CRSP)
- *Amihud Illiquidity* (Amihud, Table III, Panel B) – the Amihud (2002) illiquidity measure is defined as the ratio of daily absolute returns to the dollar trading volume. For quarterly analyses, this variable represents the average across different windows including one week (from day [*t+1*] to [*t+5*]), two weeks (from day [*t+1*] to [*t+10*]), one month (from day [*t+1*] to [*t+20*]), and one quarter (from day [*t+1*] to [*t+60*]) after the earnings call. (Source: CRSP)
- For earnings calls issued after trading hours (4:00 p.m.), the day zero return is the CRSP return for the trading day following the earnings call. For reports issued or before market close, the day zero return is the CRSP return for the intraday of the earnings call. We measure the cumulative returns from day one through *x* trading days. (Source: CRSP)
    - *Market-Adjusted Returns* – Ret [1, X] less the return on the CRSP value-weighted market portfolio over the same holding period.
    - *FF3 Adjusted Returns* – Ret [1, X] less the return on the Fama-French (1993) three factors over the same holding period.
    - *DGTW-Adjusted Returns* – Ret [1, X] less the equally-weighted return on a portfolio matched on size, book-to-market, and momentum over the same holding period. See Daniel et al. (1997) for a more detailed discussion of the construction of the DGTW benchmark portfolio.

A.2 Other Control Variables

- *Size* – the log of market capitalization computed as share prices times total shares outstanding at the end of the year. (Source: CRSP)
- *Market-to-Book (MTB)* – the book-to-market ratio computed as the market capitalization at the end of the calendar year scaled by the book value of equity during the calendar year. Negative values are deleted, and positive values are winsorized at the 1st and 99th percentile. (Source: CRSP/Compustat).
- *Volatility* – the standard deviation of daily returns over a week before the earnings call (Source: CRSP)
- *Leverage* –the calculation of (long-term debts + long-term liabilities)/total assets (Source: Compustat)
- *Turnover* – the average daily turnover (i.e., share volume scaled by shares outstanding) of over one quarter prior to the earnings call. (Source: CRSP)

- *Profitability* – Operating Income Before Depreciation scaled by the book value of assets. (Source: Compustat)
- *Earnings Surprise (SUE) – We* compute three versions of standardized unexpected earnings (SUE). The first version (SUE1) is based on earnings reported in Compustat and measures the surprise as the difference between the current earnings and the earnings in the same quarter of the previous year. The second version (SUE2) is computed in the same way as (SUE1), but uses earnings adjusted for special items. The main specification, SUE3, is calculated as the current quarter earnings minus earnings from the corresponding quarter a year ago and scaled by the market value one quarter prior to the earnings announcement date. Measured in percentage. (Source: IBES & CRSP)
- $Ret_{[-5,-1]}$ – the buy-and-hold return over the one to five days prior to the earnings call. (Source: CRSP)
- $Ret_{[-20,-1]}$ – the buy-and-hold return over the previous month. (Source: CRSP)
- *Momentum (MOM)* – the buy-and-hold return over the previous two to 12 months. (Source: CRSP)
- *Institutional Ownership (I. O.)* – the percentage of the firm's shares held by institutions at year end. (Source: Thomson Reuters Institutional Holdings S34)
- *Analysts' Coverage (Coverage)* – the number of unique active analysts issuing earnings forecasts for a firm during the quarter following the earnings call. (Source: IBES)
- *Sales Growth* –sales at current quarter *t* minus the sales at *t*-1 divided by the sales in quarter *t*-1. (Source: Compustat)
- *Analysts' Forecast Dispersion (DISP)* – Standard deviation of the analyst forecasts scaled by the mean of the previous one month stock price. (Source: CRSP & IBES)
- *Absolute Forecast Errors (AFE)* – Absolute value of the difference between announced earnings and the mean of the forecast earnings scaled by the mean of the previous one month stock price. (Source: IBES & CRSP)
- *Uncertainty* – the count of uncertainty words from the Loughran-McDonald Sentiment Dictionary scaled by the total words in the earnings call. (Source: Loughran-McDonald)
- *TSD* – the topic sentiment dictionary is the cosine similarity of sentiment by various topics in the earnings call's Q&A segments between the analysts' questions and the managers' answers. The sentiment measures by counting the words' frequency using Loughran-McDonald's sentiment dictionary. The alternative method is leveraging the FinBert model. (Source: Seeking Alpha)
- $Tone_{MGRS}$ – the sentiment of managers' answers in the Q&A section of the earnings call. (Source: Seeking Alpha)
- $Tone_{Anlysts}$ – the sentiment of analysts' questions in the Q&A section of the earnings call. (Source: Seeking Alpha)
- $Tone_{PRES}$ – the sentiment of the presentation section of the earnings call. (Source: Seeking Alpha)
- $Tone_{Q\&A}$ – the sentiment of the Q&A section of the earnings call. (Source: Seeking Alpha)
- # *Topics* – the number of unique topics discussed in the earnings call. (Source: Seeking Alpha)
- β_{MKT} – the market coefficient of the Fama French (1993) three factor regression on daily stock returns over one to 12 months prior to the earnings conference call date. (Source: CRSP)

- *$IVOL_{[t-X, t-1]}$ – Idisosyncratic Volatility:* the *x* days average standard deviation of residuals over window *t-x* to *t*-1. Residuals are computed using the Fama French (1993) three factor regression model on daily stock returns over one to 12 months prior to the earnings conference call date (Ang et al., 2006). (Source: CRSP)
- *FOG – Gunning Fog Index* - estimates the number of years of formal education a person needs to understand the text on the first reading. (Source: texstat)
- *Disagreement – disagreement is across all groups on Stocktwits following the method of Cookson and Neilson (2019). It is aggregated to a daily average over x trading days surrounding the earnings conference call. (Source: link)*
- *Complexity* – measures firm complexity following the method proposed by Loughran and McDonald (2023). (Source: link)

A.3 Lists of Figures

Figure 1: Flowchart of Constructing TAD

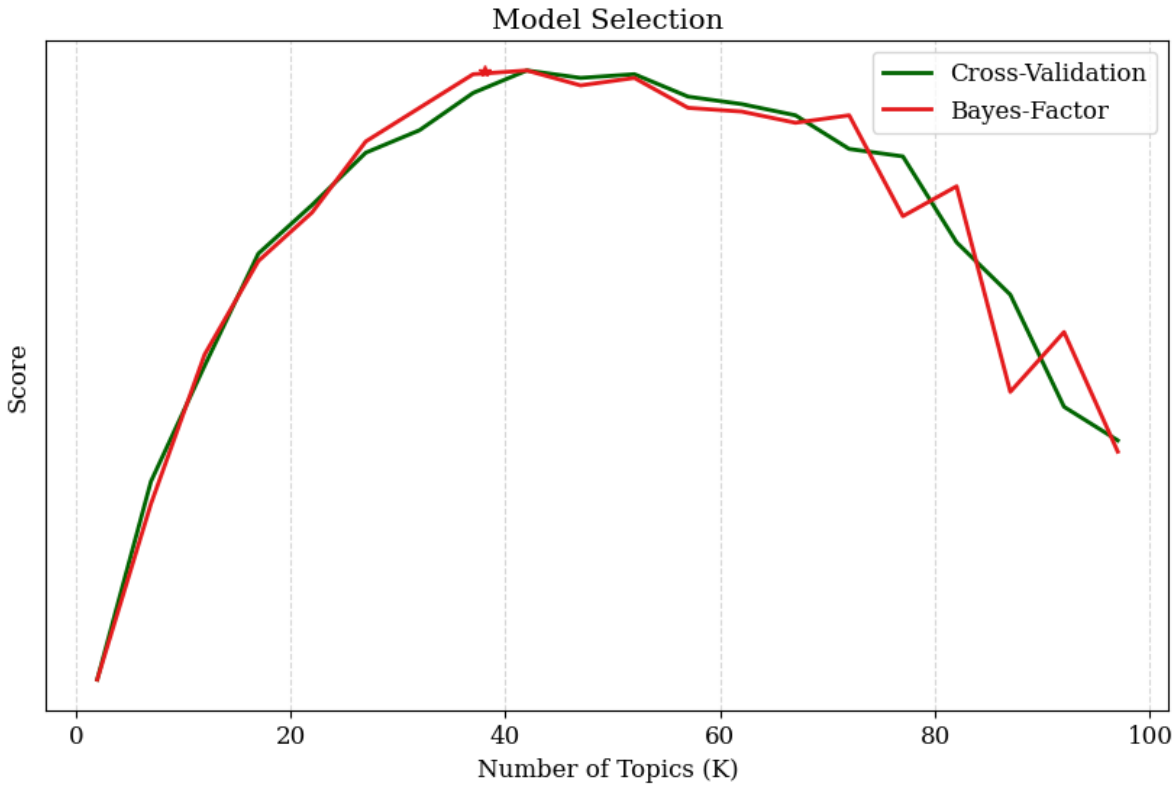


Figure 2: The Correlation Matrix of Topics in Earnings Calls



Figure 2 displays the correlation of topics generated by the model in the earnings conference call providing further visual evidence that the 37-topic model achieves a balance between interpretability and parsimony.

Figure 3: Topic Sensitivity Test of TAD

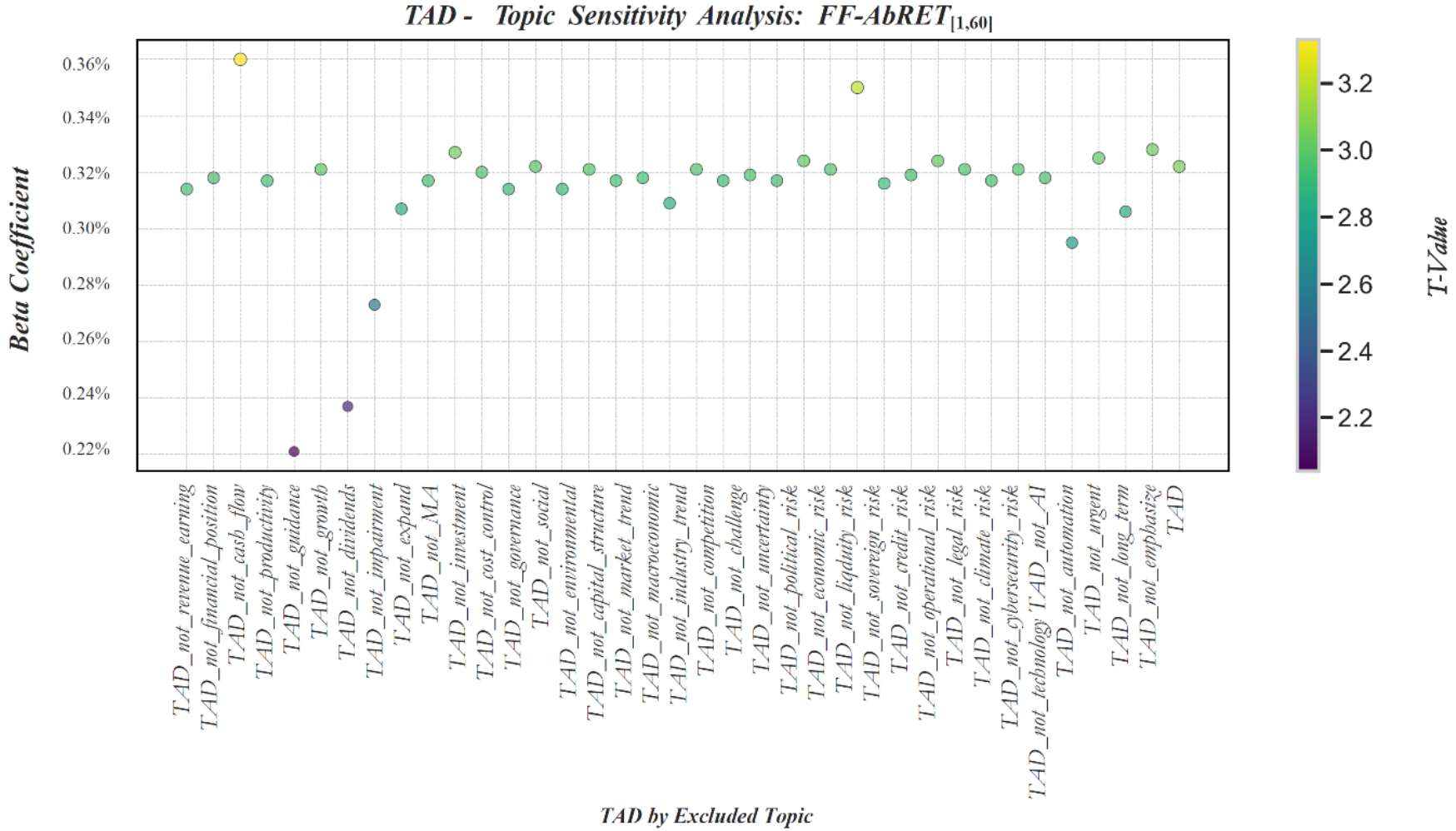


Figure *3* describes the Topic Sensitivity Test of Topic Attention Divergence (TAD)

Figure 4: The Predictability of TAD and TSD on Cross-Sectional Returns Post-Earnings

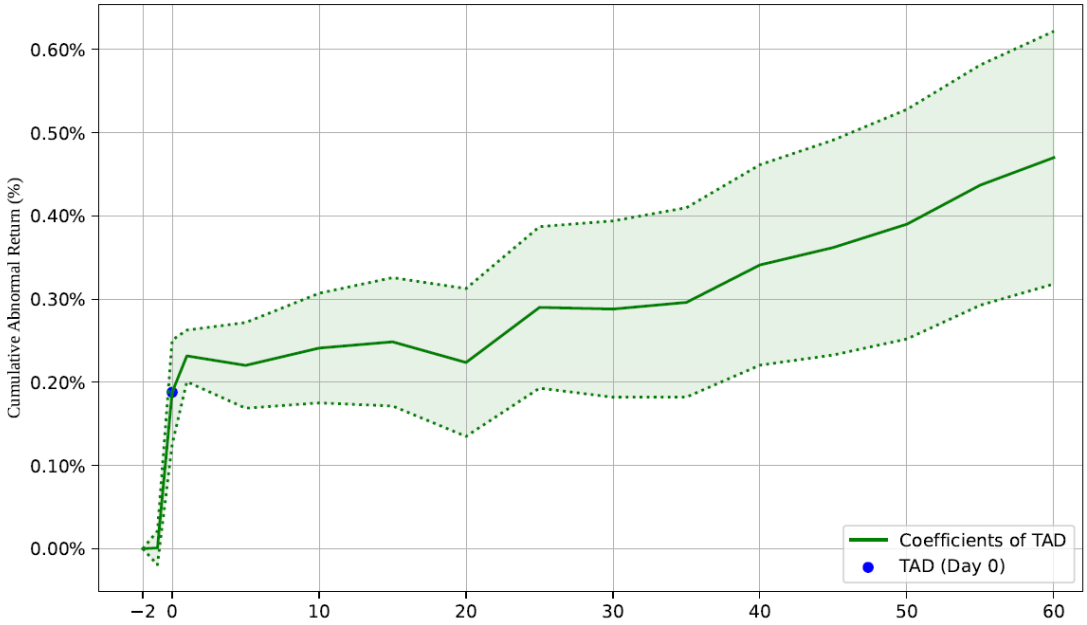


Figure 4a presents the coefficients from a series of panel regressions analyzing future stock returns, $Ret_{i[t+1, t+X]}$, as a function of the TAD and the control variables incorporating firm fixed effects and year-quarter fixed effects. Standard errors are clustered by permnos. *X* varies from 1- 60 trading days post-earnings call.

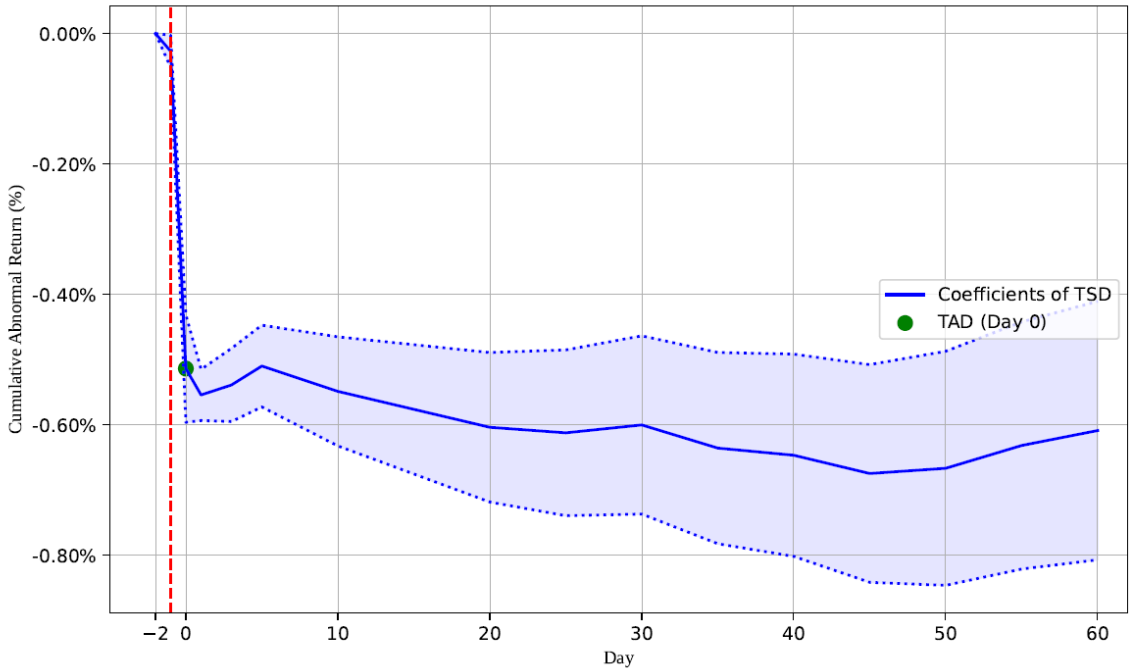


Figure 4b presents the coefficients from a series of panel regressions analyzing future stock returns, $Ret_{i[t+1, t+X]}$, as a function of TSD and the control variable, incorporating firm fixed effects and year-quarter fixed effects. Standard errors are clustered by permnos. *X* varies from 1-60 trading days post-earnings call.

A.4 Lists of Tables

Table I. Earning Conference Calls' Topic Frequencies

Table 1 presents the descriptive statistics of the top five and bottom five topics in the presentation and the Q&A sample, respectively.

	Top 5 Topics					Bottom 5 Topics				
	1st	2nd	3rd	4th	5th	1st	2nd	3rd	4th	5th
Presentation	Guidance	Growth	Investment	Cash Flow	Urgency & Emphasis	Liquidity Risk	Impairment	Market Trend	Credit Risk	Cyber-Risk
mean	19.43	9.24	8.97	8.71	7.32	0.2	0.13	0.10	0.09	0.02
std	7.42	4.05	6.17	9.39	3.52	0.78	0.49	0.37	0.52	0.16
p1	7.42	1.64	0.00	0.00	1.05	0	0	0.00	0.00	0.00
p25	5.64	6.44	4.88	0.00	4.98	0	0	0.00	0.00	0.00
p50	14.21	8.85	7.74	6.05	6.85	0	0	0.00	0.00	0.00
p75	23.68	11.56	11.64	13.58	9.07	0	0	0.00	0.00	0.00
p99	40.78	20.851	30.13	39.05	18.35	3.67	3.0	1.71	2.37	0.48
Q&A	Guidance	Urgent & Emphasis	Investment	Growth	Geo-Risk	Climate Risk	Market Trend	AI	Credit Risk	Cyber-Risk
mean	21.53	9.19	7.50	6.94	6.35	0.20	0.10	0.10	0.06	0.02
std	9.15	4.91	7.34	3.84	8.35	0.89	0.50	0.42	0.49	0.17
p1	5.57	1.46	0.00	0.41	0.00	0.00	0.00	0.00	0.00	0.00
p25	15.18	6.00	2.43	4.32	0.00	0.00	0.00	0.00	0.00	0.00
p50	20.25	8.39	5.62	6.42	3.26	0.00	0.00	0.00	0.00	0.00
p75	26.44	11.36	10.43	8.92	9.20	0.00	0.00	0.00	0.00	0.00
p99	48.92	25.46	33.02	18.48	36.35	3.80	2.25	1.84	2.02	0.43

Table II. Sample Descriptive Statistic

Sample statistics from 2006-2022 are reported. Panel A presents the distribution of firm characteristics, proxies of disagreement, and the characteristics of earnings conference calls. The sample consists of all CRSP common stocks with records of earnings conference calls. *N* is the total number of firm-quarter observations. Return is the Fama-French (1993) three factor model adjusted return. Idiosyncratic volatility (*IVOL*) is the standard deviation of residuals from a regression of daily stock returns in month *t*-1 on the Fama and French (1993) factors. *SUE* uses analyst forecasts to measure deviations in reported earnings from expectations scaled by one-month average prices. *Complexity* measures firm complexity following Loughran and McDonald (2023). β_{MKT} is the regression coefficient of the past one year of daily returns on market returns. Momentum (*MOM*) is the buy-and-hold month m-12 to m-1 return. *DISP* is the analyst forecast standard deviation scaled by the prior one month stock price. $DISP_{chg}$ uses one to five days' average analysts' forecast dispersions subtracting the average daily dispersion from -45 trading days -3 days prior to the call. *Disagreement* is within group disagreement on Stocktwits following Cookson and Neilson (2019).

Panel A: Distribution of firm characteristics

Varname	N	Mean	Std	1%	25%	50%	75%	99%
Nature of Conference Call								
TAD_{BoW}	110,882	0.240	0.146	0.037	0.130	0.207	0.319	0.692
TAD_{GPT}	110,882	0.228	0.144	0.033	0.120	0.194	0.304	0.677
$Tone_{PRES}$	110,882	0.018	0.013	-0.013	0.010	0.018	0.026	0.048
$Tone_{Q\&A}$	110,882	0.004	0.009	-0.020	-0.002	0.004	0.010	0.026
$Tone_{Analyst}$	110,882	0.001	0.133	-0.333	-0.068	0.000	0.067	0.333
$Tone_{Managers}$	110,882	0.088	0.102	-0.143	0.022	0.080	0.146	0.371
TSD	110,882	1.266	0.667	0.000	0.835	1.175	1.582	3.393
Call Length	110,882	7,287	2,259	2,238	4,909	6,840	8,670	12,967
$Uncertainty_{PRES}$	110,882	0.751	1.396	0.000	0.259	0.543	0.961	3.396
$Uncertainty_{Q\&A}$	110,882	1.159	1.160	0.000	0.488	0.906	1.517	4.994
FOG_{PRES}	110,882	12.937	1.589	9.188	11.903	12.922	13.958	16.809
$FOG_{Q\&A}$	110,882	11.359	1.692	8.047	10.247	11.131	12.243	16.465
$SMOG_{PRES}$	110,882	7.921	2.128	2.907	6.479	8.016	9.426	12.551
$SMOG_{Q\&A}$	110,882	4.797	1.573	1.613	3.647	4.700	5.902	8.467
Firm Characteristics								
Ret[-5, -1]	110,882	0.15%	5.50%	-14.36%	-2.20%	0.10%	2.39%	15.77%
Ret[0]	110,882	0.02%	7.80%	-22.60%	-3.20%	0.00%	3.30%	21.50%
Ret[1]	110,882	-0.05%	3.90%	-10.80%	-1.60%	-0.10%	1.50%	11.60%
Ret[1, 5]	110,882	0.02%	6.72%	-17.34%	-3.04%	-1.32%	2.76%	20.18%
Ret[1, 60]	110,882	-1.70%	24.70%	-62.20%	-12.60%	-1.80%	8.20%	69.40%
FF3-Ret[1,	110,882	-1.94%	20.61%	-62.20%	-12.59%	-1.84%	8.22%	69.37%
IVOL	87,647	0.007	0.082	0	0	0.001	0.005	0.072
Amihud	89,044	7.108	5.397	0.001	0.020	0.087	0.416	43.661
Bid-ask	98,125	0.193	0.006	0.000	0.000	0.001	0.002	0.020
SUE	110,882	0.0007	0.0127	-0.0678	-0.0006	0.0006	0.0026	0.0552
Log(Size)	110,882	7.903	1.882	3.717	6.636	7.881	9.106	12.493
MOM	109,882	0.046	0.685	-0.811	-0.228	-0.024	0.19	2.064
MTB	106,145	0.924	0.931	-0.999	0.305	0.828	1.419	3.727
β_{MKT}	88,231	1.052	0.383	0.174	0.815	1.027	1.266	2.104
Firm	98,152	0.48	0.184	0.173	0.344	0.453	0.586	1.031
I. O.	81,113	0.765	0.243	0.006	0.667	0.819	0.925	1.188
Profitability	110,882	0.002	0.04	-0.206	-0.001	0.008	0.019	0.083
Sales Growth	110,882	0.191	0.238	-0.604	-0.043	0.02	0.092	1.354
Coverage	110,882	4.375	4.71	0	1	3	6	21
Leverage	110,882	0.28	0.243	0	0.087	0.25	0.411	0.97
Information								
DISP	92,497	0.340	0.760	0.000	0.050	0.110	0.280	5.560
AFE	110,882	0.638	1.509	0.000	0.060	0.176	0.503	10.917

Table II - Continued

Panel B Pairwise Correlation												
Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
(1) TAD_{BoW}	1.000											
(2) TAD_{GPT}	0.639***	1.000										
	(0.000)											
(3) AFE	0.011***	0.019***	1.000									
	(0.000)	(0.000)										
(4) DISP	-0.007**	-0.002	0.591***	1.000								
	(0.040)	(0.468)	(0.000)									
(5) FOG	-0.200***	-0.209***	0.068***	0.092***	1.000							
	(0.000)	(0.000)	(0.000)	(0.000)								
(6) Uncertainty	-0.052***	-0.061***	0.051***	0.069***	0.145***	1.000						
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)							
(7) TSD	-0.169***	-0.155***	-0.025***	-0.064***	0.012***	-0.080***	1.000					
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)						
(8) Tone	-0.112***	-0.121***	-0.127***	-0.140***	0.103***	-0.268***	0.116***	1.000				
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)					
(9) Call Length	-0.205***	-0.225***	-0.143***	-0.119***	0.177***	-0.010***	0.057***	0.109***	1.000			
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)	(0.000)	(0.000)				
(10) AbLogVol	-0.038***	-0.038***	0.004	-0.064***	-0.020***	-0.032***	0.084***	0.096***	0.114***	1.000		
	(0.000)	(0.000)	(0.147)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)			
(11) Disagreement	-0.059***	-0.070***	-0.001	0.012***	0.014***	-0.040***	0.031***	0.069***	0.217***	0.275***	1.000	
	(0.000)	(0.000)	(0.753)	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)		
(12) SUE	-0.007**	-0.007**	-0.127***	-0.038***	-0.017***	-0.016***	0.008***	0.092***	0.001	0.015***	-0.010***	1.000
	(0.030)	(0.017)	(0.000)	(0.000)	(0.000)	(0.000)	(0.006)	(0.000)	(0.865)	(0.000)	(0.003)	

*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table III. TAD and Stock Illiquidity

This table examines the relationship between TAD and stock illiquidity based on firm-quarter observations between 2006 and 2022. Panel A presents the results of the effects of TAD on bid-ask spread. The dependent variable is intraday bid-ask spread computed from daily high and low prices following Corwin and Schultz (2012). In Panel A, the dependent variable is the daily average bid-ask spread across different windows ranging from the intraday of the earnings call and one week, two weeks, one month, and one quarter after the earnings call (excluding intraday). Panel B presents the results of the effects of TAD on *Amihud Illiquidity* (Amihud, 2002). The dependent variable is the daily average of *Amihud Illiquidity* over different time ranges during the post earnings call periods including intraday, one week, two weeks, one month, and one quarter's daily average illiquidity. We controlled for analyst forecast dispersion from -45 trading days to -3 days prior to the earnings call. *SUE*, standard unexpected earnings, is calculated as the most recent announced earnings minus earnings one year before scaled by stock prices following the rolling seasonal random walk model in Livnat and Mendenhall (2006). Coverage, the number of analysts within a quarter prior to earnings call. *MTB* is the market to book ratio. *Leverage* is calculated as (long-term debts + long-term liabilities)/total assets. *Size* is the natural log of total assets. *Sales growth* is computed as ($sales_t$ – $sales_{t-1}$)/$sales_{t-1}$. *Mom* is defined as the one- to 12-month cumulative abnormal returns prior to an event. #*Topic* is the number of unique topics discussed in the earnings call. *I. O.* is the institutional ownership. *TSD* is the average tone variance over all topics in the Q&A section. $Volatility_{[-5,-1]}$ is the return volatility within the one to five trading day window prior to an earnings call. Untabulated control variables include tone of managers, tone of analysts, uncertainty of managers, uncertainty of analysts; Fog score in Presentation section, Fog in Q&A section, and *IVOL, which* is the idiosyncratic volatility representing the standard deviation of residuals within one quarter prior to an event (see details in the appendix) The Huber-White heteroskedasticity-robust t-statistics, which are adjusted for clustered standard errors, are shown in parentheses below the estimated coefficients. The symbols *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively. In this table, the dependent variable has been standardized by subtracting the mean and dividing by the standard deviation to ensure comparability across different scales.

Panel A. Bid-Ask Spread

Dependent Variable:	$Illiquidity_{i,t \to t+X}$				
	Current Day	1 week	2 weeks	1 month	1 quarter
	(1)	(2)	(3)	(4)	(5)
TAD	**0.0033**	**0.0085*** **	**0.0077*** **	**0.0086*** **	**0.0075*** **
	(0.0028)	**(0.0023)**	**(0.0022)**	**(0.0022)**	**(0.0021)**
DISP	0.0546***	0.0644***	0.0662***	0.0670***	0.0636***
	(0.0058)	(0.0050)	(0.0050)	(0.0050)	(0.0048)
SUE	-0.0056	-0.0217***	-0.0247***	-0.0282***	-0.0263***
	(0.0052)	(0.0045)	(0.0044)	(0.0044)	(0.0042)
Log(Size)	-0.4083***	-0.4789***	-0.4678***	-0.4611***	-0.4521***
	(0.0157)	(0.0133)	(0.0131)	(0.0129)	(0.0127)
Coverage	0.0001	-0.0024	-0.0033	-0.0013	0.0019
	(0.0029)	(0.0022)	(0.0021)	(0.0021)	(0.0021)
MTB	-0.1550***	-0.1786***	-0.1790***	-0.1787***	-0.1763***
	(0.0064)	(0.0056)	(0.0055)	(0.0054)	(0.0051)
Leverage	0.1101***	0.1290***	0.1302***	0.1322***	0.1366***
	(0.0067)	(0.0057)	(0.0056)	(0.0055)	(0.0053)
Profitability	-0.0552***	-0.0665***	-0.0697***	-0.0707***	-0.0662***
	(0.0053)	(0.0049)	(0.0049)	(0.0048)	(0.0045)
Sales Growth	-0.0020	-0.0013	-0.0024	-0.0020	-0.0032
	(0.0029)	(0.0025)	(0.0024)	(0.0024)	(0.0022)
Mom	-0.0437***	-0.0460***	-0.0455***	-0.0452***	-0.0480***
	(0.0031)	(0.0027)	(0.0026)	(0.0026)	(0.0025)
Call Length	-0.0948***	-0.0934***	-0.1034***	-0.1045***	-0.0970***
	(0.0133)	(0.0112)	(0.0110)	(0.0109)	(0.0106)
#Topics	0.0056**	0.0084***	0.0083***	0.0077***	0.0087***
	(0.0028)	(0.0022)	(0.0022)	(0.0022)	(0.0021)
I. O.	-0.1255***	-0.1572***	-0.1590***	-0.1556***	-0.1517***
	(0.0056)	(0.0052)	(0.0051)	(0.0050)	(0.0049)
TSD	-0.0047*	0.7395***	0.7375***	0.8575***	0.8575***
	(0.0026)	(0.0218)	(0.0211)	(0.0311)	(0.0314)

$Volatility_{[-5, -1]}$	0.0432***	0.0335***	0.0333***	0.0299***	0.0293***
	(0.0045)	(0.0036)	(0.0035)	(0.0035)	(0.0033)
Controls	YES	YES	YES	YES	YES
Observations	85,542	85,542	85,542	85,542	85,542
Firm FE	YES	YES	YES	YES	YES
Year-quarter FE	YES	YES	YES	YES	YES
Adj R^2	0.725	0.739	0.747	0.765	0.725

Panel B. Amihud Illiquidity

Dependent Variable:			$Illiquidity_{i,t \rightarrow t+X}$		
	Current Day	1 week	2 weeks	1 month	1 quarter
	(1)	(2)	(3)	(4)	(5)
TAD	**0.0144*** **	**0.0144*** **	**0.0145*** **	**0.0146*** **	**0.0148*** **
	(0.0026)	**(0.0026)**	**(0.0026)**	**(0.0026)**	**(0.0026)**
DISP	0.0352***	0.0354***	0.0355***	0.0358***	0.0371***
	(0.0060)	(0.0060)	(0.0060)	(0.0060)	(0.0060)
SUE	0.0045	0.0041	0.0036	0.0030	-0.0005
	(0.0054)	(0.0054)	(0.0054)	(0.0054)	(0.0054)
Size	-0.4161***	-0.4157***	-0.4148***	-0.4152***	-0.4104***
	(0.0158)	(0.0158)	(0.0158)	(0.0158)	(0.0157)
Coverage	0.0073***	0.0072***	0.0072***	0.0071***	0.0069***
	(0.0019)	(0.0019)	(0.0019)	(0.0019)	(0.0019)
MTB	-0.1114***	-0.1121***	-0.1125***	-0.1141***	-0.1187***
	(0.0059)	(0.0059)	(0.0059)	(0.0059)	(0.0059)
Leverage	0.0879***	0.0879***	0.0878***	0.0881***	0.0896***
	(0.0062)	(0.0061)	(0.0061)	(0.0061)	(0.0061)
Profitability	-0.0260***	-0.0266***	-0.0272***	-0.0281***	-0.0313***
	(0.0055)	(0.0055)	(0.0056)	(0.0056)	(0.0056)
Sales Growth	0.0018	0.0018	0.0019	0.0017	0.0009
	(0.0027)	(0.0027)	(0.0027)	(0.0027)	(0.0026)
Mom	0.0340***	0.0320***	0.0302***	0.0269***	0.0147***
	(0.0038)	(0.0038)	(0.0038)	(0.0038)	(0.0036)
Call Length	-0.0989***	-0.0994***	-0.0996***	-0.1001***	-0.1018***
	(0.0131)	(0.0131)	(0.0132)	(0.0132)	(0.0132)
Number of Topics	0.0112***	0.0113***	0.0114***	0.0116***	0.0116***
	(0.0026)	(0.0026)	(0.0026)	(0.0026)	(0.0026)
I. O.	-0.1789***	-0.1798***	-0.1805***	-0.1823***	-0.1873***
	(0.0062)	(0.0062)	(0.0062)	(0.0062)	(0.0062)
TSD	0.0175***	0.0175***	0.0175***	0.0175***	0.0177***
	(0.0063)	(0.0063)	(0.0063)	(0.0063)	(0.0064)
$Volatility_{[-1,-5]}$	0.0168**	0.0168**	0.0169**	0.0168**	0.0168**
	(0.0079)	(0.0079)	(0.0078)	(0.0078)	(0.0077)
Controls	YES	YES	YES	YES	YES
Observations	88,468	88,468	88,468	88,468	88,468
Firm FE	YES	YES	YES	YES	YES
Year-quarter FE	YES	YES	YES	YES	YES
Adj R^2	0.689	0.689	0.689	0.690	0.693

Table IV. Inverse Information Quality Proxies with Stock Illiquidity

Table IV presents the results of regression analyses examining the cross-sectional impact of information risk (indicative of poor information quality) on stock illiquidity. This risk is represented by absolute forecast error (*AFE*), analysts' forecast dispersion (*DISP*), and topic attention divergence (*TAD*). Analysts' Forecast Dispersion (*DISP*) measures the variability in analyst forecasts normalized by the mean stock price of the preceding month. Absolute Forecast Errors (*AFE*) calculate the absolute difference between announced earnings and the average forecasted earnings, also scaled by the prior month's mean stock price. The analysis adheres to Equation (2-a) anticipating a positive association between illiquidity and the variables *AFE*, *DISP*, and *TAD*. $Illiqduity_{i,t+1\to t+n} = \beta_1 \times TAD_{i,t} + \beta_2 \times Other\ Information\ Risk_{i,t} + Controls + Firm\ FE + YearQ\ FE + \varepsilon_{it}$, Eq.$(2-\text{a})$; $Illiqduity_{i,t+1\to t+n} = \beta_1 \times Information\ Risk_{i,t} + Controls + Firm\ FE + YearQ\ FE + \varepsilon_{it}$, $Eq.(2-b)$. Control variables include *turnover*, the prior month's return ($Ret_{[-20,\,-1]}$), *volatility*, *sales growth*, *coverage*, *market-to-book ratio*, *leverage*, *profitability*, market capitalization (*Size*), and institutional ownership (*I. O.*). *Turnover* is the trading volume relative to the number of outstanding shares. Prior return denotes the stock's return from the previous month. *Volatility* is the standard deviation of daily stock returns over the past one week prior to the earnings call. *Size* refers to the log of market capitalization. The book-to-market ratio indicates the equity value in comparison to its book value, while sales growth represents the change in sales. For an in-depth explanation of these terms, refer to Appendix A. The regression in Specification (1) follows Equation (2-a), while the subsequent specifications in Table IV utilize Equation (2-b). In Panel A, the dependent variable is the intraday bid-ask spread calculated from daily high and low prices following Corwin and Schultz (2012). Panel B uses the daily average of *Amihud Illiquidity* over various periods' post-earnings call (Amihud, 2002) as the dependent variable. The Huber-White heteroskedasticity-robust t-statistics, which are adjusted for clustered standard errors, are shown in parentheses below the estimated coefficients. The symbols *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively. In this table, the dependent variable has been standardized by subtracting the mean and dividing by the standard deviation to ensure comparability across different scales.

Panel A: Bid-Ask Spread

Dependent Variable:			$Illiquidity_{i,t\to t+60}$	
	(1)	(2)	(3)	(4)
TAD	0.011***	0.012***		
	(0.002)	(0.002)		
AFE	0.0789**		0.0945***	
	(0.171)		(0.079)	
DISP	0.0538***			0.0783***
	(0.005)			(0.004)
Observations	97,365	87,187	87,581	87,581
Controls	YES	YES	YES	YES
Firm FE	YES	YES	YES	YES
Year-quarter FE	YES	YES	YES	YES
Adj R^2	0.720	0.7763	0.768	0.766

Panel B: Amihud Illiquidity

	(1)	(2)	(3)	(4)
	0.0204***	0.0215***		
	(0.003)	(0.002)		
AFE	0.0772***		0.0836***	
	(0.006)		(0.006)	
DISP	0.0171***			0.0412***
	(0.006)			(0.006)
Observations	88,439	79,452	79,807	79,807
Controls	YES	YES	YES	YES
Firm FE	YES	YES	YES	YES
Year-quarter FE	YES	YES	YES	YES
Adj R^2	0.697	0.694	0.699	0.696

Table V. Monthly Excess Returns and Risk-adjusted Portfolio alphas (%) of TAD Portfolios

This table presents the equally-weighted average future one quarter excess returns (Ret) and the Fama–French–Carhart four-factor risk-adjusted portfolio alphas (Ret_{Adj}) of stock portfolios sorted based on topic attention divergence (TAD). Panel A details Ret and Ret_{Adj} for single-sorted portfolios that are formed quarterly based on the most recent quarterly TAD. Panel B outlines Ret_{Adj} for double-sorted portfolios that are sorted monthly, initially by the previous quarter's size (Size) and subsequently by the recent quarterly TAD. Panels C and D display the equally weighted portfolios sorted by TAD. Similar to Panel A, Panel C demonstrates the monthly excess return (Ret) and risk-adjusted portfolio alphas (Ret_{Adj}) sorted by TAD. Panel D reports the value weighted double-sorted risk-adjusted alphas (Ret_{Adj}). Additionally, the differences in Ret and Ret_{Adj} between the highest and lowest TAD-sorted portfolios are provided, accompanied by Newey-West (1987) standard errors in parentheses. The analysis covers a sample period from January 2006-December 2022. The symbols *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively. The portfolio is monthly stock returns (in %).

Value-Weighted	Panel A: Single-sorted		Panel B: Double-sorted by Size and TAD				
	Ret	Ret_{Adj}	Small Size	2	3	4	Large Size
Low TAD	0.381***	0.404***	1.574***	1.147***	0.658**	0.286	0.397**
	(0.113)	(0.155)	(0.428)	(0.259)	(0.264)	(0.212)	(0.181)
2	0.513***	0.536***	1.861***	1.600***	0.954***	0.565**	0.494***
	(0.103)	(0.144)	(0.429)	(0.329)	(0.257)	(0.237)	(0.162)
3	0.746***	0.740***	1.913***	1.616***	0.941***	0.700***	0.731***
	(0.175)	(0.250)	(0.402)	(0.255)	(0.226)	(0.202)	(0.245)
4	0.712***	0.745***	2.169***	1.374***	1.049***	0.702***	0.662***
	(0.111)	(0.141)	(0.390)	(0.302)	(0.198)	(0.237)	(0.163)
High TAD	0.917***	1.064***	2.676***	1.570***	0.844**	1.066***	1.086***
	(0.177)	(0.229)	(0.301)	(0.303)	(0.366)	(0.284)	(0.278)
High-Low	**0.537****	**0.657****	**1.102****	**0.424**	**0.186**	**0.780****	**0.690****
Newey-West SE	**(0.226)**	**(0.296)**	**(0.484)**	**(0.328)**	**(0.399)**	**(0.350)**	**(0.347)**
Equal-Weighted	Panel C: Single-sorted		Panel D: Double-sorted by Size and TAD				
	Ret	Ret_{Adj}	Small Size	2	3	4	Large Size
Low TAD	-0.357*	-0.219*	-0.295	-0.252	-0.407	-0.275	-0.0354
	(0.184)	(0.116)	(0.349)	(0.255)	(0.258)	(0.226)	(0.199)
2	-0.045	0.096	0.0856	0.222	0.156	0.147	0.234
	(0.172)	(0.115)	(0.272)	(0.241)	(0.261)	(0.234)	(0.180)
3	-0.031	0.117	0.152	0.288	0.151	0.207	0.242
	(0.177)	(0.110)	(0.279)	(0.218)	(0.215)	(0.245)	(0.251)
4	0.138	0.284*	0.222	0.0226	0.341*	0.282	0.305*
	(0.208)	(0.149)	(0.321)	(0.245)	(0.199)	(0.215)	(0.178)
High TAD	0.391*	0.560***	0.754***	0.329	0.331	0.207	0.376**
	(0.220)	(0.159)	(0.273)	(0.240)	(0.357)	(0.210)	(0.171)
High-Low	**0.658*****	**0.689*****	**0.936****	**0.469***	**0.625***	**0.467**	**0.298**
Newey-West SE	**(0.221)**	**(0.221)**	**(0.394)**	**(0.248)**	**(0.319)**	**(0.297)**	**(0.239)**

Table VI. TAD-Based Portfolios: Spanning Testing Using Factor Models

This table presents the estimates of regressing trading strategies sorted by TAD on risk factors. The dependent variables include: (i) the "Long" portfolio, comprised of a value-weighted portfolio of the top quintile of topic attention divergence (TAD), is based on the TAD from all firms' conference calls within the past quarter; (ii) the "Short" portfolio, similarly, is composed of the top quintile TAD, but focused on short positions; and (iii) the "Long-Short" portfolio represents the differential returns between the long and short portfolios. The factor models incorporate various factors including the market excess return (RMRF), size (SMB), value (HML), momentum (UMD), profitability (RMW), the investment (CMA) factor, and the liquidity (LIQ) factor from Ken French's website. Using Newey-West (1987) adjusted standard errors, these are provided in parentheses beneath the estimates. The estimation period spans from January 2006-December 2022. The dependent variable is the monthly stock return in (%). The symbols *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Factors	Short	Long	L-S	Short	Long	L-S	Short	Long	L-S	Short	Long	L-S
Alpha	0.454***	0.917***	0.464***	0.470***	0.903***	0.433**	0.402***	0.768***	0.366**	0.509***	1.064***	0.555**
	(0.104)	(0.141)	(0.174)	(0.103)	(0.141)	(0.171)	(0.104)	(0.142)	(0.177)	(0.131)	(0.178)	(0.225)
RMRF	0.917***	1.003***	0.085**	0.894***	1.027***	0.133***	0.915***	1.052***	0.137***	0.911***	1.063***	0.152***
	(0.022)	(0.030)	(0.037)	(0.024)	(0.032)	(0.039)	(0.024)	(0.033)	(0.041)	(0.025)	(0.034)	(0.043)
SMB				0.092**	-0.106*	-0.198***	0.110**	-0.043	-0.154*	0.108**	-0.042	-0.149*
				(0.044)	(0.060)	(0.073)	(0.046)	(0.063)	(0.078)	(0.046)	(0.062)	(0.078)
HML				0.034	-0.017	-0.052	-0.054	-0.122**	-0.068	-0.078*	-0.136**	-0.058
				(0.033)	(0.045)	(0.054)	(0.041)	(0.056)	(0.070)	(0.044)	(0.059)	(0.075)
RMW							0.018	0.179**	0.161*	0.016	0.176**	0.160
							(0.057)	(0.078)	(0.097)	(0.057)	(0.077)	(0.097)
CMA							0.229***	0.241***	0.012	0.249***	0.257***	0.008
							(0.067)	(0.092)	(0.115)	(0.068)	(0.092)	(0.116)
MOM										-0.039	-0.023	0.016
										(0.026)	(0.035)	(0.044)
LIQ										-0.083	-0.241***	-0.158
										(0.067)	(0.091)	(0.115)
N months	204	204	204	204	204	204	204	204	204	204	204	204
Adj R^2	0.894	0.845	0.021	0.897	0.846	0.058	0.901	0.854	0.062	0.903	0.858	0.062

Table VII. The Effect of TAD on Cross-sectional Stock Returns

Panel A of Table VII presents the Fama-MacBeth (1973) quarterly cross-sectional returns and panel long-term cross-sectional cumulative returns results for the following models. $Ret_{i[t+1,t+n]} = \beta_1 TAD_{i,t} + \beta_2 Controls_{it} + \varepsilon_{it+1}$; $\frac{Ret_{i[t+1,t+m]}}{m} = \beta_1 TAD_{i,t} + \beta_2 Controls_{it} + YearQ\ FE + \varepsilon_{it+1}$ where $Ret_{i[t+1,t+n]}$ in Panel A is the future one quarter cross-sectional returns adjusted by Fama-French (1993) three factor model. $\frac{Ret_{i[t+1,t+m]}}{m}$ in Panel B is the average weekly risk adjusted return of firm *i* over the next *m* weeks and TAD is the topic attention divergence. The control variables are as follows. β_{MKT} is market beta, *Size* is market capitalization, *MTB* is the market-to-book ratio, and $IVOL_{[-5,-1]}$ is the standard deviation of daily residuals from -5 to -1 days prior to the earnings call. Daily residuals are computed from a rolling window daily Fama-French three factor (FF3) model regression (AHZX, 2006). For further details regarding control variable definitions, please see the appendix), *Amihud Illiquidity* follows the measure by Amihud (2002), $Ret_{[-5,-1]}$ is past one-week stock returns and *MOM* is past 11-month stock returns from -2 to -12 months prior to the earnings call day. *DISP* represents the analysts' forecast dispersion between -45 to -3 days prior to the earnings call. $DISP_{miss}$ is a dummy variable that is equal to one if the number of analysts' forecasts is below two. *Coverage* is the number of active financial analysts in given window from -45 to -3 days prior to the earnings conference call. Length is the log of the total number of words in an earnings call for firm *i* at time *t*. $Tone_{Mgrs}$ and $Tone_{Anlyst}$ are the sentiment (or Tone) of management and analyst in the earnings call. *FOG* is the readability score of the earnings call. Uncertainty indicates the percentage of uncertain words in the earnings call following Loughran and McDonald's dictionary. TSD measures the average value of tone variance within each topic. *#Topic* represents the number of unique topics in an earnings call. β_{MKT} is the market beta computed from FF3 regression on daily data from 12 months to one month prior to the earnings call. The standard errors are based on Newey-West (1987) standard errors. Specifications (1) and (4) examine a sample period from January 2006-December 2022. The panel presents time-series averages of the estimated slope coefficients from the above regression. Adj R^2 is the time-series average of adjusted R^2 in the cross-sectional regression. Firms denotes the time-series average of the number firms in the cross-sectional regression. The symbols *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

Panel A: Future one quarter cross-sectional returns (%) using Fama-MacBeth regression

	$Ret_{i,t\rightarrow t+60}$ (%)			
	(1)	(2)	(3)	(4)
TAD	**0.220*** **	**0.233*** **	**0.259*** **	**0.181** **
	(0.065)	**(0.071)**	**(0.082)**	**(0.085)**
MTB	-2.539***	-0.248	-0.239	-0.168
	(0.294)	(0.308)	(0.246)	(0.249)
Size	0.398	-0.148	0.265	0.045
	(0.238)	(0.327)	(0.169)	(0.274)
Leverage	0.746	0.350	0.166	0.177
	(0.469)	(0.512)	(0.336)	(0.296)
IO	-0.100	0.003	0.242*	0.260*
	(0.313)	(0.287)	(0.141)	(0.131)
IO_{miss}	-0.887*	-0.248	1.163	1.015
	(0.482)	(0.459)	(0.966)	(1.158)
DISP		-0.074	0.020	0.312
		(0.228)	(0.222)	(0.208)
$DISP_{miss}$		-0.531	-0.494	-0.384
		(-1.92)	(-1.85)	(-1.32)
SUE		2.173***	2.047***	2.512***
		(0.237)	(0.231)	(0.251)
Profitability		0.934***	0.753***	0.503***
		(0.165)	(0.163)	(0.171)
MOM		-8.317***	-8.335***	-8.625***
		(0.472)	(0.524)	(0.608)
Length			-0.765*	-0.667
			(0.444)	(0.513)
$Tone_{Mgrs}$			0.688***	0.741***
			(0.119)	(0.123)

$Tone_{Anlyts}$			1.109***	1.108***
			(0.118)	(0.129)
FOG			-0.467***	-0.447***
			(0.091)	(0.081)
Uncertainty			0.190	0.277**
			(0.118)	(0.115)
TSD			-0.212	-0.308*
			(0.181)	(0.178)
Coverage				0.063
				(0.194)
# Topics				0.147
				(0.092)
$Ret_{[-5,-1]}$				-0.897***
				(0.263)
β_{MKT}				-1.364*
				(0.796)
$IVOL_{[-5,-1]}$				-0.629**
				0.273
Amihud $Illiq_{[-5,-1]}$				-0.070
				0.396
Observations	95,621	95,236	95,227	78,622
Adj R^2	0.046	0.194	0.204	0.229

Panel B: Long-term cumulative cross-sectional returns (%) using panel regression

	$Ret_{t, t+1}$	$Ret_{t, t+2}/2$	$Ret_{t, t+3}/3$	$Ret_{t, t+4}/4$	$Ret_{t, t+8}/8$	$Ret_{t, t+12}/12$	$Ret_{t, t+26}/26$	$Ret_{t, t+52}/52$
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
TAD	**0.0696*** **	**0.0502*** **	**0.0290** **	**0.0193** **	**0.0152** **	**0.0165*** **	**0.0170*** **	**0.0159*** **
	(0.0235)	**(0.0151)**	**(0.0118)**	**(0.0084)**	**(0.0068)**	**(0.0057)**	**(0.0042)**	**(0.0032)**
SUE	0.3841***	0.2631***	0.1896***	0.1662***	0.0826***	0.0637***	0.0369***	0.0133**
	(0.0427)	(0.0283)	(0.0222)	(0.0190)	(0.0128)	(0.0107)	(0.0075)	(0.0056)
MTB	-0.4639***	-0.3785***	-0.2970***	-0.2794***	-0.2486***	-0.2355***	-0.2244***	-0.2325***
	(0.0560)	(0.0358)	(0.0282)	(0.0241)	(0.0168)	(0.0140)	(0.0101)	(0.0076)
SIZE	-1.1933***	-0.8979***	-0.7800***	-0.7296***	-0.7147***	-0.6096***	-0.5900***	-0.5710***
	(0.1232)	(0.0790)	(0.0623)	(0.0531)	(0.0367)	(0.0310)	(0.0220)	(0.0168)
Leverage	0.0597	0.0111	-0.0168	0.0121	0.0823***	0.1220***	0.1236***	0.1565***
	(0.0666)	(0.0427)	(0.0336)	(0.0287)	(0.0197)	(0.0164)	(0.0116)	(0.0088)
Profitability	0.1336***	0.0986***	0.0773***	0.0404*	0.0013	-0.0192	-0.0414***	-0.0525***
	(0.0480)	(0.0314)	(0.0252)	(0.0214)	(0.0145)	(0.0123)	(0.0086)	(0.0064)
MOM	-0.2538***	-0.1785***	-0.1456***	-0.1357***	-0.0748***	-0.0471***	-0.0571***	-0.0597***
	(0.0314)	(0.0204)	(0.0159)	(0.0137)	(0.0094)	(0.0078)	(0.0055)	(0.0042)
DISP	0.1459***	0.1646***	0.1386***	0.0992***	0.0722***	0.0852***	0.0998***	0.0643***
	(0.0529)	(0.0345)	(0.0277)	(0.0236)	(0.0165)	(0.0137)	(0.0095)	(0.0071)
$DISP_{miss}$	0.0329	-0.0013	-0.0009	-0.0026	-0.0097	-0.0251***	-0.0276***	-0.0409***
	(0.0350)	(0.0229)	(0.0180)	(0.0153)	(0.0105)	(0.0088)	(0.0063)	(0.0046)
Coverage	0.2493**	0.2346***	0.1627***	0.0996**	0.0208	0.0411	-0.0108	-0.0704***
	(0.1161)	(0.0750)	(0.0592)	(0.0503)	(0.0344)	(0.0284)	(0.0199)	(0.0147)
Length	0.2560***	0.1395***	0.0976***	0.0692***	0.0509***	0.0256***	0.0067	-0.0094**
	(0.0312)	(0.0201)	(0.0158)	(0.0134)	(0.0091)	(0.0076)	(0.0054)	(0.0040)
$Tone_{Mgrs}$	0.2084***	0.1485***	0.1148***	0.0987***	0.0597***	0.0465***	0.0277***	0.0079**
	(0.0296)	(0.0191)	(0.0150)	(0.0128)	(0.0087)	(0.0071)	(0.0051)	(0.0038)
$Tone_{Anlyst}$	0.0972***	0.0649***	0.0439***	0.0446***	0.0332***	0.0313***	0.0125**	0.0057*
	(0.0326)	(0.0210)	(0.0164)	(0.0140)	(0.0095)	(0.0079)	(0.0056)	(0.0031)
FOG	-0.0454	0.0045	-0.0012	0.0073	0.0159*	0.0185**	0.0195***	0.0123***

	(0.0289)	(0.0188)	(0.0149)	(0.0127)	(0.0086)	(0.0072)	(0.0050)	(0.0037)
Uncertainty	0.0251	0.0030	-0.0065	-0.0083	0.0109	0.0056	0.0073	0.0051
	(0.0265)	(0.0171)	(0.0134)	(0.0114)	(0.0077)	(0.0064)	(0.0045)	(0.0033)
TSD	0.1093	0.0492	0.1015**	0.0784*	0.0760***	0.0260	0.0139	0.0266**
	(0.0982)	(0.0620)	(0.0491)	(0.0417)	(0.0281)	(0.0232)	(0.0166)	(0.0121)
#Topics	0.0284	0.0279*	0.0293**	0.0167	0.0055	0.0081	0.0020	0.0010
	(0.0255)	(0.0165)	(0.0130)	(0.0111)	(0.0075)	(0.0063)	(0.0044)	(0.0032)
Ret[-5, -1]	-1.9672***	-1.5594***	-0.8772***	-0.9389***	-0.1369	-0.1805	-0.1721*	-0.1923**
	(0.6221)	(0.3727)	(0.2909)	(0.2477)	(0.1747)	(0.1480)	(0.1045)	(0.0814)
β_{MKT}	0.1838*	0.0790	0.0257	0.0236	0.0313	0.0190	0.0122	0.0248**
	(0.0975)	(0.0636)	(0.0496)	(0.0420)	(0.0286)	(0.0236)	(0.0170)	(0.0126)
I. O.	0.0300	0.0736**	0.0734***	0.0620***	0.0310**	-0.0016	-0.0167**	-0.0262***
	(0.0452)	(0.0297)	(0.0233)	(0.0198)	(0.0136)	(0.0114)	(0.0081)	(0.0060)
$I. O._{Miss}$	0.5927***	0.4164***	0.3709***	0.3549***	0.1246***	0.0604***	0.0227	0.1005***
	(0.0842)	(0.0560)	(0.0442)	(0.0376)	(0.0251)	(0.0206)	(0.0142)	(0.0111)
$IVOL_{[-5,-1]}$	0.1587***	0.1083***	0.0930***	0.1423***	0.0919***	0.0876***	0.0962***	0.0937***
	(0.0490)	(0.0313)	(0.0247)	(0.0214)	(0.0143)	(0.0120)	(0.0086)	(0.0064)
Amihud	-0.0211	0.0372	0.0428	0.0566**	0.0661***	0.0857***	0.0775***	0.0562***
	(0.0548)	(0.0364)	(0.0286)	(0.0241)	(0.0162)	(0.0139)	(0.0103)	(0.0078)
Observations	78,635	78,635	78,635	78,635	78,635	78,593	76,534	78,635
Year-Qtr FE	YES	YES	YES	YES	YES	YES	YES	YES
Adj R^2	0.029	0.040	0.045	0.042	0.048	0.054	0.048	0.029

Table VIII. Cross Sectional Test of Future Stock Return on TAD by of Information Processing Costs

Table VIII presents the results of regression analyses examining the cross-sectional impact of information risk (indicative of poor information quality) on future stock return by information processing costs. The classification of firms into high and low information processing groups following work (Kim and Liu, 2018), which claims firms with high institutional ownership have lower level of information asymmetry and informed trading behaviours due to reducing information advantage of corporate insiders over outsiders. My analysis reveals a significant relationship between TAD and future cross-section risk adjusted returns ($Ret_{i,[t+1,\ t+60]}$), particularly pronounced in firms characterized by higher information processing costs. Specifically, Table VIII divides our sample based on institutional ownership (Specifications 1 and 2), financial analysts' coverage (Specifications 3 and 4), and market capitalization (Specification 5 and 6), using the median as a benchmark for classification. Control variables include *turnover*, the prior week's return ($Ret_{i,[-5,\ -1]}$), *volatility*, *sales growth*, *analysts' coverage*, *market-to-book ratio*, *leverage*, *profitability*, *size*, *MOM*, and *I. O. Turnover* is the trading volume relative to the number of outstanding shares. Prior return denotes the stock's return from the previous month. *Volatility* is the standard deviation of daily stock returns over one week prior to earnings calls. *Size* refers to the log of market capitalization. The *book-to-market ratio* indicates the equity value in comparison to its book value, while *sales growth* represents the change in sales. For an in-depth explanation of these terms, refer to Appendix A. The Huber-White heteroscedasticity-robust t-statistics, which are adjusted for clustered standard errors, are shown in parentheses below the estimated coefficients. The symbols *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

Dependent Variable	$Ret_{i,t\rightarrow t+60}$ (%)					
Sub-sample:	Low IO	High IO	Low Coverage	High Coverage	Low Size	High Size
	(1)	(2)	(3)	(4)	(5)	(6)
TAD	**0.428*** **	**0.005**	**0.239****	**0.132**	**0.329*** **	**0.060**
	(0.121)	**(0.090)**	**(0.105)**	**(0.098)**	**(0.121)**	**(0.081)**
DISP	0.280	0.755***	0.561***	0.371	0.255	0.946***
	(0.249)	(0.224)	(0.203)	(0.286)	(0.208)	(0.257)
SUE	2.514***	2.024***	2.093***	2.609***	2.307***	2.042***
	(0.192)	(0.136)	(0.135)	(0.215)	(0.142)	(0.178)
MTB	-3.160***	-3.497***	-3.546***	-2.780***	-3.977***	-2.647***
	(0.315)	(0.202)	(0.246)	(0.220)	(0.258)	(0.202)
Size	-11.017***	-10.130***	-10.353***	-8.347***	-11.136***	-7.449***
	(0.920)	(0.541)	(0.670)	(0.584)	(0.760)	(0.560)
Leverage	1.884***	1.745***	2.019***	1.537***	2.240***	1.369***
	(0.417)	(0.247)	(0.290)	(0.286)	(0.292)	(0.280)
Profitability	-0.054	0.343*	0.283	0.295	0.281	0.160
	(0.260)	(0.193)	(0.209)	(0.226)	(0.192)	(0.246)
Mom	-8.508***	-7.702***	-8.006***	-7.302***	-8.246***	-6.975***
	(0.180)	(0.117)	(0.132)	(0.139)	(0.128)	(0.137)
IO	-0.163	-0.239	-0.393*	0.326	-0.223	0.172

	(0.286)	(0.460)	(0.216)	(0.209)	(0.235)	(0.172)
Observations	31,348	51,243	44,835	37,419	40,867	41,981
Controls	YES	YES	YES	YES	YES	YES
Firm FE	YES	YES	YES	YES	YES	YES
Year-quarter FE	YES	YES	YES	YES	YES	YES
Adj R^2	0.217	0.190	0.199	0.186	0.203	0.157

Internet Appendix A

In this appendix, We discuss and tabulate results from select robustness tests referenced in the paper. The set of figures and tables are as follows:

- Table IA1. Table I Cross-Sectional Stock Return on SUE and TAD (Robustness)
- Figure IA1 Flowchart of Constructing TAD
- Figure IA2 Samples of Earnings Conference Call and Topic Attention by Different Segments
- Figure IA3 Trend of TAD By Industry

Table IA.1. Double Sorted Cross-Sectional Stock Return on SUE and TAD

Intraday Risk Adjusted Return

TAD

SUE	1(L)	2	3	4	5(H)	**H-L**	**p-Value**
1	-3.752%	-3.206%	-3.309%	-3.317%	-3.081%	**0.671%*** **	0.00%
2	-1.711%	-1.466%	-1.368%	-1.558%	-1.199%	**0.513%*** **	0.00%
3	0.028%	0.280%	0.189%	0.385%	0.620%	**0.592%*** **	0.00%
4	1.292%	1.555%	1.582%	1.703%	1.583%	**0.291%** **	4.50%
5	2.660%	3.274%	2.914%	3.383%	3.049%	**0.389%** **	4.90%

1 Quarter Buy-and-Hold Risk Adjusted Return

TAD

SUE	1(L)	2	3	4	5(H)	**H-L**	**p-Value**
1	-4.877%	-4.698%	-5.047%	-4.742%	-3.787%	**1.090%***	7.60%
2	-4.251%	-3.986%	-3.493%	-3.430%	-3.057%	**1.193%*** **	0.20%
3	-2.483%	-2.276%	-1.715%	-1.988%	-1.293%	**1.190%*** **	0.00%
4	-1.860%	-1.419%	-0.790%	-0.433%	-0.232%	**1.629%*** **	0.00%
5	1.568%	2.180%	2.843%	2.380%	3.744%	**2.175%*** **	0.20%

Panel C: Double Sorted Cross-Sectional Risk Adjusted Stock Return on Size and TAD

Intraday Risk Adjusted Return

TAD

SIZE	1(L)	2	3	4	5(H)	**H-L**	**p-Value**
1	-0.887%	0.023%	-0.117%	0.200%	-0.041%	**0.846%*** **	0.00%
2	-0.090%	0.313%	0.319%	0.072%	0.361%	**0.450%*** **	0.10%
3	-0.234%	0.125%	-0.051%	0.086%	0.117%	**0.351%** **	2.30%
4	-0.074%	0.190%	-0.011%	0.197%	0.163%	**0.237%***	7.90%
5	-0.179%	-0.069%	-0.092%	0.106%	0.066%	**0.246%** **	1.40%

1 Quarter Buy-and-Hold Risk Adjusted Return

TAD

SIZE	1(L)	2	3	4	5(H)	**H-L**	**p-Value**
1	-4.572%	-3.770%	-2.747%	-2.701%	-1.530%	**3.042%*** **	0.00%
2	-2.275%	-1.996%	-1.990%	-2.236%	-0.796%	**1.480%*** **	0.60%
3	-2.058%	-1.521%	-1.196%	-1.220%	-0.765%	**0.80%** **	0.30%
4	-1.750%	-1.412%	-1.290%	-1.176%	-1.117%	**0.439%**	14.70%
5	-1.334%	-1.358%	-0.835%	-0.821%	-0.843%	**0.312%**	11.60%

Figure IA1. Flowchart of Constructing TAD

Figure IA1 depicts the data pipeline for converting real-time earnings conference calls into a TAD index, which measures the information quality of voluntary disclosures. This can enhance market efficiency by providing information in a more timely manner and reducing the information advantage of corporate insiders vs. outsiders.

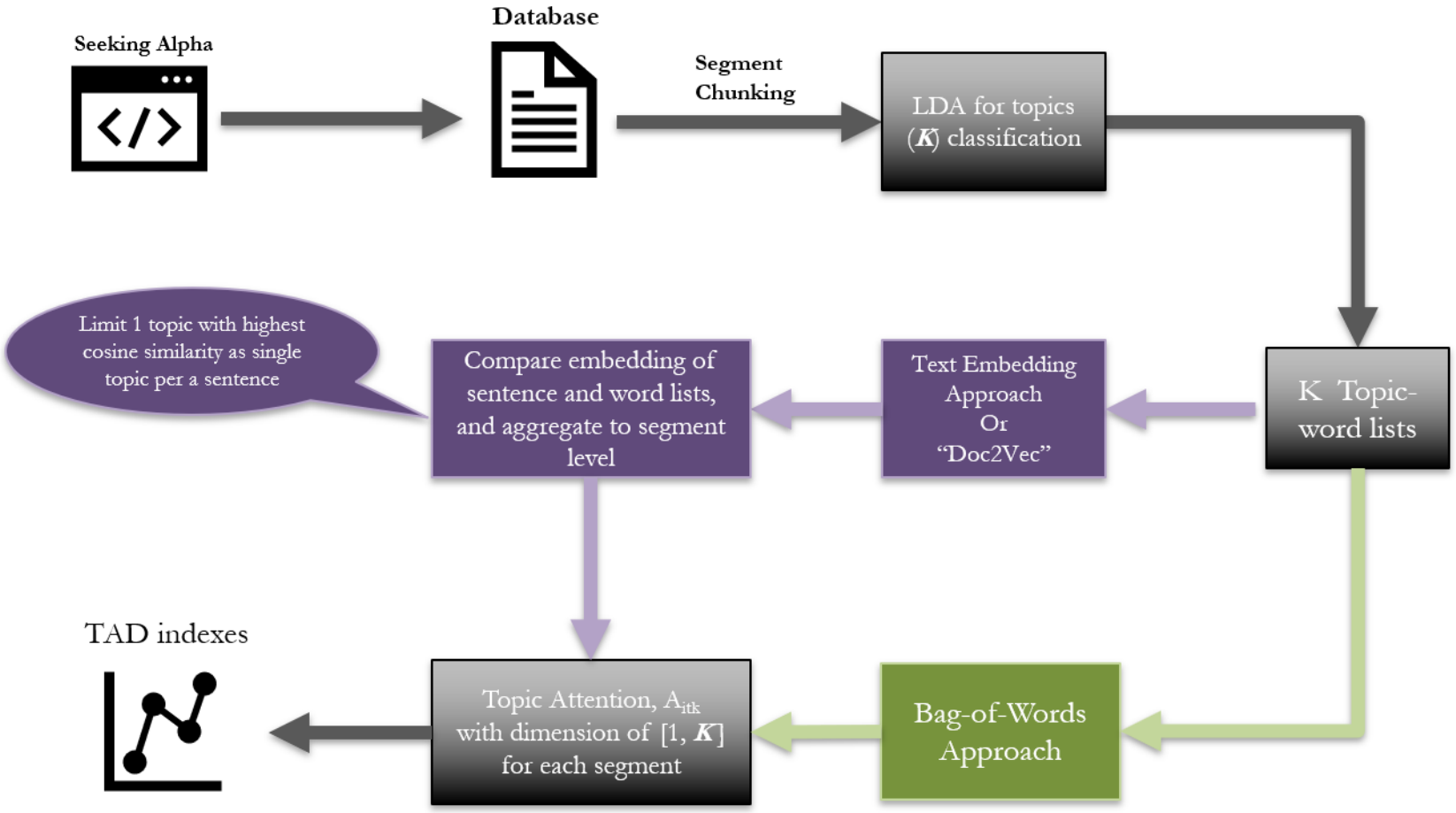


Figure IA2. Samples of Earnings Conference Call and Topic Attention by Different Segments

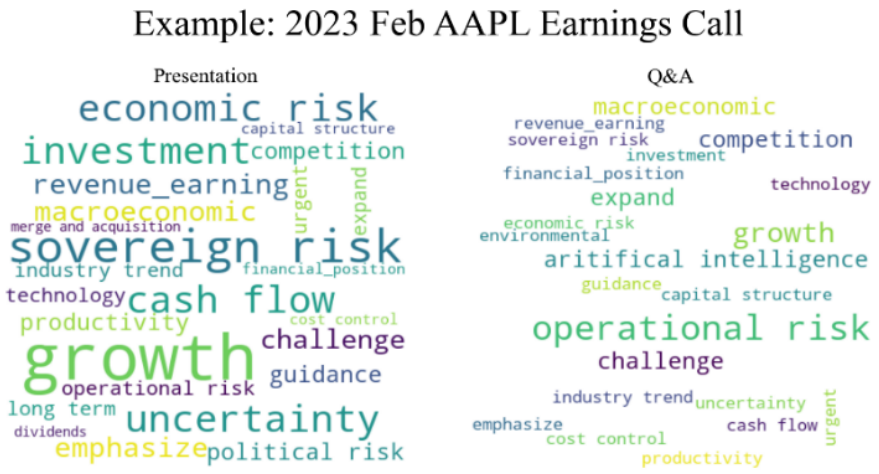


Figure 4a. presents an example of the Apple Earnings Call in February 2023. It demonstrates how the focus of management's disclosures differs from the information content that investors demand.

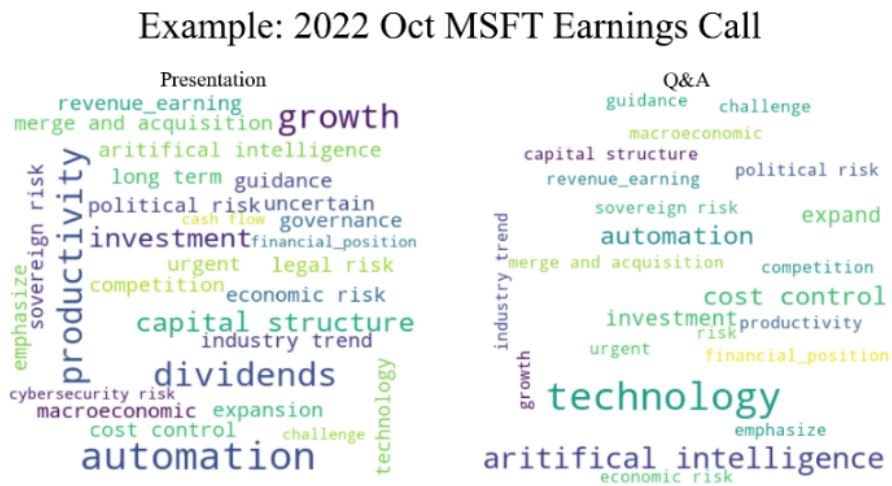


Figure 4b. illustrates the example of the Microsoft Earnings Call in October 2022. It shows the topic divergence between managers and market investors.

Summary of AAPL 2023 Jan, Earnings Call

In January 2023, Apple CEO Tim Cook began the earnings call by providing an overview of the company's financial performance for the quarter. He highlighted strong revenue growth, driven by the iPhone and services, as well as strong margins and a healthy balance sheet. Cook acknowledged that Apple had been facing supply chain challenges, but he said that the company was working to mitigate them.

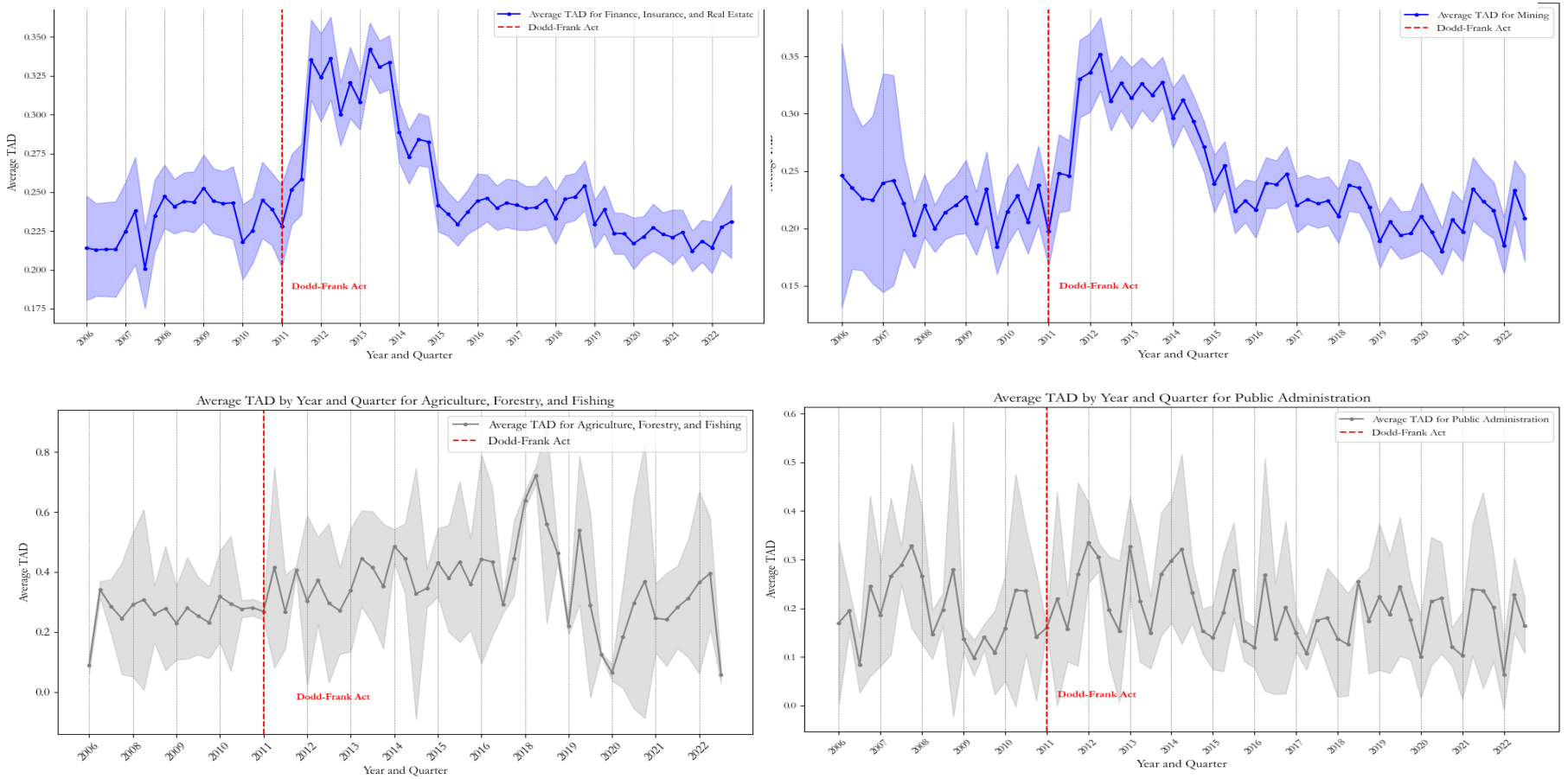
In the Q&A section, Cook fielded questions from analysts about Apple's supply chain challenges, the new features and improvements in the next generation of iPhones, and the competitors and challenges in the industry. Cook reiterated that Apple was working to mitigate supply chain challenges, and he provided details about the new features and improvements in next generation of iPhones. He also acknowledged the challenges posed by competitors, but he expressed confidence in Apple's ability to continue to innovate and grow.

Summary of MSFT 2022 Feb Earnings Call

In October 2022, Microsoft CEO Satya Nadella provided an overview of Microsoft's financial performance for the quarter. He then took questions from analysts. One analyst asked about Microsoft's cloud computing business, which has been growing rapidly in recent years. Nadella said that Microsoft is committed to its cloud computing business and that the company is investing heavily in new features and services for Azure. He also said that Microsoft is working to expand its cloud computing business into new markets, such as China and India. Another analyst asked about Microsoft's plans to compete with Amazon Web Services. Nadella said that Microsoft is confident in its ability to compete with Amazon Web Services and that the company is focused on delivering high-quality products and services to its customers. He also said that Microsoft is working to make Azure more affordable and accessible to businesses of all sizes.

Figure IA3. TAD Index by Industry (Sample Period: 2006-2022)

Figure IA3 presents the average TAD index by SIC 1-digit code. The TAD index shows a significant increase at the end of 2010 implying that the Dodd-Frank Act has impacted corporate voluntary disclosure. In particular, the Mining and Finance industries show a sudden increase in TAD compared to that of Public Administration and Agriculture. This might imply that the Dodd-Frank Act has a more pronounced effect on the disclosure quality of specific industries.



Internet Appendix B. Dataset Construction Detail

This section introduces the detailed data processing process including three parts.

- IB.1. Merge earnings call transcript with the other database
- IB.2 Earnings transcript text pre-processing
- IB.3 Identify the earnings call date and time

IB.1 Merge earnings call transcript with the other database

I construct the corpus of earnings conference call transcripts using the following process. A.1 Earnings conference call transcripts are hand collected from Seeking Alpha. We start with all conference call transcripts held between 2000 and 2022 from Seeking Alpha. N = 213,380. We followed the process used by Hassan et al. (2019) to process earnings transcripts. In the process, our sample loses:

- 3,896 transcripts because they either have a missing or incorrect earnings announcement date.
- 26,206 transcripts because they are either not earnings conference transcripts or are duplicate transcripts.
- 19,154 transcripts because we could not reliably fuzzy name match them to a company name in Compustat by using iclink.
- 1,282 transcripts because they are shorter than 50 words
- 33,143 transcripts because they could not merge or link with I/B/E/S and Compustat.
- 5,093 transcripts because they could not merge with CRSP.

- 13,724 transcripts after excluding firms with only four quarters or fewer. The dataset was refined to 110,882 rows with 5,092 unique permnos spanning from the first quarter of 2006 to the fourth quarter of 2022.

IB.2 Earnings transcript text pre-processing

Following Hassan et al. (2019), we made some modifications to the transcripts, such as changing "Bill" to "bbill," "Constitution" to "cconstitution," and removing "risk officer" and "risk credit officer" to avoid the synonyms "risk" and "credit" from catching these person/positions. We also removed "unknown speaker." "unknown participant," "unknown caller," and "unknown firm analyst" to avoid the synonym "unknown" from catching these persons. We removed "in the states."

In addition, we removed 17,750 "safe harbor" snippets from the transcripts. A safe harbor snippet is a snippet that contains two or more of the following keywords: safe, harbor, forward-looking, forward, looking, actual, statements, statement, risk, risks, uncertainty, uncertainties, future, events, sec, and results. If a snippet from the first half of the transcript contains two or more of these keywords or if it contains the word "forward looking" and fewer than two of these keywords, then the snippet is removed. Safe harbor statements are formulaic legal language that are used at the beginning of earnings calls to remind participants that forward-looking information disclosed in the call will not be considered fraudulent unless it is made in bad faith or without reasonable basis.

Using a part-of-speech tagger, we also removed all of the following bigrams from the training libraries:

- Pronoun-pronoun bigrams, such as "hers herself" and "him himself."
- Preposition-preposition bigrams, such as "among upon" and "whether out."
- Adverb-adverb bigrams, such as "occasionally unabatingly" and "maddeningly adventurously."

- Wh-adverb-adverb bigrams, such as "how however" and "whence whither."
- These bigrams were removed because they are often not informative and can be noise in the training data.

IB.3 Identify the earnings call date and time

Earnings calls can be held before or after the market closes. This affects which daily return should be used to measure the performance of the stock on the day of the earnings call. For calls that happen in the morning (Before 9:00 a.m.) of day t (the day of the earnings call), the return on day t (the price change between the market close on day t+1 and the market close on day t) is the most appropriate measure. However, if the call happens after the market closes (After 4:00 p.m.) on day t, then the current day of earnings call is t+1 (the price change between the market close on day t+1 and the market open on day t+1) is the more accurate measure.

Internet Appendix C. World List by Topics

Metatopic	Topic	Key Words
Company Performance	**Bottom line**	bottom line growth, bottom line, COGS, core business, cost of goods sold, earnings driver, earnings growth, earnings per share growth rate, earnings per share growth, earnings per share, earnings, eps growth, EPS, gross margin, gross profit, gross revenue, gross sales, inventory adjustment, inventory turnover, inventory, net income growth, net income, net profit growth, net profit margin, net profit, net sales, operating income, operating margin growth, opex, operating margin, profit growth, profit margin growth, profit margin, profitability ratio growth, profitability ratio, revenue growth, sales growth, sales turnover, top line, total revenue.
	Financial Position	accounts payable turnover, accounts payable, accounts receivable turnover, accounts receivable, capital structure, cash conversion cycle, cash equivalents, cash reserves, credit rating, current ratio, debt capacity, debt ratio, debt to equity ratio, debt-to-equity ratio, default risk, financial distress, financial leverage, financial risk, financial stability, financial strength, financial sustainability, financial viability, financial vulnerability, intangible asset, interest coverage, inventory turnover, inventory, leverage, liquidity, refinance, solvency, working capital
	Cash Flow	cash balance, cash burn rate, cash conversion cycle, cash flow, cash generation, cash position, cash reserves, free cash flow, negative cash flow, net cash flow, operating cash flow, operating cash, positive cash flow
	Productivity	automation, capacity utilization, capacity, cost cutting, cost efficiency, cost reduction, cost saving, digital transformation, efficiency, labor cost, labor efficiency, labor layoff, labor productivity, laid off, lay off, layoff, more effective, more efficient, more productive, more timely, output employee, output per unit of input, output worker, outsource, outsourced, outsourcing, process improvement, production, productivity, productivity growth, productivity improvement, productivity ratio, resource allocation, sales square foot, technology adoption, timely manner, workflow management, workflow, workforce management, workforce optimization, workforce reduction, workforce restructuring
	Guidance	anticipate result, anticipate that, anticipated future, assume, believe, bottom line guidance, continue, earnings guidance, estimate, expectation, expected EPS, expected growth, financial guidance, forecast that, forecasting, forward guidance, forward looking, forward outlook, forward projection, future, going forward, guidance forecast, guidance outlook, guidance projection, guidance range, guidance statement, market expansion, predict that, predict to, predictability, product development pipeline, project that, projecting, projection, revenue guidance, top line guidance
	Growth	acquisition, bottom line growth, capacity expansion, cash flow growth, client base growth, client growth, core growth, customer base growth, customer growth, double digit, double-digit, earning growth, earnings growth, earnings per share growth rate, earnings per share growth, earnings-growth, eps growth, eps-growth, expand, expansion, grew, growth forecast, growth opportunity, growth outlook, growth plan, growth potential, growth rate, growth strategy, improvement, increase, last quarter, last year, growth, muted growth, net income growth, net profit growth, of growth, operating margin growth, opportunity, organic growth, past quarter, growth, productivity growth, profit growth, profit margin growth, ratio growth, quarter-over-quarter, revenue growth, revenue-growth, sales growth, single digit, single-digit, solid growth, strength, strong growth, sustainable growth, sustained performance, tailwind, triple-digit, upward trajectory, user base growth, user growth, versus prior quarter, prior year, growth, year-over-year

	Dividends	dividend, repurchase, share repurchase, share buyback, share split, stock split, stock dividend, stock buyback, stock repurchase, stock split
	Impairment	impairment of goodwill, impairment of intangible asset, impairment of long-lived asset, impairment of property, plant and equipment, impairment of property and equipment, impairment of property, impaired, write off, write-off, impairment, impairment loss, goodwill
Strategy and Operation	**Expansion**	acquisition, advertisement, advertising, brand development, brand strategy, brand, branding strategy, branding, business diversification, business expansion, capital investment, competitive positioning, customer acquisition, expansion strategy, game changer, game changing, geographic expansion, global expansion, goals, growth initiative, growth strategy, headcount, hiring, implementation, initiative, innovation cost, innovation, inorganic growth, international expansion, joint ventures, long term, long-term strategy, long-term, management plan, management strategy, market development, market expansion, market penetration, market share growth, merger and acquisition, mergers, mid term, mid-term strategy, mid-term, milestone, new branch, new business, new customer, new market entry, new product, nimble, objective, organic growth, partnership and alliance, partnership development, partnership, partnerships, portfolio expansion, priority, process improvement, product development, product diversification, product expansion, product launch, R&D, scaling operations, short term, short-term strategy, short-term, strategic alliances, strategic corporate, strategic direction, strategic focus, strategic framework, strategic goals, strategic initiative, strategic investment, strategic objectives, strategic outlook, strategic plan, strategic planning, strategic priority, strategic review, strategic roadmap, strategic vision, strategic, strategy, tactic, target, timely basis, utilization rate, vision
	Investment	advertisement, advertising, amortization, brand, branding strategy, branding, capex, capital allocation, capital efficiency, capital expenditure, customer acquisition cost, customer lifetime value, depreciation and amortization, depreciation expense, depreciation, invest, investing activities, investing cash flow, investing, investment, marketing investment, net promoter score, NPV, product launch, r&d expense, r&d spending, reinvestment, research and development spending, research and development, return on assets, return on equity, return on investment, ROI
	Cost Control	cost control, cost cutting, cost efficiency, cost management, cost minimization, cost optimization, cost reduction, cost saving, decrease in investment, digital transformation, direction, downsize, efficient, execution, exit unattractive, expenditure reduction, expense control, expense management, expense reduction, focus, headcount management, headcount optimization, headcount reduction, headcount freeze, hiring freeze, labor cost, labor efficiency, labor management, labor optimization, lean operate, lower spend, operating margin, outsource, overhead reduction, production cost, reduce cost, reduce expenditure, reduce expense, reduce investment, reduce marketing, reduce overhead, reduce R&D, reduce SG&A, reduce spending, reducing capital expenditure, restructure, restructuring cost, roadmap, robotic process automation, RPA, spend control, streamline, supply chain optimization, utilization
ES&G	**Environment**	biodiversity, carbon emissions, carbon footprint, carbon neutrality, carbon offset, carbon reduction, carbon tax, carbon trading, clean energy, climate change, climate risk, eco-friendly, energy efficiency, environmental impact, environmental sustainability, ESG, global warning, green energy, green product, green technology, greenhouse gas, greenhouse, low-carbon, natural resource, ozone layer, pollution control, pollution, sustainability, sustainable
	Social	community, corporate culture, culture diversity, gender diversity, gender equality, gender inequality, income inequality, job creation, LGBT, social impact, social responsibility, social sustainable, sustainable development, workforce diversity
	Governance	bonus plans, conservatism, conservative, corporate governance, corporate culture, discipline, effective, efficient, employee engagement, employee productivity, employee retention, employee satisfaction, employee stock option, employee turnover, equity

		compensation, executive compensation, gender inequality, incentive compensation, incentive plan, income inequality, insider, LGBT, optimal, organizational culture, organizational structure, performance evaluation, performance-based compensation, predictability, predictable, race equality, shareholder activis, stock option plan, stock option program, stock option, succession planning, talent acquisition, talent development, talent management, talent retention, transparent, turnover, workforce diversity
Capital Structure	**Capital Structure**	bond rating, capital market condition, capital structure management, capital structure optimization, capital structure, capitalization, cost debt, cost equity, credit rating, debt capacity, debt covenants, debt financing, debt maturity, debt restructure, debt restructuring, debt-to-equity ratio, dividend policy, equity financing, financial flexibility, financing plan, financing strategy, financing, interest coverage ratio, interest coverage, interest expense, interest rate risk, interest rate sensitivity, interest rate, leverage, long-term debt, merge and acquisition, share buyback, share repurchase, short-term debt, stock buyback, stock repurchase, strategic acquisition, strategic alliance, strategic investment, strategic merger, strategic partnership, WACC, weighted average cost of capital
Trend	**Market Trend**	industry contraction, industry expansion, industry forecast, industry growth forecast, industry growth outlook, industry growth projection, industry growth trend, industry growth, industry outlook, industry projection, industry trend, investor sentiment, market contraction, market expansion, market forecast, market growth, market outlook, market projection, market trend, sector forecast, sector growth, sector outlook, sector projection, sector trend
	Macro Economics	bernanke, biden, china economy, the china, composite pmi, consumption, construction pmi, consumer confidence, consumer price index, consumer sentiment, consumer spending, core consumer price index, core cpi, core ppi, core producer price index, corona, coronavirus, covid, CPI, difficult time, economic crisis, economic growth, economic sanctions, economy contraction, economy crisis, economy development, economy expansion, economy recession, economy, emerging market, energy price, FED, federal reserve, financial crisis, fiscal policy, foreign exchange, GDP, global economic condition, government spend, government spending, government stimulus, inflation, interest rate, jobs data, jobs report, labor market, lockdown, macro, manufacturing pmi, market demand, market fear, market sentiment, market uncertainty, market volatility, monetary policy, obama, pandemic, PCE, PMI, powell, PPI, producer price index, purchasing managers index, recession, recovery, remote work, retail sales growth, retail sales, services PMI, shutdown, soft demand, supply chain, trade agreement, trade deal, trade dispute, trade policy, trade sanctions, trade tariff, trade war, trump, unemployment, yellen
	Industry Trend	bbill, commodity price, competitive advantage, competitive environment, competitive landscape, competitive position, competitive risk, competitive threat, consumer growth, consumer trend, consumption growth, consumption trend, consumption, consolidation, consumer behavior, consumer confidence, consumer preference, consumer sentiment, consumer spending, consumer trend, govern subsidy, government bill, government contract, government subsidy, heat up, heats up, industry contraction, industry expansion, industry forecast, industry growth trend, industry growth, industry outlook, industry projection, industry trend, innovation, innovative approach, innovative business, innovative manner, innovative product, innovative service, innovative solution, innovative technology, market demand, market supply, market trend, PPI, price competition, price trend, price war, price, producer price index, product innovation, R&D, subsidy war, supply chain, supply-chain, sustainability, technology development, technology innovation, technology trends
Competition & Challenge	**Competition**	adjust strategy, barrier entry, big shift, brand awareness, brand loyalty, brand recognition, catch up, challenge, competition, competitive advantage, competitive edge, competitive environment, competitive intensity, competitive landscape, competitive position, competitive risk, competitive threat, competitor, consolidation, continuous innovation, continuous optimization, continuous transformation, customer loyalty, customer retention, customer satisfaction, differentiate, differentiation, difficult

		environment, difficult position, difficult situation, dynamic environment, dynamic industry, dynamic market, grow change, headwind, heat up, heats up, price competition, price war, resilience, rival, rivalry, threat, tough decision
	Challenge	adjustment period, back-end loaded, be behind, bear the brunt, bear the burden, behind the curve, behind the eight ball, behind the times, belly up, big question mark, big question, bizarre decision, bleak, bloodbath, brake, bump in the road, bumpy, cautious outlook, challenge, choppy, close its doors, compression, conservative, contract, cost growth, cost pressure, creep, curve ball, curveball, cut numbers, dead horse, directly affected, directly hit, directly impacted, downtime, dramatically affected, dramatically impacted, evaporate, expense growth, fall apart, fall off a cliff, falloff, flattish, flatten, fluid situation, gap in the road, have an issue, headwind, heavily impacted, hiccup, impacted visibility, in the hole, in the hot seat, in the red, In uncharted waters, Inventory adjustment, keep up at night, lack of clarity, lack of transparency, lack of visibility, less visibility, limited visibility, low visibility, lumpiness, materially affected, materially hit, materially impacted, mixed bag, mixed result, nightmare, sustainability impacted
Risk & Uncertainty	**Uncertainty**	ambiguity, ambiguous, ambivalence, ambivalent, apprehension, bet, chance, chancy, changeability, changeable, danger, dangerous, debatable, defenseless, dicey, diffidence, diffident, dilemma, disquiet, dodgy, doubt, doubtful, doubtfulness, dubious, endanger, equivocating, equivocation, erratic, exposed, exposure, faltering, fear, fickleness, fitful, fluctuant, fluctuating, foolhardy, gamble, gnarly, hairy, halting, hazard, hazardous, hazy, headwind, hesitancy, hesitant, hesitating, hesitation, hiccup, iffy, imperil, in the hole, in the red, in uncharted waters, incalculable, incertitude, indecision, indecisive, indeterminate, insecure, insecurity, instability, inventory adjustment, irregular, jeopardize, jeopardy, likelihood, lumpiness, menace, misgiving, niggle, oscillating, parlous, pending, peril, perilous, possibility, possible, precarious, precariousness, probability, probable, prospect, qualm, quandary, query, reckless, reservation, risk, risked, riskier, riskiest, riskiness, risking, risky, skepticism, speculative, sticky, suspense, suspicion, tentative, tentativeness, threat, torn, treacherous, tricky, uncertain, uncertainties, uncertainty, unclear, unconfident, undecided, undependable, undetermined, unforeseeable, unknown, unpredictability, unpredictable, unreliability, unreliable, unresolved, unsafe, unsettled, unstable, unsure, unsureness, untrustworthy, vacillating, vacillation, vague, vagueness, variability, variable, varying, volatility, vulnerability, wager, wariness, wavering
	Political Risk	a government, a political, a president, a yes, african americans, american political, and congress, and party, and political, and senate, and social, any state, argued that, care reform, civil service, civil war, clause of, congress the, congress to, congress, constitution and, court has, due process, economic policy, elected officials, executive branch, executive privilege, federal bureaucracy, federal courts, federal reserve, first amendment, for governor, free market, general election, geopolitical attension, geopolitical change, geopolitical policy, geopolitical reform, geopolitical risk, geopolitical uncertainty, government and, government contract, government in, government is, government policy, government the, governor and, governor in, groups and, in congress, in government, interest group, interest groups, islamic state, judicial review, law suit, legal risk, legal uncertainty, limits on, most americans, national and, national government, national security, NATO, of citizens, of civil, of government, of politics, of religion, of representatives, of social, of speech, office of, other nations, passed by, political risk, political uncertainty, public policy, policy goals, policy is, policy, political parties, political party, political process, political system, politics and, politics is, politics of, powers and, president has, president obama, proposed by, public opinion, regulatory change, regulatory environment, regulatory policy, regulatory reform, regulatory risk, regulatory uncertainty, ruled that, sanction, senate and, shall have, shall not, social policy, state the, states or, struck down, support for, tax risk, tax tariff, tea party, that congress, the bureaucracy, the campaign, the candidates, the civil, the congress, the constitution, the constitutional, the democratic, the electoral, the epa, the faa, the gop, the governments, the house, the islamic, the legislative, the legislature, the nation, the party, the political, the politics, the presidency, the presidential, the republican, the south 1, the states, the taliban, the va, this election, to enact, vote for, war risk, war uncertainty, war, white house, yes vote, tariff

	Economic Risk	economic risk, economic uncertainty, economic change, economic reform, economic policy, economic environment, economic slowdown, economic recession, economic contraction, economic expansion, economic growth, economic recovery, headwind, financial risk, market risk, interest rate risk, downturn, slow down, systematic risk, bank run, crsis, macroeconomic, market crush, inflation, deflation, global recession, fiscal policy, monetary policy, unemployment
	Climate Risk	blizzard, climate disaster, climate risk, drought, dust storm, earthquake, environmental hazard, environmental risk, fire-storm, flood, flooded, flooding, forest fire, freeze, frigid, frost, frostbite, frostbitten, frozen, global warming, heatstroke, heatwave, hotness, hurricane, hyperthermia, hypothermia, landslide, natural disaster, precipitation, rainfall, severe weather, shivering, snowfall, snowstorm, solar-flare, storm, thunder, thunderstorm, tornado, torrid, torridity, tropical cyclone, tsunami, typhoon, volcanic, volcano, weather condition, wildfire, windy, winter storm, wintriness
	Legal Risk	legal risk, legal uncertainty, legal change, legal reform, legal policy, legal environment, corruption, antitrust, antitrust risk, anti-monopoly, monopoly, monopoly risk, antitrust law, compliance, regulator, litigation, lawsuit, settlement, legal proceed, legal action, legal dispute, legal claim, legal case, legal process, judicial proceed
	Cybersecurity Risk	cybersecurity, cyberattack, cybercrime, cybersecurity risk, cybersecurity threat, cybersecurity vulnerability, cybersecurity exposure, cybersecurity incident, cybersecurity breach, cybersecurity attack, computer attack, computer breach, computer break-in, computer joint security, computer security, computer system security, computer virus, confidential data, confidentiality of data, corruption of data, crimeware, cyber attack, cyber fraud, cyber incident, cyber insurance, cyber risk, cyber security, cyber terrorist, cyber threat, cyber-attack, cyber-fraud, cyber-incident, cyberinsurance, cyber-risk, cyber-security, cyber-terrorist, cyber-threat, cyber-based attack, cyberfraud, cyberincident, cyberrisk, cyberterrorist, cyberthreat, cyberbullying, ddos, data-breach, data breach, data confidentiality, data corruption, data theft, denial of service, dark web, encryption, espionage, hack, hacker, hacked, hacking, hacktivism, phishing, information security, information technology attack, information technology security, infosec, intrusion, keylogger, keystroke logging, malware, network break-in, network joint security, network security, ransomware, security breach, security expenditure, security incident, security management, security monitoring, social engineering, system security, unauthorized access
	Liquidity Risk	liquidity risk, liquidity uncertainty, financial tightness, financial constraint, financial stress, financial strain, financial pressure, financial difficulty, solvency, default, capital adequacy, credit crunch, credit rating, credit risk, counterparty risk, stress test, repo market, interest coverage, quick ratio, current ratio, collateral, counterparty risk, cash reserve
	Sovereign Risk	the America, the usa, the british, the china, chinese, europe, european, great britain, iran, israel, israelian, north korea, north korean, russia, russian, sovereign change, sovereign credit rating, sovereign credit risk, sovereign debt, sovereign default, sovereign downgrade, sovereign environment, sovereign policy, sovereign reform, sovereign risk, sovereign uncertainty, the eu, the uk, the us, united kingdom, united states
	Operational Risk	brand reputation, business continuity, business operation, contingency, control environment, crisis management, customer satisfaction, cybersecurity, CEO transition, CFO transition, fraud, health and safety, incident management, information risk, insurance, internal control, logistic, loss provision, manufacturing efficiency, network downtime, operational change, operational constraint, operational difficulty, operational downgrade, environment, operational policy, operational pressure, operational rating, operational reform, operational risk, operational strain, operational stress, operational tightness, operational uncertainty, process efficiency, process improvement, optimization, quality, production safety, regulatory compliance, resilience, supply chain constraint, supply chain disorder, supply chain disruption, supply chain management, supply chain risk, supply chain shortage, supply shortage, supply constraint, step away

Technology	**Technology**	agile methodology, AI, API, AR, artificial intelligence, augmented reality, automation, automative, autonomous vehicle, AWS, azure, big data, biometrics, blockchain, chatbot, cloud computing, contactless technology, crowdsourcing, cryptocurrency, cybersecurity, data analytics, devops, digital marketing, digital transformation, digital twin, drones, e-commerce, edge computing, fintech, gaming, geolocation technology, google cloud, health tech, IAAS, industry 4, innovation, innovative, internet censorship, internet of things, IOT, machine learning, mobile payments, mobile technology, nanotechnology, open source, PaaS, patent, personalization, predictive analytics, prototype, quantum computing, quantum encryption, robot, robotic, robotics, SaaS, self-driving cars, smart devices, smart home technology, social media, technology, user-generated content, virtual reality, VR, wearable technology
	Artificial Intelligence	adaptive system, adversarial attack, adversarial learning, adversarial networks, ai tools, AI, algorithmic decision-making, ambient intelligence, analytics, artificial intelligence, artificial intelligence, association rules, attention mechanisms, augmented reality, automated reasoning, automation, autonomous systems, bayesian networks, BERT, big data, chatbot, chatgpt, clustering algorithms, cognitive analytics, cognitive computing, computer vision, content generate, contextual understand, conversational ai, data science, decision trees, deep learning, deepfake, digital assistant, digital transformation, dimensionality reduction, edge computing, expert systems, fake news detection, fine-tuning, generative models, human-robot collaboration, image recognition, intelligent agents, intelligent assistant, intelligent automation, intelligent edge, intelligent interface, intelligent optimization, intelligent sensors, internet of things, keywords, large language model, LLM, machine intelligence, machine learning, model selection, natural language processing, neural architecture search, neural networks, neural processing, open ai, openai, pattern recognition, pre-trained, predictive modeling, principal component analysis, random forests, reinforcement learning, robotic process automation, robotics, self-learning systems, semi-supervised learning, smart algorithms, smart systems, supervised learning, support vector machines, text classification, text completion, text generation, text summarization, transfer learning, transformer, unsupervised learning, virtual reality, voice recognition
	Automation	automation, robotics, RPA, robotic process automation, artificial intelligence, machine learning, process automat, digital transformation, digitalization, digit process automat, autonomous drive, autonomous vehicle, autonomous system
Urgency and Emphasis	**Urgency**	urgent, immediate, immediately, quick, fast, fastest, rapid, rapidly, speed, speed up, sped up, catch up, speeding up, accelerate, quickly, priority, swift, prompt, promptly, expedite, expeditiously, expeditious, time-critical, time-sensitive, red code, short term, short-term, short run
	Emphasis	emphasize, highlight, highlighted, significant, important, critical, key, must, have to, has to, essential, vital, imperative, mandatory, prominent, pronounced, marked, eminent, noticeable, striking, conspicuous, remarkable, salient, notable, distinctive, pivotal, crucial, decisive, indispensable, integral, necessity, requisite, urgent, compelling, forceful, persuasive, strong, assertive, definitive, unmistakable, expressive, intense, resounding, strident, sure, confident, insistent, positive, surprise, dog and pony show, misunderstanding