Earnings Call Sentiment and Stock Returns in Brazil *

Derek Poustka
derekdp@al.insper.edu.br

Rodrigo De Losso da Silveira Bueno
delosso@usp.br

Abstract

This paper examines whether corporate sentiment extracted from earnings call transcripts (ECTs) affects the short-term stock market returns of Brazilian firms listed on the IBovespa B3 index. We compare a traditional lexicon-based method with modern transformer-based Natural Language Processing (NLP) models to classify call-level sentiment. Specifically, we use the Loughran and McDonald (2011) dictionary, FinBERT, and GPT to extract sentiment from corporate disclosures. The analysis is conducted under two specifications: one using the full transcript as a single document and another separating the text into *Presentation* and *Q&A* sections. Overall, we find that sentiment helps explain cumulative abnormal returns (CAR), although the strength and direction of the effect vary with the sentiment classification method and transcript structure. These findings highlight the importance of methodological choice and transcript segmentation in extracting market-relevant information from corporate disclosures.

Keywords: *Corporate sentiment; Stock returns; Earnings call transcripts; Natural language processing; Domain-specific dictionary; FinBERT; GPT; Brazil.*
JEL Classification: *G12; G14; G34; G41.*

1 Introduction

The modern financial landscape is characterized by an unprecedented volume of information, which goes far beyond traditional quantitative reports. Corporations and financial institutions continuously publish large amounts of unstructured textual data through press releases, regulatory filings, and, notably, earnings call transcripts (ECTs). These communications serve as critical tools to inform markets beyond formal reporting periods, helping shape investor expectations and influencing asset prices (Frankel, Johnson, & Skinner, 1999; F. Li, 2010).

Over the past decade, Natural Language Processing (NLP) techniques have made it possible to transform this type of qualitative content into structured variables suitable for empirical analysis. This has allowed researchers in economics and finance to capture insights embedded in corporate narratives and use them to understand investor behavior and asset price dynamics (Gentzkow, Kelly, & Taddy, 2019). In particular, sentiment analysis has become a prominent technique for extracting the tone of financial disclosures, enabling the measurement of optimism or pessimism conveyed in language (Loughran & McDonald, 2016).

Earlier approaches to analyzing sentiment relied heavily on dictionary-based methods. For instance, Loughran and McDonald (2011) developed a domain-specific lexicon that reclassified words that general-purpose dictionaries often misclassified. Terms like *"tax"*, *"liability"*, or *"cost"* which are typically marked as negative in generic sentiment dictionaries, are frequent and often neutral in financial documents. The specialized nature of their dictionary made it

*This version was submitted for presentation at the XXVI Brazilian Finance Meeting.

widely adopted in financial research, forming the basis for much of the early empirical literature.

Despite their influence, dictionary-based approaches face limitations. They typically quantify sentiment by counting word frequencies, disregarding contextual nuances, sentence structure, and grammatical features. This can lead to misclassification, particularly in the presence of negation, uncertainty, or conditional statements. To address these issues, later methods introduced richer word representations, known as embeddings, which encode semantic similarity by mapping words into continuous vector spaces (Gentzkow et al., 2019). These embeddings improve flexibility and allow sentiment to be modeled in a more context-sensitive manner.

Building on this foundation, more advanced approaches began to model language at the sequence level. While early applications of embeddings relied on static representations, newer architectures introduced mechanisms to capture word meaning in context. Transformer-based models such as BERT (Devlin, Chang, Lee, & Toutanova, 2019) significantly enhanced text analysis by incorporating bidirectional attention across documents. In finance, models like FinBERT (Araci, 2019) extended this architecture to domain-specific corpora, improving sentiment classification in financial contexts. More recently, generative models such as GPT (T. Brown et al., 2020) have demonstrated the ability to assess sentiment using broad contextual understanding.

In this paper, we study whether sentiment expressed in ECTs can explain cumulative abnormal returns (CARs). We apply three distinct methodologies to measure sentiment: (i) a domain-specific dictionary approach using the Loughran and McDonald (2011) dictionary,[1] (ii) FinBERT, a fine-tuned transformer model adapted to financial texts, and (iii) GPT, a generative large language model using zero-shot prompting. These approaches enable comparison across lexicon-based, fine-tuned, and generative paradigms for sentiment detection.

We collect earnings call transcripts (ECTs) for the 30 largest firms in terms of current weight in the Ibovespa B3 index, covering the period from 2002 to 2024. Each transcript is cleaned and preprocessed before being classified into a ternary sentiment category (positive, negative, or neutral) according to the applied model. This classification scheme is adopted to ensure comparability across methods, following the approach of Hasani and Zaw (2024). The resulting sentiment labels are then converted into dummy variables and included as explanatory variables in panel regressions. The dependent variable is the firm's cumulative abnormal return (CAR) around the earnings call date, capturing the short-term market response to the information disclosed in the call (S. Brown & Warner, 1985; MacKinlay, 1997).

To assess whether the informativeness of sentiment varies across different parts of the transcript, we separate each ECT into two components: the *Presentation* and the *Q&A* sections. Prior research has shown that spontaneous interactions between analysts and executives during the Q&A often reveal more value-relevant information than the scripted portion of the call (Matsumoto, Pronk, & Roelofsen, 2011; Mayew & Venkatachalam, 2012). By analyzing each

[1] Henceforth referred to as the LM dictionary.

segment independently, we aim to identify whether the market reacts differently to sentiment depending on the structure and spontaneity of the communication.

This study contributes to the literature in several ways. First, while prior studies have applied established NLP methodologies to the Brazilian market, our paper provides one of the first systematic comparisons that includes a latest-generation generative model (GPT) alongside a domain-specific transformer (FinBERT) and a traditional dictionary-based approach. This comparative framework offers novel insights into the relative performance and specific capabilities of these evolving techniques for financial text analysis. Finally, the paper underscores the importance of disaggregating corporate disclosures, demonstrating that analyzing sentiment within distinct structural components of the communication can yield more nuanced insights than treating the document as a single unit.

The remainder of this article is organized as follows. Section 2 reviews the relevant literature, with a particular focus on applications of natural language processing (NLP) in finance and sentiment analysis. Section 3 describes the implementation of the NLP models and techniques. Section 4 presents the data sources, including the earnings call transcripts and firm-level variables, as well as the construction of the outcome variable. Section 5 reports the main empirical results. Finally, Section 6 concludes and discusses implications for future research.

2 Related Work

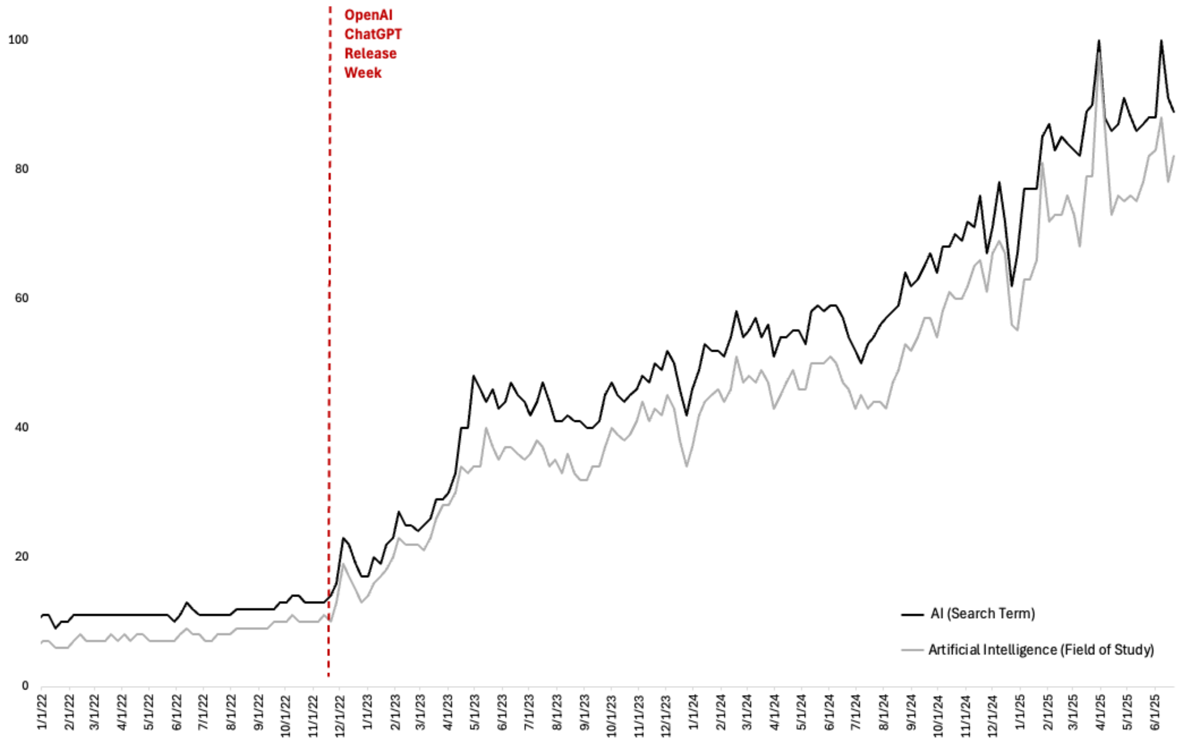
2.1 Applications of *Natural Language Processing* (NLP) in Finance

In recent years, several key developments have brought *Natural Language Processing* (NLP) models to the forefront of computational research. Foundational advances such as *Word2Vec* (Mikolov, Chen, Corrado, & Dean, 2013) and GloVe (Pennington, Socher, & Manning, 2014) introduced distributed word representations that significantly improved machines' ability to capture semantic meaning. Subsequent models like BERT (Devlin et al., 2019) further enhanced language understanding by incorporating contextual embeddings based on transformer architectures, setting new performance benchmarks across a wide range of NLP tasks.

Over the past few years, the release of large-scale generative models, particularly OpenAI's GPT series, has underscored the potential of transformer-based architectures for producing fluent and human-like text. The public launch of ChatGPT in late 2022 marked a key inflection point in the diffusion of AI and generative AI (GenAI). Global interest in "AI" and artificial intelligence more broadly rose sharply after the release, as evidenced by Google search trends presented in Figure 1. This surge reflects not only growing public interest but also a marked acceleration in academic research focused on language models.

These developments have also shaped the financial sector, where *Large Language Models* (LLMs) are increasingly used in applied contexts. According to Zhao et al. (2024), models such as GPT-4 are already being employed in the automation of financial reports, market forecasting, sentiment analysis, and the delivery of personalized financial guidance. Other applications include insolvency prediction, portfolio optimization, and fraud detection. These examples demonstrate how LLMs enhance the ability of financial institutions to extract insights from unstructured data while improving scalability and operational efficiency.

Figure 1: Global Interest in AI and Artificial Intelligence (Topic)



Notes: This figure displays Google search trends over the past five years for two distinct query types: the term "AI" (black line) and the broader topic "Artificial Intelligence" (gray line), which aggregates semantically related searches across languages. The data show a marked increase in interest beginning in late 2022, coinciding with the public release of ChatGPT. Available at Google Trends. Accessed in June, 2025.

2.2 Vectorization

Vectorization is a fundamental step in natural language processing (NLP), as it transforms unstructured textual data into numerical formats that can be processed by statistical models and machine learning algorithms. Since most algorithms operate on numeric input, an appropriate transformation of text into vectors is essential for any quantitative text analysis.

2.2.1 Frequency-Based Approaches

The most traditional vectorization techniques rely on the frequency of word occurrences in a given corpus. One of the simplest and most widely used methods is the *Bag of Words* (BoW) approach. In BoW, each document is represented as a vector whose components correspond to the count of each word from a fixed vocabulary. This method is straightforward to implement and can be effective for basic classification or clustering tasks. However, it ignores the order of words and treats each word independently, failing to capture context or semantics.

A refinement of BoW is the *Term Frequency-Inverse Document Frequency* (TF-IDF) method. TF-IDF adjusts the raw frequency of a word in a document by penalizing terms that are frequent across the entire corpus. Specifically, it assigns higher weights to words that appear often in a specific document but are rare in the collection as a whole. The TF-IDF weight of a word w in document d is given by:

$$\text{TF-IDF}(w, d) = \text{TF}(w, d) \times \text{IDF}(w), \tag{1}$$

where

$$\text{TF}(w, d) = \frac{f_{w,d}}{\sum_{w'} f_{w',d}} \quad \text{and} \quad \text{IDF}(w) = \log\left(\frac{N}{1 + n_w}\right). \tag{2}$$

In this formulation, $f_{w,d}$ is the frequency of word w in document d, N is the total number of documents in the corpus, and n_w is the number of documents containing the word w. The logarithmic scaling of the inverse document frequency prevents extremely common words from dominating the representation. As demonstrated by Loughran and McDonald (2011), TF-IDF can be particularly effective in financial text analysis, where it is crucial to emphasize domain-specific terminology while downweighting generic language.

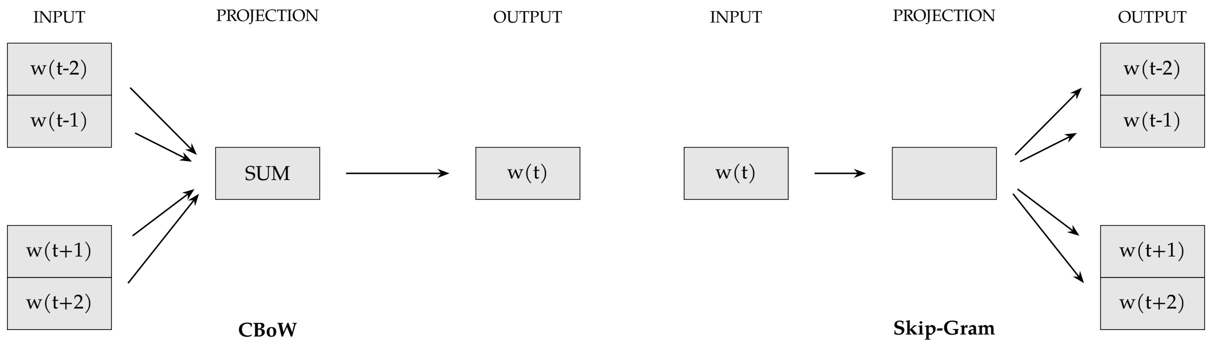
2.2.2 Embeddings

To overcome the limitations of sparse and context-independent representations, researchers have developed dense vector representations known as embeddings. These approaches build on the *distributional hypothesis*, famously summarized by Firth (1957) as "you shall know a word by the company it keeps." The core idea is that words appearing in similar contexts tend to share semantic meaning and, therefore, can be mapped to nearby positions in a continuous vector space.

One of the most influential implementations of this idea is the *Word2Vec* framework developed by Mikolov et al. (2013). Word2Vec offers two primary architectures for learning word embeddings: the *Skip-gram* model and the *Continuous Bag of Words* (CBoW) model. In the Skip-gram model, the objective is to predict the surrounding context words given a central target word, enabling the model to capture how the target contributes to its local semantic environment. Conversely, CBoW reverses this logic by predicting the target word based on its neighboring context. Both models are implemented using shallow neural networks and learn word vectors by maximizing the probability of accurate context prediction. These architectures are visually summarized in Figure 2.

Unlike BoW or TF-IDF, these embedding-based techniques produce dense, low-dimensional representations that encode both syntactic and semantic information. This makes them especially effective in domains like finance, where the same term can carry different meanings depending on the context. In addition to improved predictive performance, these embeddings also support useful linguistic operations. For instance, vector arithmetic on word embeddings can capture relationships, such as vector("king") $-$vector("man") $+$ vector("woman") $\approx$

Figure 2: Word2Vec Model Architectures



Notes: This figure illustrates two Word2Vec variants with a context window of two: Continuous Bag of Words (CBoW) leverages the neighboring word vectors $\{w(t-2), w(t-1), w(t+1), w(t+2)\}$ to infer the vector $w(t)$, whereas Skip-Gram uses $w(t)$ to infer each of its neighboring vectors. Here $w(t)$ denotes the vector at position t, and $w(t \pm k)$ denotes vectors at positions offset by k. The figure is inspired by Mikolov et al. (2013).

vector("queen"). Those capabilities have become widely used for both analytical and visualization purposes. One common technique is to project high-dimensional embeddings into two dimensions using dimensionality reduction techniques like t-SNE or PCA.[2] This enables the construction of word clouds or scatter plots that highlight clusters of semantically related words.

More recently, the development of *contextualized embeddings* has significantly advanced the representation of language. Unlike Word2Vec, which assigns a single, static vector to each word, models such as BERT and GPT generate representations that vary dynamically with the context. These models are built upon the Transformer architecture, which relies on self-attention mechanisms (Vaswani et al., 2017). In this framework, each token is represented by a *Query*, a *Key*, and a *Value* vector. The attention mechanism computes similarity scores between queries and keys to determine the relative importance of other tokens in the sequence. These scores are then used to weight and aggregate the corresponding value vectors, producing context-sensitive embeddings that capture both local and global linguistic dependencies.

2.3 Sentiment Analysis

According to Liu (2020), sentiment analysis is the computational study of people's opinions, sentiments, emotions, moods, and attitudes. Also known as opinion mining, it is a Natural Language Processing (NLP) technique used to determine the sentiment expressed in textual data. It enables classification into categories such as positive, negative, or neutral, and has extensive applications in finance, marketing, and social media analytics. Methodologically, sentiment analysis approaches are generally grouped into three main categories: lexicon-based, machine learning-based and deep learning-based models.

[2] Principal Component Analysis (PCA) is a linear dimensionality reduction technique that transforms high-dimensional data into a lower-dimensional space by projecting it onto the directions of maximum variance. In contrast, t-Distributed Stochastic Neighbor Embedding (t-SNE) is a non-linear technique that preserves local structure and is particularly effective for visualizing high-dimensional data in two or three dimensions.

2.3.1 Rule-Based and Dictionary-Based Approaches

Dictionary and rule-based sentiment analysis methods represent foundational techniques in the financial text analysis literature. These approaches determine the polarity of a given text by counting occurrences of predefined positive and negative words from lexicons, assuming that a higher frequency of positive (or negative) terms corresponds to the tone of the document (Todd, Bowden, & Moshfeghi, 2024). Early implementations used general-purpose dictionaries, such as Harvard IV[3] and DICTION[4], but these often misclassified financial terms like "liability" or "cost". To address this, domain-specific dictionaries were developed, most notably the Loughran and McDonald (2011) lexicon, which redefined sentiment categories based on the actual usage of terms in financial filings. These refined dictionaries improved classification accuracy in financial contexts and have been frequently used in studies analyzing earnings press releases, financial news, and earnings call transcripts. Despite their simplicity and low computational requirements, dictionary methods are limited by fixed word sentiment orientations and a lack of contextual adaptability.

The literature recognizes these limitations and has gradually shifted toward more sophisticated NLP techniques. However, dictionary-based approaches remain prevalent due to their transparency and replicability. The LM dictionary, for instance, was manually constructed by examining the frequency and meaning of words in 10-K filings, and is continuously updated to reflect financial discourse. These word lists provide a rule-based framework that does not require annotated training data and are still frequently used as benchmarks or components within hybrid systems. While recent advances in sentiment analysis emphasize deep learning and contextual embeddings, the interpretability and robustness of dictionary-based sentiment scores continue to offer value, especially in settings where model transparency is critical or when computational resources are limited.

2.3.2 Machine Learning-Based Approaches

Machine learning (ML) approaches have considerably enhanced sentiment analysis in financial applications by improving classification accuracy and enabling the analysis of large volumes of unstructured data. Among the most common methods is the Naïve Bayes classifier, which estimates the probability of a sentiment category based on word frequencies in labeled corpora. These models are computationally efficient and require minimal training data, making them appealing for early-stage implementations (Todd et al., 2024). Other tools, such as sentiment engines that combine ML with rule-based NLP components, have been used to assess the relevance and polarity of news articles in real time, often producing sentiment indicators that correlate with contemporaneous market activity. Applications span a variety of data sources, including earnings announcements, corporate filings, and financial news, with results showing that machine-learned sentiment can be associated with abnormal returns, trading volume, and

[3] Harvard IV is a general-purpose sentiment dictionary developed for content analysis across various domains. See inquirer.sites.fas.harvard.edu. Accessed in August, 2025.

[4] DICTION is a text analysis software that uses word lists to measure tone and other linguistic properties in written communication. See methods.sagepub.com. Accessed in August, 2025.

volatility.

Despite these advances, traditional ML models are constrained by their dependence on manual feature engineering and their inability to fully capture semantic context. While they outperform dictionary-based methods in many settings, they remain less flexible than more recent neural architectures. The application of ML to financial sentiment has also expanded into social media, where real-time signals from platforms such as X (formerly known as Twitter) are used to forecast intraday or short-horizon returns. However, the accounting and finance literature continues to adopt such models at a slower pace than the broader NLP field. Challenges include the lack of large, domain-specific annotated datasets and the difficulty of aligning financial language with generic sentiment training objectives (Todd et al., 2024). As a result, while ML approaches offer clear improvements over earlier methods, their effectiveness remains dependent on data quality and model design.

2.3.3 Deep Learning and Transformer-Based Approaches

Deep learning has played a central role in recent advances in Natural Language Processing (NLP), enabling models to achieve state-of-the-art performance across a variety of language understanding tasks. In the context of financial sentiment analysis, deep learning models based on neural networks have shown greater accuracy compared to traditional dictionary-based methods, as they can capture non-linear patterns and contextual relationships in text. Despite these advantages, the adoption of deep learning in accounting and finance remains limited (Todd et al., 2024). One reason is the lack of large, domain-specific datasets needed to train these models effectively. In addition, higher model complexity does not always guarantee improved performance, especially when the input text is noisy, short, or contextually ambiguous. As a result, the use of deep learning methods in financial applications still faces practical and methodological constraints.

The transformer architecture represented a paradigm shift in natural language processing by employing self-attention mechanisms, enabling parallel sequence processing and more efficient capture of long-range dependencies. This innovation led to the development of highly influential models such as BERT and GPT, which have been widely used for text classification and sentiment analysis. These models can be further improved through domain-specific pretraining and fine-tuning, as demonstrated by FinBERT, which was trained on documents related to financial context. Transformer-based models provide richer semantic representations and have been applied not only to stock return prediction but also to fraud detection, financial disclosure analysis, and market sentiment monitoring. Nonetheless, their deployment is often limited by computational cost, data requirements, and challenges related to interpretability and reproducibility.

2.4 Textual Analysis of Brazilian Earnings Calls

Research on Brazilian Earnings Call Transcripts (ECTs) has largely focused on measuring the tone of corporate communication through dictionaries. Ferreira, Fiorot, Motoki, and Moreira (2019) analyze managers' speeches in conference calls and find that optimism is positively related to both contemporaneous and future firm performance. Importantly, their study adapts the Loughran and McDonald (2011) dictionary to Portuguese, addressing linguistic and cultural differences that challenge the direct application of U.S.-based word lists. Tonin and Scherer (2021) also employ the LM dictionary, but refine the analysis by disaggregating tone by participant type, showing that analysts' tone has greater explanatory power for abnormal returns than managers' tone.

Beyond dictionary-based approaches, other contributions employ different strategies to capture disclosure content. Viana Junior, Arruda Castro, Rodrigues Ponte, and Chagas Lima (2019) apply IBM Watson's Natural Language Understanding platform[5] to extract sentiment from conference calls, focusing exclusively on managers' speeches. They find that optimism expressed in the Q&A section is positively associated with abnormal returns, even though sentiment levels are generally lower in this section compared to prepared remarks. Moreira, Ramos, Kozak-Rogo, and Rogo (2016), in turn, do not measure sentiment directly, but rather use the duration of calls as a proxy for informational content, documenting that firms with negative news tend to hold longer calls and disclose more information.

The empirical configurations of these studies vary substantially. Ferreira et al. (2019) analyze Brazilian firms with American Depositary Receipts (ADRs) between 2002 and 2016, while Moreira et al. (2016) focus on high-governance firms listed on the BM&F Bovespa during 2008-2015. Tonin and Scherer (2021) study 44 firms listed on the Ibovespa index between 2010 and 2017, and Viana Junior et al. (2019) examine 24 Novo Mercado[6] companies in 2016-2017. Taken together, this literature underscores both the potential and the limitations of textual analysis in the Brazilian context. Most studies rely on sentiment dictionaries, with one explicitly adapting the LM lexicon for Portuguese and another turning to a machine-learning platform like IBM Watson. Despite methodological differences, the results consistently point to the informational role of ECTs and the distinct contribution of Q&A interactions, highlighting the importance of both measurement choice and transcript segmentation in shaping empirical findings.

3 Natural Language Processing (NLP) Methods and Techniques

Our main explanatory variable is constructed using both traditional and more contemporary *Natural Language Processing* (NLP) methods and techniques. In this section, we describe

[5]IBM Watson Natural Language Understanding is a cloud-based service that uses natural language processing to analyze text for sentiment, emotion, entities, and other linguistic features. See www.ibm.com. Acessed on August, 2025.

[6]Novo Mercado is B3's highest corporate governance listing segment, requiring companies to adopt enhanced practices such as improved disclosure standards, independent board members, and strengthened shareholder protection mechanisms. See www.b3.com.br. Accessed in August, 2025.

these methodological procedures in detail. For additional details on the preprocessing procedures and the structure of the transcripts, see Appendix A.

3.1 Preprocessing and Cleaning

As is common in NLP applications, we follow a structured workflow that extends the traditional data science pipeline by incorporating an additional step known as text preprocessing. This step, positioned between data cleaning and exploratory analysis, is essential for transforming unstructured textual data into a format suitable for quantitative analysis.

3.1.1 Raw ECTs

The raw earnings call transcripts (ECTs) used in this study were obtained from LSEG Workspace and are downloaded in plain text format (`.txt`). Each file corresponds to a single earnings event and contains the full transcript of the call. The file names follow a consistent pattern that encodes metadata useful for organizing and preprocessing the data. A typical file name includes the event date, the Reuters Instrument Code (RIC)[7], and a numeric identifier, followed by a standard suffix. For example:

```
2024-Mar-08_PETR4.SA-1710428352937-transcript.txt
```

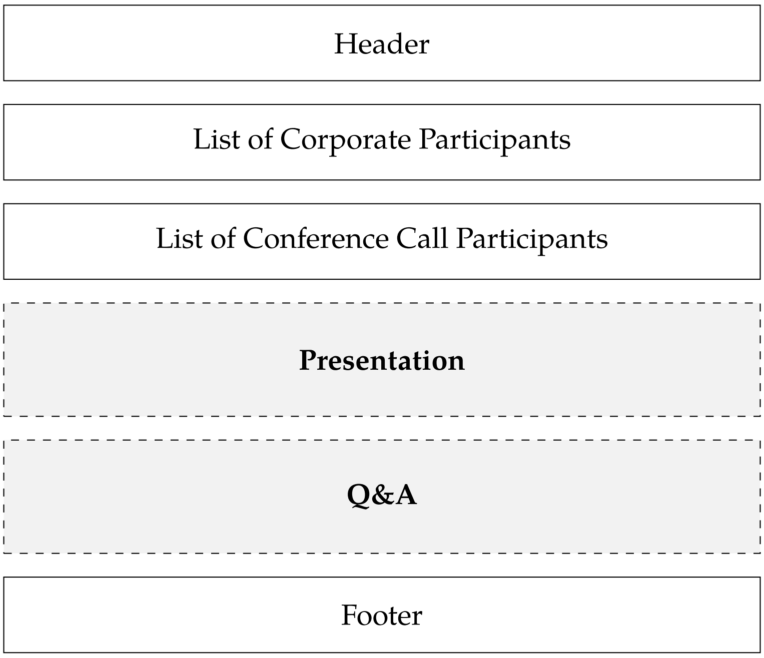
where `2024-Mar-08` represents the date of the earnings call, while `PETR4.SA` is the RIC corresponding to Petróleo Brasileiro S.A. (Petrobras). As noted by K. Li, Mai, Shen, and Yan (2021), *LSEG Workspace* retroactively updates company identifiers across historical records to reflect the most recent naming conventions, particularly in cases of mergers or ticker changes. However, the original company name is often retained within the transcript's internal title block, which provides an additional layer of historical traceability and allows researchers to mitigate distortions introduced by such backfilling.

Once a transcript is loaded, its internal structure follows a standardized format that facilitates computational analysis. Each document is divided into six main sections: *Header*, *List of Corporate Participants*, *List of Conference Call Participants*, *Presentation*, *Q&A*, and *Footer*. The *Header* contains metadata about the event, including the name of the transcript provider, the label "EDITED VERSION" (which indicates that the content has been reviewed and corrected by professional editors), the official event title, and the date and time of the call.

Following the header, the document lists the *Corporate Participants*, which includes the names and roles of company executives participating in the call, such as the Chief Executive Officer (CEO), Chief Financial Officer (CFO), or Investor Relations (IR) Officer. This is followed by the *Conference Call Participants*, typically composed of equity analysts or institutional investors who engage in the question-and-answer segment.

[7] The Reuters Instrument Code (RIC) is a market-level identifier for instruments and pricing sources issued by LSEG. See www.lseg.com. Accessed in March, 2025.

Figure 3: Raw ECT Structure



Notes: This figure illustrates the standard structure of earnings call transcripts (ECTs) as provided by LSEG Workspace. Each document follows a consistent format, beginning with a header and participant lists, followed by the presentation and Q&A sections, and concluding with a legal and informational footer. The central sections, Presentation and Q&A, form the basis for sentiment and content analysis.

The core of the transcript is divided into two primary content sections. The *Presentation* contains the company's prepared remarks, which offer structured commentary on recent financial performance and forward-looking statements. This section is generally more controlled and may include elements of strategic messaging or cheap talk. In contrast, the *Q&A* section captures unscripted interactions between analysts and company representatives. It tends to be more informative, as it elicits clarifications, follow-up questions, and spontaneous responses that often reveal management tone, priorities, and confidence levels.

The transcript concludes with a *Footer*, which includes definitions, legal disclaimers regarding forward-looking statements, and copyright information. This standardized layout enables systematic segmentation of the document, allowing for targeted preprocessing and sentiment analysis focused on the most relevant portions of managerial communication.

3.1.2 Minimally Prepocessed ECTs

To preserve the semantic and contextual integrity of the transcripts, we construct a minimally preprocessed version of the earnings call texts. This representation is particularly suited for Transformer-based language models, which benefit from exposure to unaltered or lightly preprocessed text. These models rely on long-range dependencies and nuanced patterns that can be distorted by overly aggressive cleaning. The goal of this preprocessing stage is to retain the richness of linguistic expression while removing elements that are irrelevant or disruptive to textual analysis.

The process begins with the extraction of key metadata from each transcript. Using regular expressions (`regex`), we retrieve the company name, event date, and time, which are stored separately to facilitate merging with financial data and will be used in subsequent analysis. We also parse the list of participants and distinguish between internal and external participants. The number of external participants is recorded as a proxy for coverage and analyst engagement, and will later be used as a control variable.

Next, we implement a rule-based cleaning procedure to remove structural noise from the transcripts. This includes eliminating headers and footers, speaker tags, operator remarks, and participant lists. Once cleaned, each transcript is divided into two sections: the *Presentation*, which contains the company's prepared remarks, and the *Q&A*, which captures the interactive dialogue between management and external analysts. These sections are stored separately to enable section-specific analysis. For the unified document analysis, however, the transcript is reassembled after structural noise removal in the case of Transformer-based models, and after full preprocessing in the case of the dictionary-based approach.

3.1.3 Preprocessed ECTs

Text preprocessing is a crucial step in NLP, ensuring that our data is standardized, structured, and free of noise for subsequent analysis. A conservative, multi-stage pipeline is employed, specifically designed to support dictionary-based sentiment analysis using the Loughran and McDonald (2011) financial word list. This approach prioritizes lexical fidelity and minimizes transformations that could distort sentiment measurements. The methodology closely follows the strategy proposed by Hasani and Zaw (2024), which is optimized to align with the structure and vocabulary of the LM dictionary.

The first step in the pipeline is the expansion of English contractions, a normalization step that prevents tokenization inconsistencies and ensures that different forms of the same expression are treated uniformly. Next, proper nouns are removed using part-of-speech (PoS) tagging with the `spaCy` library in Python. As these terms are typically sentiment-neutral and context-dependent, their exclusion helps reduce noise. Importantly, this step is executed before lowercasing, since capitalization plays a key role in the correct identification of named entities.

Following this, the text is converted to lowercase, and all punctuation and numeric characters are removed. These normalization procedures ensure consistent matching between tokens and dictionary entries, and eliminate non-lexical elements that carry no sentiment weight, such as formatting characters or numerical references that do not convey tone. Removing these elements reduces dimensionality and avoids artificial inflation of the vocabulary. The cleaned text is then tokenized using the `nltk` library, segmenting the transcript into individual lexical units suitable for analysis in downstream applications.

To further reduce noise while preserving meaningful content, a customized stopword list is employed. Generic stopword lists are inadequate in financial contexts, as they tend to exclude polarity-bearing terms that are crucial in interpreting tone shifts. In line with Hasani and

Zaw (2024), we modify the standard `nltk` stopword set to explicitly retain sentiment-relevant words such as "not," "no," "up," and "down," which frequently appear in financial disclosures and carry strong directional meaning. This adjustment enhances the sensitivity of the sentiment analysis while maintaining linguistic coherence in the processed text.

As a final cleaning step, tokens with fewer than two characters are discarded. These short tokens are often artifacts from earlier transformations and are highly unlikely to represent semantically meaningful terms. Eliminating them helps to streamline the token set, improving both processing efficiency and the overall signal-to-noise ratio in the analysis. Their exclusion further refines the dataset without sacrificing relevant content or interpretability.

Critically, and in direct alignment with the methodology of Hasani and Zaw (2024), we abstain from applying stemming or lemmatization. The Loughran and McDonald (2011) dictionary includes specific word inflections rather than just their root forms. Applying normalization techniques that reduce words to their stems or lemmas would result in systematic mismatches between the text and the dictionary, thereby underestimating the frequency of sentiment-bearing terms. By preserving the original lexical forms, the preprocessing pipeline maintains full compatibility with the dictionary and ensures accurate and replicable sentiment scores.

3.2 Dictionary-Based Approach

As a baseline for sentiment analysis, we implement a dictionary-based method using the domain-specific lexicon developed by Loughran and McDonald (2011). The LM dictionary was constructed manually and contains 347 positive and 2,345 negative words that were selected to reflect the unique semantics of financial texts. These classifications account for the fact that terms such as "liability," "cost," or "tax" often carry neutral or technical meanings in finance, despite being flagged as negative in general-purpose lexicons. Since its initial release, the dictionary has been periodically updated and expanded. Its application has become standard in empirical research involving financial text sentiment.

Following the methodology applied by Hasani and Zaw (2024), we compute a tone score for each earnings call transcript based on the relative frequency of positive and negative words. The *initial tone score* τ_i for document i is defined as:

$$\tau_i = \frac{\sum w_i^+ - \sum w_i^-}{\sum \omega_i}, \tag{3}$$

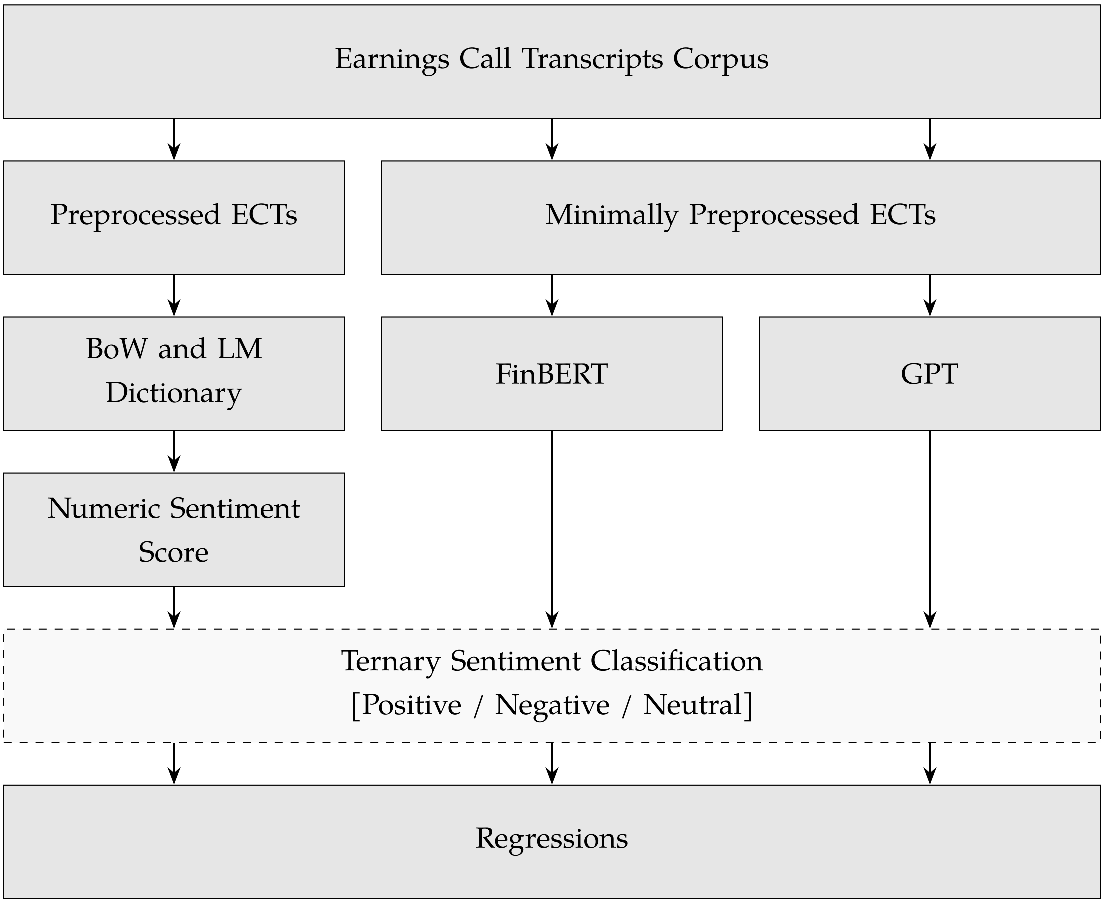
where $\sum w_i^+$ and $\sum w_i^-$ are the total counts of positive and negative words in the transcript, and $\sum \omega_i$ is the total number of words in the document. To standardize tone scores across the corpus, we apply min-max normalization to obtain a rescaled score $\tilde{\tau}_i$ in a specified range $[\min^{\text{NEW}}, \max^{\text{NEW}}]$. This is computed as:

$$\tilde{\tau}_i = \frac{\tau_i - \min(\tau)}{\max(\tau) - \min(\tau)} \cdot \left(\overset{NEW}{\max} - \overset{NEW}{\min}\right) + \overset{\text{NEW}}{\min}. \tag{4}$$

In our analysis, we normalize tone scores to the range $[-1, 1]$, setting $\min^{\text{NEW}} = -1$ and $\max^{\text{NEW}} = 1$. This normalized score $\tilde{\tau}_i$ serves as the input for classification and sensitivity procedures in the subsequent stages of the pipeline.

Unlike the other two approaches, which directly generate ternary sentiment classifications, the dictionary-based method produces a continuous numeric output. To ensure comparability across methods, this score must be discretized into one of three sentiment categories: *Positive*, *Neutral*, or *Negative*. This harmonization step is necessary for integrating the dictionary-based results into the unified sentiment classification framework illustrated in Figure 4.

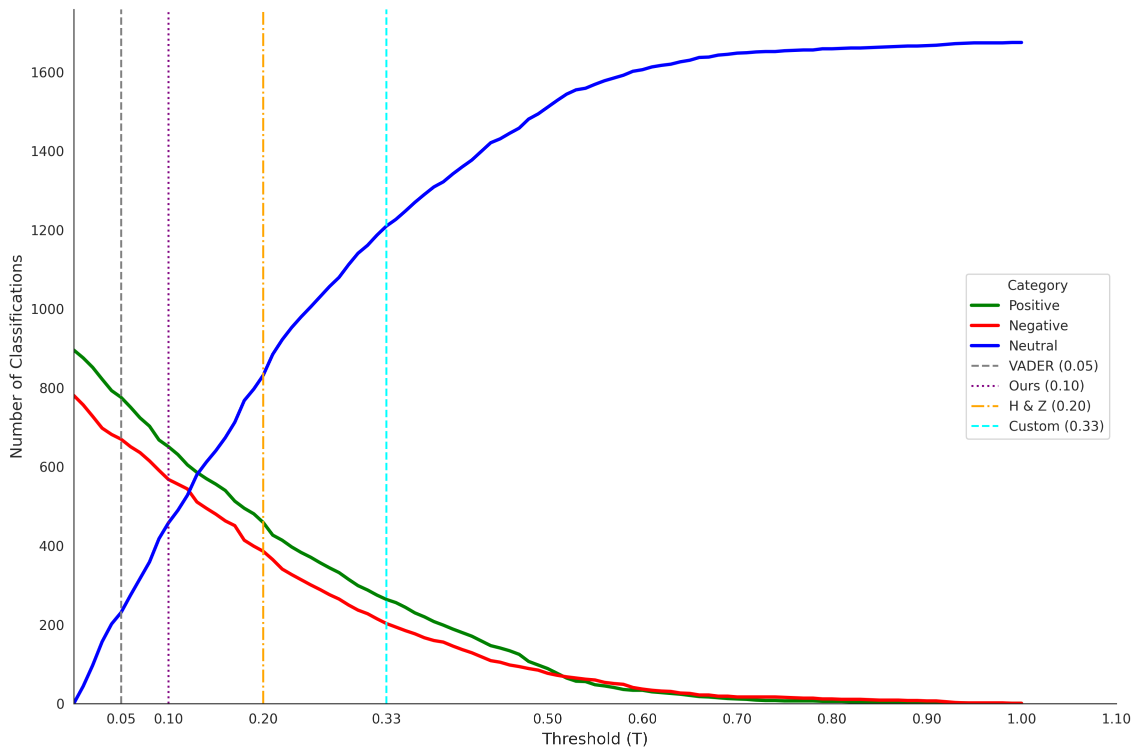
Figure 4: Overview of the Sentiment Analysis Pipeline for Earnings Call Transcripts



Notes: This figure presents the full architecture of the sentiment analysis pipeline implemented for Earnings Call Transcripts (ECTs). The process begins with the raw ECT corpus, which is directed through two distinct preprocessing streams. On the left, standard preprocessing techniques generate a more structured textual data for vectorization via BoW, followed by the computation of a numeric sentiment score based on Loughran and McDonald (2011) domain-specific dictionary. In parallel, a minimally preprocessed version of the corpus feeds into transformer-based models: FinBERT and GPT. The outputs from all three sentiment measures (dictionary, FinBERT, GPT) are mapped into a unified ternary classification scheme (positive, negative, neutral). This standardized sentiment output serves as the main explanatory variable in subsequent regression analyses.

A commonly used benchmark for this discretization is the VADER sentiment framework,[8] which defines neutrality as tone values between -0.05 and $+0.05$, classifying anything beyond that range as either positive or negative. More recent applications in financial contexts, such as Hasani and Zaw (2024), propose a stricter threshold of 0.20 to reduce ambiguity in classification. In the broader literature, threshold values as high as 0.33 are sometimes adopted.

Figure 5: Sensitivity of Sentiment Classification to Varying Thresholds



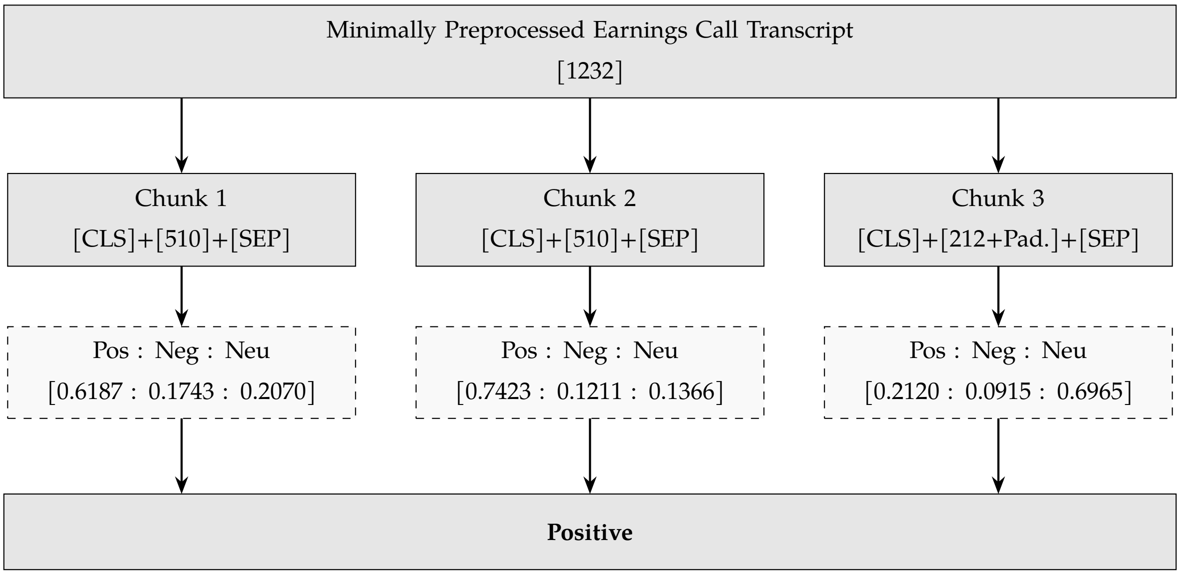
Notes: This figure shows how the number of transcripts classified as *Positive*, *Neutral*, or *Negative* evolves as the tone threshold T changes from 0.00 to 1.00. The green, red, and blue curves represent the number of documents assigned to each sentiment category at a given threshold. Vertical dashed lines indicate commonly used cutoff values in the literature: VADER (black), and three custom alternatives at 0.10 (ours), 0.20 (Hasani & Zaw, 2024), and 0.33. The figure provides a diagnostic tool to assess the trade-off between sensitivity and strictness in tone-based classification.

To inform our threshold choice, we perform a sensitivity analysis (Figure 5), which simulates how classification outcomes vary across the entire range of possible cutoff values. This diagnostic reveals the trade-off between sensitivity and strictness in classification. Based on these insights, we select 0.10 as the primary threshold for our analysis. This intermediate value balances two desirable properties: it avoids the overclassification that may arise with overly permissive thresholds (e.g., 0.05), while retaining enough sensitivity to meaningfully differentiate sentiment polarity in financial texts.

[8]The Valence Aware Dictionary and Sentiment Reasoner (VADER) uses a compound sentiment score, which combines lexical features and syntactic heuristics to produce a normalized measure of sentiment intensity. For classification purposes, VADER applies standard threshold cutoffs to determine whether a text is negative, neutral, or positive. See the VADER GitHub repository at github.com/cjhutto/vaderSentiment. Accessed in July, 2025.

3.3 FinBERT

Figure 6: FinBERT Implementation Pipeline



Notes: This figure illustrates the implementation pipeline for FinBERT-based sentiment analysis applied to a minimally preprocessed earnings call transcript. Due to the inclusion of special tokens (`[CLS]` at the beginning and `[SEP]` at the end), the text is divided into chunks of up to 510 tokens, resulting in a total of 512 tokens per chunk. In the example shown, the document contains 1232 tokens and is therefore split into three chunks. The first two chunks contain 510 tokens each, while the last chunk includes only 212 tokens and is padded to match the expected input size. Each chunk is independently processed by FinBERT, which outputs raw logits. These logits are converted into class probabilities (positive, negative, and neutral) using a softmax function. The final sentiment classification is determined by averaging the chunk-level probabilities. In this example, the transcript is classified as Positive.

FinBERT is a language model tailored for sentiment analysis in financial contexts. It builds upon the *Bidirectional Encoder Representations from Transformers* (BERT) architecture originally introduced by Devlin et al. (2019), which leverages self-attention mechanisms to jointly consider the words that appear both before and after a given target word. FinBERT adapts this architecture to financial language by undergoing additional pretraining and fine-tuning on labeled sentiment datasets (Araci, 2019). Its domain-specific training allows it to detect tone and polarity in financial texts more effectively than general-purpose language models.

Earnings call transcripts often exceed the 512-token input limit imposed by BERT-based models, which restricts the length of text that can be processed in a single forward pass. To address this constraint, two general strategies are available. One option is to summarize the transcript before classification. However, summarization risks omitting relevant information, introducing selection bias, and reducing transparency and reproducibility. As an alternative, we adopt a “chunk-and-aggregate” methodology, following the procedure used by Hasani and Zaw (2024).

In this approach, each transcript is divided into fixed-length chunks compatible with the model’s input constraints. These chunks are processed independently by FinBERT,[9] and

[9]The FinBERT model was accessed via the Hugging Face repository, using the `transformers` package in Python.

the resulting outputs are aggregated to yield a single sentiment classification for the full document. This method avoids the distortions associated with summarization while preserving the transcript's semantic content and allowing for the full exploitation of FinBERT's contextual capabilities within each chunk. This process is visually summarized in Figure 6.

The preprocessing begins by applying BERT's tokenizer with an option that returns PyTorch-native tensors (`return_tensors = 'pt'`). These tensors are then split into non-overlapping segments of 510 tokens using the `torch.split` function, reserving two positions for the special classification and separation tokens used by BERT-based models. Specifically, a `[CLS]` token is added to the beginning of each chunk to signal the start of the input, and a `[SEP]` token is added to the end to mark its boundary. If the last chunk contains fewer than 510 tokens, it is padded to reach the full 512-token length, ensuring consistent input dimensions across all chunks in the batch.[10]

All chunks are batched together and passed in parallel through the FinBERT model, which produces raw output logits for each chunk. These logits are converted into class probabilities using the softmax function:

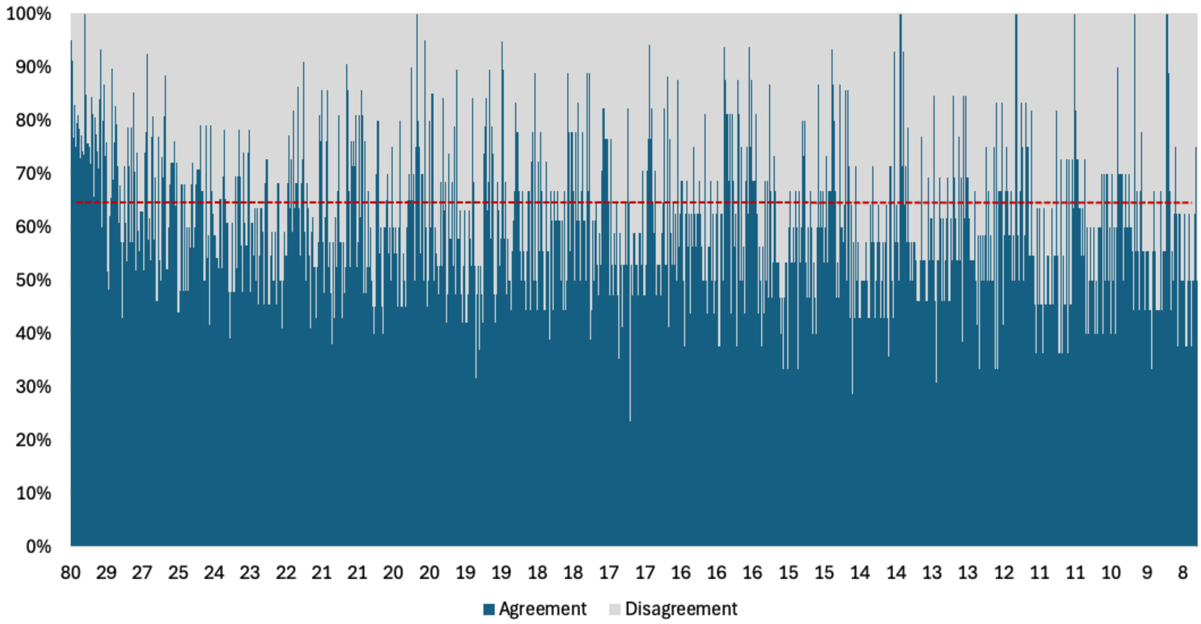
$$\text{softmax}(z_i) = \frac{e^{z_i}}{\sum_{j=1}^{K} e^{z_j}},$$

where z_i is the logit for class i, and $K = 3$ corresponds to the ternary sentiment classes: positive, negative, and neutral. The softmax transformation yields a probability distribution across these classes for each chunk. To generate a transcript-level classification, we compute the average of the chunk-level probability vectors and assign the sentiment corresponding to the highest mean probability.

FinBERT offers important advantages over dictionary-based approaches by modeling contextual dependencies and capturing sentiment at a higher semantic level. Unlike general-purpose models such as the *Generative Pre-trained Transformer* (GPT), FinBERT is fine-tuned on financial texts, enhancing its ability to detect domain-specific tone and terminology. However, the need to segment transcripts into chunks can lead to some loss of cross-sentence context. To assess the impact of this constraint, we evaluate the consistency between chunk-level predictions and the final transcript-level classification. Unified documents in our sample are typically divided into 10 to 25 chunks, with a peak at 16. On average, 65% of a transcript's chunks receive the same label as the final classification. This relatively high level of agreement suggests that the chunk-and-aggregate methodology remains reliable, despite the trade-off between input length and contextual continuity. The distribution of agreement rates is shown in Figure 7.

[10]For more information on the chunking procedure used with FinBERT, see the implementation available at github.com/rohan-paul. Accessed in May, 2025.

Figure 7: FinBERT Chunk Agreement



Notes: This figure shows the proportion of chunks within each transcript that share the same sentiment label as the final aggregated classification (agreement, in blue) versus those that do not (disagreement, in gray). The analysis is based on the unified document, i.e., the entire transcript without separating the Presentation and Q&A sections. Transcripts are ordered by descending number of chunks, from left to right. The red horizontal line indicates the average within-transcript agreement across the sample, which is approximately 65%. This metric helps assess the consistency of FinBERT's predictions under the chunk-and-aggregate approach.

3.4 GPT

Generative Pre-trained Transformer (GPT) is a family of autoregressive language models that generate text by predicting the next token in a sequence based on preceding context (T. Brown et al., 2020). Its unidirectional architecture restricts information flow to only previous tokens when computing probabilities. While this design is optimal for text generation, GPT can be adapted to classification tasks through carefully constructed prompts.

For this analysis, we implement a zero-shot prompting strategy based on Hasani and Zaw (2024), using OpenAI's model accessed through the *Application Programming Interface* (API). The prompt consists of two roles: a system message that instructs the model to act as an expert in sentiment and financial language, and a user message that provides the full transcript and asks for a sentiment label. The model is constrained to reply with exactly one word: *Positive*, *Negative*, or *Neutral*. The prompt structure is shown in Figure 8. While we closely follow the structure used by Hasani and Zaw (2024), we adopt a slightly streamlined version of the instruction.[11]

[11] We omit the word "please" from the prompt, following guidance from OpenAI CEO Sam Altman, who noted that it uses more energy without offering any apparent advantage. See x.com/sama/status/. Accessed in May, 2025.

Figure 8: Prompt Structure for OpenAI's GPT-4o Mini via API

GPT-4o Mini

[1] Role: 'System':
As an AI with expertise in language and emotion analysis, your task is to analyse the sentiment of text. Consider the financial context and setting in the text.

[2] Role: 'User':
Label the sentiment of the given text. Your answer should be exactly 'Positive', 'Neutral', or 'Negative':

`{Transcript text}`

[3] <Output>: Positive / Negative / Neutral

Notes: This figure shows the structured prompt used to query GPT-4o Mini for sentiment classification. The prompt is composed of a system role that defines the task, analyzing sentiment with attention to financial context, and a user role that provides the input text and specifies the expected response format. The model is instructed to classify the input as either "Positive", "Negative", or "Neutral". The output is a single-token label based on the model's interpretation of the sentiment expressed in the transcript.

Each transcript is processed through a single API call using the following key parameters: `model = "gpt-4o-mini"`,[12] `temperature = 0`, and `max_tokens = 1`. Setting the temperature to zero makes the model's outputs more predictable and consistent across runs, [13] while the `max_tokens=1` setting constrains the response to a single sentiment label (*Positive*, *Negative*, or *Neutral*) as required by the classification task. The model has a context window of 128K tokens,[14] allowing the processing of long transcripts in a single pass. This implementation offers an effective and scalable solution for classifying large volumes of financial text. GPT-4o Mini provides sufficient natural language understanding for sentiment classification tasks while remaining significantly more cost-effective[15] than larger models like GPT-4 or GPT-5. The use of zero-shot prompting[16] eliminates the need for labeled training data and enables rapid deployment across the entire corpus of earnings call transcripts.

[12] We use the fixed model version `gpt-4o-mini-2024-07-18` to ensure better reproducibility. OpenAI models are periodically updated, even under the same model name, which can lead to differences in behavior over time. Specifying a static version guarantees that the same model snapshot is used, allowing future researchers to replicate the results under similar conditions.

[13] Despite its stabilizing effect, a temperature of zero does not guarantee absolute determinism, as subtle variations can still occur across runs due to computational non-determinism and underlying model processes. See Atil et al. (2024).

[14] See openai.com/gpt-4o-mini for full model specifications. Accessed in July, 2025.

[15] As of August 2025, OpenAI's `gpt-4o-mini` costs USD 0.15 per million input tokens and USD 0.60 per million output tokens under the Standard plan (see Pricing - OpenAI API). With `max_tokens=1`, costs depend almost entirely on input length. The united-documents classification totaled USD 2.06, and the separated-documents classification USD 2.11, with slight additional charges from repeated runs.

[16] Zero-shot prompting refers to querying a language model without providing any labeled examples in the prompt. The model relies entirely on its pre-trained knowledge and the task instruction to generate an output. Alternatives include few-shot prompting, which includes labeled examples, and fine-tuning, which requires training the model on a labeled dataset.

4 Data and Variables

4.1 Textual Data

In this section, we present the main data sources and the variables used in the empirical exercise. Additional commentary can be found in the Appendix B.

4.1.1 Earnings Call Transcripts

Earnings Call Transcripts (ECTs) are comprehensive textual records of the discussions that take place during earnings calls, which are regularly scheduled meetings through which publicly traded companies communicate their quarterly results and strategic outlook to external stakeholders, including investors, analysts, and the media. These calls are typically divided into two main segments: the *Presentation* section, where executives deliver prepared remarks and present financial results, and the *Q&A* section, during which analysts pose questions to the management team. ECTs are a rich source of unstructured data that reflects not only quantitative disclosures but also qualitative signals such as managerial tone, forward-looking statements, and linguistic patterns. Although the transcripts can be accessed individually through company Investor Relations (IR) websites, we opted to download the transcripts directly from *LSEG Workspace* to ensure standardized formatting and broad coverage. Table 1 summarizes the availability of transcripts across firms and years.

4.2 Numerical Data

4.2.1 Stock Prices and Firm Metrics

We downloaded from *LSEG Workspace* the daily stock prices (close) used to construct the outcome variable, as well as the quarterly firm-level metrics employed as control variables in the empirical analysis, including return on assets (ROA), market capitalization, book-to-market ratio (B/M), and stock turnover. We also constructed a measure of analyst coverage by extracting the number of analysts participating in each earnings call directly from the header section of the transcripts.[17] The industry classification used to build the sector dummies was obtained directly from the B3 website,[18] and the industry breakdown for our sample is presented in Table 2.

[17]We opted to construct analyst coverage by counting the number of analysts participating in each earnings call, as reported in the transcript header, due to the limited availability of earnings forecasts in LSEG Workspace and Bloomberg. Many firms had few or no analyst estimates, which would have significantly reduced the usable sample or introduced noise into the model. Our approach represents a contribution of this work.

[18]Available at www.b3.com.br. Accessed in February, 2025.

Table 1: Top 30 IBovespa B3 Index Companies Transcript Availability

IBov Part. (%)	Company	Start	Available	Success Rate (%)
1.53	Embraer	Q2 2002	90	99%
12.45	Petrobras	Q2 2002	89	98%
2.52	Ambev	Q1 2003	87	99%
11.33	Vale	Q4 2002	86	99%
7.28	Itaú Unibanco	Q2 2003	85	98%
1.01	Gerdau	Q2 2003	85	98%
1.55	Suzano	Q4 2003	83	98%
4.51	Bradesco	Q2 2003	82	94%
1.13	Ultrapar	Q3 2003	79	92%
0.96	CEMIG	Q2 2003	76	87%
0.95	BRF *	Q4 2009	60	98%
3.86	Eletrobras	Q4 2009	57	93%
1.78	JBS *	Q2 2011	54	98%
1.57	RD Saúde	Q1 2011	54	96%
0.94	Telefônica BR	Q3 2011	54	100%
3.56	Banco do Brasil	Q4 2011	53	100%
3.06	WEG	Q4 2011	53	100%
1.80	Localiza	Q4 2011	53	100%
1.67	PRIO	Q1 2012	48	92%
2.04	BTG Pactual	Q4 2013	45	100%
1.90	Eneva	Q4 2011	45	85%
2.88	Sabesp	Q2 2011	43	78%
1.90	Equatorial	Q1 2012	38	73%
1.04	BB Seguridade	Q3 2015	37	97%
1.18	Rumo	Q1 2016	36	100%
3.11	B3	Q2 2017	31	100%
1.17	Vibra	Q3 2018	26	100%
0.90	Hapvida	Q2 2019	23	100%
2.68	ITAÚSA	Q2 2018	13	48%
1.63	Rede D'Or São Luiz	Q3 2022	10	100%
Total	**30 Companies**	**-**	**1675**	**95%**

Notes: This table lists the 30 companies with the highest weights in the IBovespa B3 Index (theoretical portfolio valid from September to December 2024), ranked by transcript availability. For each company, we report its participation percentage (%) in the index, the quarter and year when transcripts began to be consistently available on *LSEG Workspace*, the total number of transcripts retrieved, and the corresponding success rate. The start date refers to the onset of consistent availability on the platform, not to the company's first earnings call. The sample includes a total of 1,675 transcripts available with an overall success rate of 95%, with individual company success rates ranging from 48% to 100%.

(*) Our sample goes until 2024 and therefore includes BRF prior to the announced 2025 merger with Marfrig to form MBRF, and JBS prior to its delisting from the B3 in June 2025 as part of a reorganization to create JBS NV, which will trade on the NYSE and have BDRs on the B3. For details on how we handled name changes, mergers, and acquisitions, see Appendix A.1.

Table 2: Sample's Industry Breakdown

B3 Economic Sector	# Companies
Communications	1
Consumer Cyclical	1
Basic Materials	3
Capital Goods and Services	3
Consumer non Cyclical	3
Health	3
Oil, Gas and Biofuels	4
Utilities	5
Financial	7
9 sectors	30 companies

Notes: This table displays the distribution of sampled companies across the B3 economic sectors. The B3 sector classification was created to provide clear identification of company sectors from the first level of the structure. This classification system permits an overview of companies that, even though performing different activities, belong to the same production chain or produce related products/services and show similar responses to economic conditions. The data shows that the Financial sector has the highest representation (23.3%) in our sample, followed by Utilities (16.7%) and Oil, Gas and Biofuels (13.3%), while Communications and Consumer Cyclical have the lowest representation (3.3% each).

4.2.2 Outcome Variable

Cumulative Abnormal Returns (CAR) refers to the sum of abnormal returns experienced by a stock over a specific period, in this case, around the earnings call event:

$$AR_{it} = R_{it} - \mathbb{E}[R_{it}] \quad \Longrightarrow \quad CAR_i[j_1, j_2] = \sum_{t=j_1}^{j_2} AR_{it}. \tag{5}$$

Our counterfactuals, i.e., the expected or "normal" returns $(E[R_{i,t}])$, are estimated using an asset pricing model framework following the approach of Fama and French (1993) and Jegadeesh and Titman (1993):

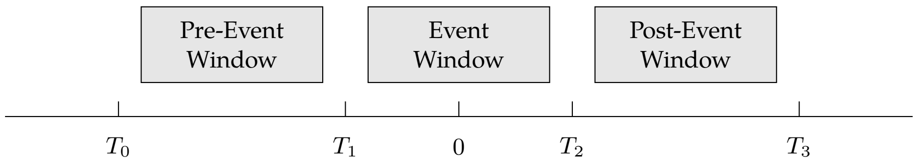
$$\begin{aligned} E[R_{it}] = \alpha_i + R_{f,t} + \beta_{i,M}(R_{M,t} - R_{f,t}) \\ + \beta_{i,SMB} \cdot SMB_t + \beta_{i,HML} \cdot HML_t + \beta_{i,WML} \cdot WML_t \end{aligned} \tag{6}$$

where $R_{M,t}$ denotes the return on the market portfolio, $R_{f,t}$ is the risk-free rate, SMB_t captures the size premium (Small Minus Big), HML_t represents the value premium (High Minus Low), and WML_t reflects the momentum effect (Winners Minus Losers). All factor series were retrieved from NEFIN's website.[19] The parameters of the model were estimated over the $[-315, -62]$ trading days prior to the event date, following Arslan-Ayaydin, Boudt, and Thewissen (2016).

The timeline defining the event study and the Asset Pricing Model parameters estimations windows is illustrated in Figure 9. For instance, $CAR[-1, +1]$ captures the three-day market reaction spanning from one day before through one day after the earnings call, isolating the immediate impact of the disclosure. Meanwhile, $CAR[+2, +61]$ tracks the longer-term

[19] Available at nefin.com.br. Accessed in March, 2025. More information in Appendix B.

Figure 9: Timeline of the Event Study and Asset Pricing Model Estimation Window



Notes: This figure illustrates the structure of the event study, dividing the sample period into a pre-event window, the event window itself, and a post-event window. The pre-event window is used to establish a baseline, the event window captures the immediate market reaction to the disclosure, and the post-event window tracks longer-term price adjustments. In addition, the parameters of the asset pricing model are estimated over a separate window spanning from 315 to 62 trading days before the event date.

price drift over approximately three trading months following the announcement, revealing how markets continue to process the earnings information after the initial reaction period. Our analysis centers on the immediate return window.

5 Empirical Evidence

5.1 Unified Documents

We begin by estimating the following panel data model using the unified version of each earnings call transcript, where the entire document is treated as a single unit of analysis. The goal is to evaluate the relationship between sentiment and short-term stock market reactions:

$$CAR[-1,+1]_{iqt} = \alpha + \beta_1\, Pos_{iqt} + \beta_2\, Neg_{iqt} + \boldsymbol{\gamma}'\, Controls_{iqt} + u_{iqt} \tag{7}$$

where the dependent variable $CAR[-1,+1]_{iqt}$ denotes the immediate cumulative abnormal return for firm i in quarter q of year t, and the key explanatory variables are dummy indicators for positive (Pos_{iqt}) and negative (Neg_{iqt}) sentiment in the transcript. The vector $Controls_{iqt}$ includes firm-level and informational variables following Arslan-Ayaydin et al. (2016): [20] ROA, $\ln$(Market Cap), $\ln$(B/M), $\ln$(1 + Coverage), and $\ln$(Turnover). We also include sector, year, and quarter dummies to absorb systematic variation across firms and reporting periods. The model is estimated separately for each of the three sentiment classification approaches: LM dictionary, FinBERT, and GPT. The results are reported in Table 3. Standard errors are two-way clustered by firm and time.

Across the three sentiment classification methods, the estimated coefficients consistently indicate that positive sentiment is associated with positive immediate cumulative abnormal returns, while negative sentiment is linked to negative returns. Nonetheless, the magnitude and statistical significance of these effects vary across methods. The LM dictionary shows a statistically significant increase in returns for Pos_{iqt}, whereas the corresponding coefficient for

[20]While several prior studies, including Arslan-Ayaydin et al. (2016), incorporate earnings surprise as a control, our attempts to include such a measure were limited by data availability in LSEG Workspace and Bloomberg. The number of analyst estimates was insufficient or missing for a substantial portion of the sample, which would have significantly reduced the number of usable observations or introduced noise.

Neg_{iqt} is negative but not significant. FinBERT yields coefficients that align in sign with these results but are smaller in magnitude and lack statistical significance. The GPT-based estimates display the largest absolute effects, with negative sentiment showing a strong and significant reduction in returns, while the positive sentiment effect is larger than that of the LM dictionary but only marginally significant. These results suggest that although all methods agree on the direction of sentiment's impact, the strength and size of the estimated effects depend on the classification approach.

Table 3: Impact of Sentiment on Immediate Stock Returns using Unified Documents: Full Sample (2002-2024)

Dependent Variable: Immediate Cumulative Abnormal Returns (CAR)					
	(1)	(2)	(3)	(4)	(5)
Intercept	0.0433	0.0493*	0.0371	0.0365	0.0486*
	(0.0277)	(0.0280)	(0.0296)	(0.0305)	(0.0271)
Pos_LM	0.0071***			0.0071***	
	(0.0021)			(0.0025)	
Neg_LM	-0.0050			-0.0032	
	(0.0047)			(0.0044)	
Pos_FinBERT		0.0015		-0.0022	
		(0.0031)		(0.0027)	
Neg_FinBERT		-0.0113		-0.0065	
		(0.0120)		(0.0124)	
Pos_GPT			0.0097*	0.0076	
			(0.0055)	(0.0051)	
Neg_GPT			-0.0193***	-0.0197***	
			(0.0053)	(0.0049)	
Controls	Yes	Yes	Yes	Yes	Yes
Industry Dummies	Yes	Yes	Yes	Yes	Yes
Year Dummies	Yes	Yes	Yes	Yes	Yes
Quarter Dummies	Yes	Yes	Yes	Yes	Yes
N	1385	1385	1385	1385	1385
R^2	0.0376	0.0329	0.0393	0.0431	0.0323
Adjusted R^2	0.0082	0.0034	0.0100	0.0110	0.0042
Sample Period	2002-2024	2002-2024	2002-2024	2002-2024	2002-2024

Notes: This table reports the results from panel regressions analyzing the impact of sentiment on cumulative abnormal returns (CAR), using the unified version of each earnings call transcript, i.e., treating the full document as a single unit of analysis. The dependent variable is CAR_{iqt}, calculated over a three-day window around the call date (day $t = 0$, plus one day before and one day after). Columns (1)–(3) display results for the LM Dictionary, FinBERT, and GPT sentiment measures, respectively. Column (4) includes all sentiment measures jointly. Column (5) includes only control variables and no sentiment. For detailed information on control variable coefficients, see Appendix C.1. Standard errors are two-way clustered by firm and time. ***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively.

5.2 Subsample Estimation: Analysis from the Mass Adoption Phase

We next re-estimate the model described in Section 5.1 for the subsample with reference years from 2012 onward. The reference year corresponds to the fiscal year for which the earnings call reports results, rather than the calendar year in which the call itself takes place. Approximately 78% of the calls in our dataset refer to this period, when the adoption of earnings calls by sample firms becomes increasingly widespread. The specification remains identical to the unified-document regression outlined above, with the same set of controls and estimation procedures.

Table 4: Impact of Sentiment on Immediate Stock Returns using Unified Documents: Mass Adoption Phase (2012-2024)

Dependent Variable: Immediate Cumulative Abnormal Returns (CAR)					
	(1)	(2)	(3)	(4)	(5)
Intercept	0.0570**	0.0633**	0.0531*	0.0526*	0.0620**
	(0.0280)	(0.0283)	(0.0297)	(0.0295)	(0.0279)
Pos_LM	0.0070***			0.0071**	
	(0.0025)			(0.0029)	
Neg_LM	-0.0047			-0.0030	
	(0.0052)			(0.0048)	
Pos_FinBERT		0.0006		-0.0030	
		(0.0034)		(0.0028)	
Neg_FinBERT		-0.0122		-0.0071	
		(0.0122)		(0.0127)	
Pos_GPT			0.0101*	0.0081	
			(0.0060)	(0.0057)	
Neg_GPT			-0.0166***	-0.0171***	
			(0.0058)	(0.0056)	
Controls	Yes	Yes	Yes	Yes	Yes
Industry Dummies	Yes	Yes	Yes	Yes	Yes
Year Dummies	Yes	Yes	Yes	Yes	Yes
Quarter Dummies	Yes	Yes	Yes	Yes	Yes
N	1226	1226	1226	1226	1226
R^2	0.0355	0.0313	0.0371	0.0407	0.0307
Adjusted R^2	0.0104	0.0061	0.0121	0.0125	0.0072
Sample Period	2012-2024	2012-2024	2012-2024	2012-2024	2012-2024

Notes: This table reports the results from panel regressions analyzing the impact of sentiment on cumulative abnormal returns (CAR) for the mass adoption phase subsample (2012-2024), using the unified version of each earnings call transcript. The dependent variable is CAR_{iqt}, calculated over a three-day window around the call date (day $t = 0$, plus one day before and one day after). Columns (1)–(3) display results for the LM Dictionary, FinBERT, and GPT sentiment measures, respectively. Column (4) includes all sentiment measures jointly. Column (5) includes only control variables and no sentiment. For detailed information on control variable coefficients, see Appendix C.1. Standard errors are two-way clustered by firm and time. ***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively.

Building on the previous results, restricting the analysis to the post-2012 period, when the adoption of earnings calls among sample firms became widespread, does not alter the core directional relationships. Across all three sentiment classification methods, the direction of the coefficients remains unchanged, with positive sentiment associated with positive immediate cumulative abnormal returns and negative sentiment linked to negative returns. While some variation in magnitude is observed, particularly for the GPT-based estimates, the statistical significance of the coefficients is generally consistent with the full-sample results. Notably, the adjusted R^2 increases across all model specifications, indicating a modest improvement in explanatory power in the mass adoption phase, without fundamentally affecting the inference regarding the role of sentiment. This suggests that the expanded and more standardized use of earnings calls in this period does not materially alter the relationship between sentiment and short-term market reactions, but may improve overall model fit.

5.3 Divided Documents

5.3.1 Presentation and Q&A Sentiment

We then estimate a panel data model based on the divided version of the transcripts, in which each earnings call is split into two segments: the *Presentation* and the *Q&A*. This specification allows us to assess to evaluate the relationship between sentiment conveyed in each part of the call and short-term stock market reactions:

$$\begin{aligned} CAR[-1,+1]_{iqt} = \; & \alpha + \beta_1 \, Pos_{Pres_{iqt}} + \beta_2 \, Neg_{Pres_{iqt}} \\ & + \rho_1 \, Pos_{Q\&A_{iqt}} + \rho_2 \, Neg_{Q\&A_{iqt}} + \boldsymbol{\gamma}' \, \text{Controls}_{iqt} + u_{iqt} \end{aligned} \tag{8}$$

where CAR_{iqt} is the dependent variable and denotes the cumulative abnormal return for firm i in quarter q of year t. The explanatory variables include dummy indicators for positive and negative sentiment in both the Presentation section ($Pos_{Pres_{iqt}}$, $Neg_{Pres_{iqt}}$) and the Q&A section ($Pos_{QA_{iqt}}$, $Neg_{QA_{iqt}}$). The vector $Controls_{iqt}$ contains the same firm-level and informational controls as before, following Arslan-Ayaydin et al. (2016): $\ln$(Market Cap), ROA, $\ln$(B/M), $\ln$(Coverage), and $\ln$(Turnover). Sector, year, and quarter dummies are also included. The model is estimated separately for each of the three sentiment classification approaches: LM dictionary, FinBERT, and GPT. The results are presented in Table 5. Standard errors are two-way clustered by firm and time.

When using the divided version of the transcripts, results diverge more across sentiment measures than in the unified-document specification. Although the literature views the Q&A as more informative due to its unscripted nature, our estimates reveal only partial consistency across methods. For the LM dictionary, Q&A sentiment behaves largely as expected: positive sentiment is linked to higher returns and negative sentiment to lower returns, both statistically significant in several cases. FinBERT shows a strong and significant negative effect for negative sentiment but no significant impact for positive sentiment. GPT yields the largest and most robust positive effect for positive sentiment, while the negative sentiment coefficient is small and insignificant. Presentation sentiment effects are generally weaker, less consistent, and often insignificant. Overall, the evidence suggests that sentiment in the Q&A carries more informational weight than in the Presentation, but the magnitude and significance of these effects remain sensitive to the sentiment classification method.

Table 5: Impact of Sentiment on Immediate Stock Returns using Divided Documents

Dependent Variable: Immediate Cumulative Abnormal Returns (CAR)					
	(1)	(2)	(3)	(4)	(5)
Intercept	0.0561	0.0472	0.0230	0.0396	0.0486
	(0.0500)	(0.0435)	(0.0420)	(0.0516)	(0.0419)
Pos_pres_LM	-0.0096			-0.0162	
	(0.0087)			(0.0091)	
Neg_pres_LM	-0.0254			-0.0158	
	(0.0185)			(0.0179)	
Pos_qa_LM	0.0094***			0.0047	
	(0.0035)			(0.0039)	
Neg_qa_LM	-0.0200**			-0.0143*	
	(0.0100)			(0.0086)	
Pos_pres_FinBERT		0.0066*		0.0035	
		(0.0036)		(0.0034)	
Neg_pres_FinBERT		0.0019		0.0010	
		(0.0088)		(0.0106)	
Pos_qa_FinBERT		-0.0002		-0.0049	
		(0.0058)		(0.0072)	
Neg_qa_FinBERT		-0.0474***		-0.0274*	
		(0.0122)		(0.0166)	
Pos_pres_GPT			0.0071	0.0084	
			(0.0073)	(0.0071)	
Neg_pres_GPT			-0.0226*	-0.0231*	
			(0.0136)	(0.0130)	
Pos_qa_GPT			0.0169***	0.0154**	
			(0.0054)	(0.0060)	
Neg_qa_GPT			-0.0027	0.0005	
			(0.0078)	(0.0076)	
Controls	Yes	Yes	Yes	Yes	Yes
Industry Dummies	Yes	Yes	Yes	Yes	Yes
Year Dummies	Yes	Yes	Yes	Yes	Yes
Quarter Dummies	Yes	Yes	Yes	Yes	Yes
N	1385	1385	1385	1385	1385
R^2	0.0418	0.0353	0.0560	0.0632	0.0323
Adjusted R^2	0.0111	0.0043	0.0257	0.0274	0.0042
Sample Period	2002-2024	2002-2024	2002-2024	2002-2024	2002-2024

Notes: This table reports the results from panel regressions analyzing the impact of sentiment on cumulative abnormal returns (CAR), using the divided version of each earnings call transcript, i.e., separating the document into Presentation and Q&A sections. The dependent variable is CAR_{iqt}, calculated over a three-day window centered on the call date (day $t = 0$, plus one day before and after). Columns (1)–(3) display estimates for the LM Dictionary, FinBERT, and GPT sentiment measures, respectively. Column (4) includes all sentiment measures jointly, while Column (5) includes only control variables and no sentiment terms. For detailed information on control variable coefficients, see Appendix C.1. Standard errors are two-way clustered by firm and time. ***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively.

5.3.2 Q&A Sentiment Only

In a further step, we restrict the specification to include only the sentiment extracted from the *Q&A* segment of the earnings call as explanatory variables. This design isolates the effect of unscripted exchanges on short-term market reactions, in line with the literature that emphasizes the Q&A as a richer source of value-relevant information compared to the scripted *Presentation*. The model is estimated separately for each sentiment classification approach (LM dictionary, FinBERT, and GPT):

$$CAR[-1,+1]_{iqt} = \alpha + \rho_1\, Pos_{Q\&A_{iqt}} + \rho_2\, Neg_{Q\&A_{iqt}} + \gamma' \text{Controls}_{iqt} + u_{iqt} \tag{9}$$

where $Pos_{Q\&A_{iqt}}$ and $Neg_{Q\&A_{iqt}}$ are dummy indicators for positive and negative sentiment in the *Q&A* section of the call, respectively. The specification includes the same firm-level and informational controls as in prior estimations, along with sector, year, and quarter dummies. Standard errors are two-way clustered by firm and time.

Table 6: Impact of Sentiment on Immediate Stock Returns using Q&A Section Only

Dependent Variable: Immediate Cumulative Abnormal Returns (CAR)					
	(1)	(2)	(3)	(4)	(5)
Intercept	0.0440	0.0495*	0.0301	0.0309	0.0486*
	(0.0287)	(0.0269)	(0.0229)	(0.0231)	(0.0271)
Pos_qa_LM	0.0091***			0.0048	
	(0.0033)			(0.0039)	
Neg_qa_LM	-0.0179***			-0.0131*	
	(0.0068)			(0.0070)	
Pos_qa_FinBERT		0.0009		-0.0043	
		(0.0068)		(0.0070)	
Neg_qa_FinBERT		-0.0474***		-0.0329*	
		(0.0112)		(0.0169)	
Pos_qa_GPT			0.0177***	0.0160***	
			(0.0047)	(0.0053)	
Neg_qa_GPT			-0.0074*	-0.0053*	
			(0.0043)	(0.0030)	
Controls	Yes	Yes	Yes	Yes	Yes
Industry Dummies	Yes	Yes	Yes	Yes	Yes
Year Dummies	Yes	Yes	Yes	Yes	Yes
Quarter Dummies	Yes	Yes	Yes	Yes	Yes
N	1385	1385	1385	1385	1385
R^2	0.0400	0.0332	0.0511	0.0546	0.0323
Adjusted R^2	0.0107	0.0037	0.0222	0.0228	0.0042
Sample Period	2002-2024	2002-2024	2002-2024	2002-2024	2002-2024

Notes: This table reports the results from panel regressions analyzing the impact of sentiment on cumulative abnormal returns (CAR), using only the Q&A section of each earnings call transcript as explanatory variables. The dependent variable is CAR_{iqt}, calculated over a three-day window centered on the call date (day $t = 0$, plus one day before and after). Columns (1)–(3) display estimates for the LM Dictionary, FinBERT, and GPT sentiment measures, respectively. Column (4) includes all sentiment measures jointly, while Column (5) includes only control variables and no sentiment terms. For detailed information on control variable coefficients and intercept, see Appendix C.1. Standard errors are two-way clustered by firm and time. ***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively.

The results, presented in Table 6, indicate that the LM dictionary produces coefficients aligned with theoretical expectations: positive Q&A sentiment is associated with higher cumulative abnormal returns and negative sentiment with lower returns, both statistically significant. FinBERT exhibits a strong and significant negative association for negative sentiment, while the coefficient for positive sentiment remains small and statistically insignificant. GPT produces the largest positive coefficient for positive sentiment in this specification, coupled with a statistically significant negative effect for negative sentiment in most specifications.

Compared to the divided-document results, focusing solely on the Q&A sentiment increases the clarity of the signal for certain methods, particularly GPT and the LM dictionary, while preserving the overall pattern that sentiment in unscripted exchanges plays a key role in shaping immediate market reactions. However, the magnitude and statistical significance of these effects remain sensitive to the sentiment extraction methodology, reinforcing the view that both the section of the transcript analyzed and the model used are critical determinants of empirical findings.

6 Conclusion and Discussion

This paper investigates the relationship between corporate sentiment, extracted from earnings call transcripts (ECTs) of Brazilian firms, and immediate stock market reactions. By employing three distinct natural language processing methodologies, a dictionary-based approach (Loughran & McDonald, 2011), a domain-specific transformer model (FinBERT), and a generative large language model (GPT), we provide a comparative analysis of how different sentiment measures explain cumulative abnormal returns (CARs). Our findings yield several key insights.

First, in the unified document specification, the sentiment in ECTs consistently shows the same directional relationship with short-term stock returns across all three methods: positive sentiment is associated with higher CARs, and negative sentiment with lower CARs. However, the statistical significance and magnitude of these effects vary across methods. The dictionary-based approach detects a significant market reaction to positive tone but fails to capture a comparable response to negative sentiment. GPT identifies both a positive association for positive sentiment and a statistically significant negative association for negative sentiment, whereas FinBERT's coefficients, although aligned in sign with these patterns, are not statistically significant. Restricting the sample to the post-2012 mass adoption phase does not materially alter these conclusions, although the adjusted R^2 values improve across specifications, indicating a modest gain in explanatory power.

While GPT-based sentiment measures demonstrate notable explanatory power for short-term stock returns in certain specifications, part of this performance may be attributable to the way large language models (LLMs) are trained. GPT-4o Mini, for example, has a knowledge cutoff in October 2023, creating the possibility that classifications for historical earnings calls could be indirectly influenced by knowledge of subsequent events. This look-ahead bias undermines causal interpretation, as the sentiment score may embed information not available to market participants at the time of the call. Moreover, the "black box" nature of these models makes it difficult to determine the extent to which their classifications differ from those of models such as FinBERT because of superior linguistic analysis, as opposed to the influence of pre-training cutoffs and the specific materials used during pre-training and fine-tuning.

Second, the performance gap between sentiment measures reflects a broader methodological shift in financial text analysis. Transformer-based models represent an advance over lexicon-based approaches by capturing context, syntax, and linguistic nuance, capabilities that are essential for interpreting complex corporate disclosures. However, when using large language models (LLMs), this advantage must be weighed against the potential for look-ahead bias, which can limit their applicability in real-world tasks by undermining the causal interpretation of results. At the same time, dictionary-based methods such as the Loughran and McDonald (LM) dictionary face their own challenges, as they rely on fixed word lists and struggle with negation or phrasing, also limiting their effectiveness in practice.

These limitations become particularly salient in international applications. As shown by Pinto and Siano (2025), dictionaries developed in the U.S. context, such as the Loughran and McDonald (2011) word list, tend to underperform when applied to earnings call transcripts from firms in culturally and institutionally distinct settings like Brazil. Their study demonstrates that cultural distance can significantly weaken the interpretive validity of such lexicons, even when the disclosure language is English. This helps explain our own findings, in which the LM dictionary-based sentiment measure exhibits weaker explanatory power in some specifications compared to more adaptive, context-aware approaches such as FinBERT and GPT.

Furthermore, dictionary-based and traditional machine learning methods exhibit pronounced sensitivity to text preprocessing choices, whereas transformer architectures are generally less dependent on extensive preprocessing and can operate effectively on raw or minimally processed text. This characteristic represents a significant operational advantage, as it reduces the need for analyst intervention and lowers the risk of errors arising from excessive or inadequate text cleaning. The ability to handle unprocessed text also helps preserve important contextual information that might otherwise be lost through aggressive preprocessing.

Third, in the divided-document specification, we find that the *Q&A* segment is more strongly related to immediate stock returns than the scripted *Presentation*, consistent with the literature that views unscripted exchanges as more informative. For the LM dictionary, positive Q&A sentiment is linked to higher CARs and negative sentiment to lower CARs, both significant. FinBERT shows a robust and significant negative effect for negative Q&A sentiment but no meaningful effect for positive sentiment. GPT delivers the largest and most consistent positive effect for Q&A sentiment, while its negative coefficient remains small and insignificant. In contrast, Presentation sentiment effects are generally weaker, less consistent in sign, and rarely significant. Overall, these findings indicate that investor reactions are primarily shaped by the spontaneous exchanges in the Q&A, although the size and significance of these effects vary across sentiment classification methods.

Going forward, an important avenue for research is the integration of multimodal data. As emphasized by Todd et al. (2024), sentiment analysis of earnings calls should prioritize the expansion and refinement of models that combine textual and vocal cues to better capture tone and emotion. Incorporating vocal features can enrich the analysis by revealing nuances that are often absent from written transcripts. This integration has the potential to enhance the detection of subtle shifts in sentiment and improve the explanatory power of empirical models in financial contexts. At the same time, further research is required to design and validate methods that more effectively address and quantify the extent of look-ahead bias, ensuring that sentiment measures reflect only the information available to market participants at the time of disclosure.

References

Araci, D. (2019). Finbert: Financial sentiment analysis with pre-trained language models. *CoRR*, *abs/1908.10063*. Retrieved from http://arxiv.org/abs/1908.10063

Arslan-Ayaydin, Ö., Boudt, K., & Thewissen, J. (2016). Managers set the tone: Equity incentives and the tone of earnings press releases. *Journal of Banking & Finance*, *72*, S132–S147.

Atil, B., Aykent, S., Chittams, A., Fu, L., Passonneau, R. J., Radcliffe, E., ... others (2024). Non-determinism of" deterministic" llm settings. *arXiv preprint arXiv:2408.04667*.

Brown, S., & Warner, J. (1985). Using daily stock returns: The case of event studies. *Journal of financial economics*, *14*(1), 3–31.

Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., ... others (2020). Language models are few-shot learners. *Advances in Neural Information Processing Systems*, *33*, 1877–1901.

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). Bert: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 conference of the north american chapter of the association for computational linguistics: human language technologies, volume 1 (long and short papers)* (pp. 4171–4186).

Fama, E. F., & French, K. R. (1993). Common risk factors in the returns on stocks and bonds. *Journal of financial economics*, *33*(1), 3–56.

Ferreira, F. R., Fiorot, D. C., Motoki, F. Y. S., & Moreira, N. C. (2019). Voluntary disclosure: Empirical analysis of the tone used in conference calls. *Revista de Administração de Empresas*, *59*, 271–283.

Firth, J. (1957). A synopsis of linguistic theory, 1930-1955. In J. Firth (Ed.), *Studies in linguistic analysis.* Oxford, UK: Basil Blackwell.

Frankel, R., Johnson, M., & Skinner, D. J. (1999). An empirical examination of conference calls as a voluntary disclosure medium. *Journal of Accounting Research*, *37*(1), 133–150.

Gentzkow, M., Kelly, B., & Taddy, M. (2019). Text as data. *Journal of Economic Literature*, *57*(3), 535–574.

Hasani, P., & Zaw, M. S. T. A. (2024). *Beyond words: An applied study of sentiment analysis on scandinavian earnings call transcripts comparing a traditional lexicon, machine learning, finbert, and gpt-4 turbo* (Unpublished master's thesis). NORWEGIAN SCHOOL OF ECONOMICS.

Jegadeesh, N., & Titman, S. (1993). Returns to buying winners and selling losers: Implications for stock market efficiency. *The Journal of finance*, *48*(1), 65–91.

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

Li, K., Mai, F., Shen, R., & Yan, X. (2021). Measuring corporate culture using machine learning. *The Review of Financial Studies*, *34*(7), 3265–3315.

Liu, B. (2020). *Sentiment analysis: Mining opinions, sentiments, and emotions*. Cambridge University Press.

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

Loughran, T., & McDonald, B. (2016). Textual analysis in accounting and finance: A survey.

Journal of Accounting Research, *54*(4), 1187–1230.

MacKinlay, A. C. (1997). Event studies in economics and finance. *Journal of Economic Literature*, *35*(1), 13–39.

Matsumoto, D., Pronk, M., & 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.

Mayew, W. J., & Venkatachalam, M. (2012). The power of voice: Managerial affective states and future firm performance. *The Journal of Finance*, *67*(1), 1–43.

Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). *Efficient estimation of word representations in vector space.* Retrieved from https://arxiv.org/abs/1301.3781

Moreira, N. C., Ramos, F., Kozak-Rogo, J., & Rogo, R. (2016). Conference calls: An empirical analysis of information content and the type of disclosed news. *BBR. Brazilian Business Review*, *13*, 291–315.

Pennington, J., Socher, R., & Manning, C. D. (2014). Glove: Global vectors for word representation. In *Proceedings of the 2014 conference on empirical methods in natural language processing (emnlp)* (pp. 1532–1543).

Pinto, J., & Siano, F. (2025). Beyond words: The challenge of measuring financial text across borders. *Journal of International Business Studies, Forthcoming*.

Todd, A., Bowden, J., & Moshfeghi, Y. (2024). Text-based sentiment analysis in finance: Synthesising the existing literature and exploring future directions. *Intelligent Systems in Accounting, Finance and Management*, *31*(1), e1549.

Tonin, J. M. d. F., & Scherer, L. M. (2021). Market reaction to the tones of earnings conference calls. *Revista de Administração de Empresas*, *62*, e2020–0301.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... Polosukhin, I. (2017). Attention is all you need. *Advances in neural information processing systems*, *30*.

Viana Junior, D. B. C., Arruda Castro, L., Rodrigues Ponte, V. M., & Chagas Lima, M. (2019). Do the words matter? an analysis of the linguistic sentiment of earnings conference calls and abnormal stock returns in brazilian companies. *Advances in Scientific & Applied Accounting*, *12*(3).

Zhao, H., Liu, Z., Wu, Z., Li, Y., Yang, T., Shu, P., ... Liu, T. (2024). *Revolutionizing finance with llms: An overview of applications and insights.* Retrieved from https://arxiv.org/abs/2401.11641

A More on ECTs Preprocessing, NLP Techniques and Sentiment Measures

A.1 Raw ECT Retrieval and Descriptive Temporal Patterns

We obtain Earnings Call Transcripts (ECTs) from the *Thomson Reuters Street Events* (SE) database. This dataset was originally branded under Thomson Reuters. Following its acquisition of Refinitiv, the *London Stock Exchange Group* (LSEG) integrated the platform into its suite of products. In 2024, LSEG officially retired the Refinitiv brand and rebranded the platform as *LSEG Workspace*, consolidating its financial data interfaces, including the former Eikon terminal, under a single name.[21] Although earnings call transcripts can also be individually accessed through each company's investor relations website (typically in PDF format, as illustrated in Figure A.10), we opted to use the standardized data available through *LSEG Workspace*. This choice offers important advantages for large-scale textual analysis, particularly due to the consistent formatting across firms and the availability of files in plain text (.txt) format (as shown in Figure A.11), which significantly facilitates automated processing.

However, despite these advantages, the use of *LSEG Workspace* transcripts was not without limitations. A small number of files in the dataset were entirely empty in terms of substantive content: while they retained the standard header and footer sections, both the presentation and Q&A segments were missing. Examples included the transcripts for Itaú Unibanco Holding on August 3, 2005, Ultrapar on May 15, 2009, and RD Saúde on November 10, 2011. In addition, a few transcripts presented structural or metadata inconsistencies. In most of these cases, the event title did not explicitly state the reference quarter, as observed in Petrobras for Q3 2009 and Q1 2007, and Ultrapar for Q4 2004, requiring inference from the content of the discussion. In one instance, the title explicitly indicated the quarter but was incorrect: the Suzano transcript dated May 14, 2021, was labeled as "Q2 2020" while the discussion clearly referred to Q1 2021 results. Although these occurrences represented a very small share of the dataset, they required manual verification and, in some cases, correction before the transcripts could be reliably used in large-scale textual analysis.

As K. Li et al. (2021) note, *LSEG Workspace* retroactively updates company identifiers across historical records whenever a firm undergoes a name or ticker change, typically due to a merger or acquisition. This ensures alignment with the most recent entity. However, the original company name is often preserved in the event title of the earnings call (e.g., "Q4 2014 ALL America Latina Logistica SA Earnings Call"), which allows researchers to trace corporate identity over time and mitigate distortions caused by this backfilling process.

[21] More information available at standard.co.uk. Accessed in April, 2025.

Figure A.10: Earnings Call Transcript (ECT) as found in Investor Relations Webpage

PETROBRAS

Webcast - 4th Quarter Results 2023
March 8, 2023

Operator:

Good morning, and welcome to Petrobras' earnings call for the 4Q23. It's a pleasure to have you with us today.

This event will be in Portuguese, and we have simultaneous translation into English. You can find the links to both on our Investor Relations page.

All participants will be in listen-only mode during the Company's presentation, and then we will have a Q&A session, and you may send us your questions to petroinvest@petrobras.com.br.

We have Jean Paul Prates, President of Petrobras, Clarice Coppetti, Corporate Affairs; Sergio Caetano Leite, Financial and Investor Relations Director; Mario Spinelli, Governance and Compliance Director; Tolmasquim, Energy Transition and Sustainability Director; Joelson Mendes, Exploration and Production Director; William França, Industrial Processes and Products; Claudio Schlosser, Logistics, Trade and Markets; Carlos Travassos, Engineering, Technology and Innovation; and Maiza Goulart. We will start with award from our CEO. Go ahead, sir.

Jean Paul Prates:

Thank you. Good morning everyone, dear shareholders, investors. I am very happy to talk to you today, not only because of the excellent economic and financial results we are going to present today, but especially because I am convinced we are building a more robust Petrobras that is more resilient and is better able to generate value on the long term to its shareholders and for society facing the numerous challenges that we know and do not know created by this world in transition.

Notes: This figure shows the initial remarks from the Earnings Call Transcript (ECT) of Petrobras' 4th Quarter 2023 results, held on March 8, 2024. It includes the opening statement by the operator, the full list of executive board members present at the event, and a brief introductory comment by then-CEO Jean Paul Prates. The webcast was conducted in Portuguese with real-time English translation and is available at Petrobras' Investor Relations webpage (www.investidorpetrobras.com.br. Accessed in February, 2025.). Note that the original document contains an error regarding the event date, incorrectly indicating that the call took place in 2023. This mistake refers to the date of the webcast itself, not the quarter being reported. The corrected version, reflecting the accurate date of March 8, 2024, is available through *LSEG Workspace*. This highlights one of the advantages of using LSEG's platform, which offers verified and adjusted transcripts for greater accuracy in financial analysis.

Figure A.11: Earnings Call Transcript (ECT) as found in *LSEG Workspace*

```
Refinitiv StreetEvents Event Transcript
E D I T E D   V E R S I O N

Q4 2023 Petroleo Brasileiro SA Petrobras Earnings Call
MARCH 08, 2024 / 2:30PM GMT

==============================================================================
Corporate Participants
==============================================================================

 * Carlos Jose do Nascimento Travassos
   Petróleo Brasileiro S.A. - Petrobras - Chief Engineering, Technology & Innovation Officer and Member of Executive Board
 * Mario Vinicius Claussen Spinelli
   Petróleo Brasileiro S.A. - Petrobras - Chief Governance & Compliance Officer and Member of Executive Board
 * Cláudio Romeo Schlosser
    -
 * William Franca da Silva
   Petróleo Brasileiro S.A. - Petrobras - Chief Industrial Processes & Products Officer and Member of Executive Board
 * Jean Paul Terra Prates
   Petróleo Brasileiro S.A. - Petrobras - CEO & Director
 * Sergio Caetano Leite
   Petróleo Brasileiro S.A. - Petrobras - Chief Financial & IR Officer and Member of Executive Board
 * Joelson Falcao Mendes
   Petróleo Brasileiro S.A. - Petrobras - Chief Exploration & Production Officer and Member of Executive Board
 * Mauricio Tiomno Tolmasquim
   Petróleo Brasileiro S.A. - Petrobras - Chief Energy Transition & Sustainability Officer and Member of Executive Board
 * Unidentified Company Representative
    -

==============================================================================
Conference Call Participiants
==============================================================================

 * Bruno Montanari
   Morgan Stanley, Research Division - Equity Analyst
 * Gabriel Coelho Barra
   Citigroup Inc., Research Division - Research Analyst
 * Pedro Soares
   Banco BTG Pactual S.A., Research Division - Analyst
 * Rodrigo Reis de Almeida
   Santander Investment Securities Inc., Research Division - Research Analyst
 * Vicente Falanga Neto
   Banco Bradesco BBI S.A., Research Division - Research Analyst
 * Rodolfo R. De Angele
   JPMorgan Chase & Co, Research Division - Head of Brazil Equity Research & Senior Analyst
 * Luiz Carvalho
   UBS Investment Bank, Research Division - Director and Analyst
 * Monique Martins Greco Natal
   Itaú Corretora de Valores S.A., Research Division - Research Analyst
 * Caio Burger Ribeiro
   BofA Securities, Research Division - Director in Equity Research and Head of the LatAm Metals and Mining and Pulp and Paper


==============================================================================
Presentation
------------------------------------------------------------------------------
Operator    [1]
------------------------------------------------------------------------------
Good morning, and welcome to Petrobras' earnings call for the fourth quarter of 2023. It's a pleasure to have you with us today. (Operator
Instructions)
We have Jean Paul Prates, President of Petrobras, Clarice Coppetti, Corporate Affairs; Sergio Caetano Leite, Financial and Investor
Relations Director, Mario Spinelli, Governance and Compliance Director, Tolmasquim, Energy Transition and Sustainability Director, Joelson
Mendes, Exploration and Production Director, William Franca, Industrial Processes and Products, Claudio Schlosser, Logistics, Trade and
Markets, Carlos Travassos, Engineering, Technology and Innovation, and Maiza Goulart, [Seps]. We'll start with award from our CEO. Go
ahead, sir.

------------------------------------------------------------------------------
Jean Paul Terra Prates,  Petróleo Brasileiro S.A. - Petrobras - CEO & Director    [2]
------------------------------------------------------------------------------
Thank you. Good morning everyone. Dear shareholders, investors, I'm very happy to talk to you today, not only because of the excellent
economic and financial results we're going to present today, but especially because I am convinced we're building a more robust Petrobras
that is more resilient and is better able to generate value on the long term to its shareholders and for society facing the numerous
challenges that we know and don't know created by this world in transition.
```

Notes: This figure shows the initial remarks from the Earnings Call Transcript (ECT) of Petrobras' 4th Quarter 2023 results, as provided by LSEG Workspace and downloaded in .txt format. It includes the corrected event date (March 8, 2024), the full list of corporate and market participants, and the beginning of the presentation with introductory comments from the event operator and then-CEO Jean Paul Prates. Unlike the original version available on Petrobras' Investor Relations website, which contained a date error, the transcript shown here reflects the accurate information, an example of the added value of using *LSEG Workspace* for accessing verified and structured financial transcripts.

In our empirical strategy, we maintain transcripts exclusively from the acquiring or dominant firm in the event of a merger or acquisition. For instance, we excluded the transcripts of América Latina Logística (ALL) after its 2015 incorporation into Rumo.[22] We also omitted those of SulAmérica following its 2022 acquisition by Rede D'Or.[23] Similarly, we dropped the transcripts of Votorantim Celulose e Papel (VCP), which in 2009 merged with Aracruz Celulose to form Fibria.[24] Those of Fibria following its 2019 merger with Suzano are likewise removed.[25] We retain only B3's transcripts following the 2017 merger of BM&F Bovespa and Cetip.[26] Our analysis discards Vivo Participações transcripts, retaining only Telefônica Brasil transcripts from 2011 following Telesp's incorporation of Vivo Participações and adoption of its current corporate name.[27] We consider only BRF's transcripts after the 2009 merger between Perdigão and Sadia.[28] Our sample goes until 2024 and therefore includes BRF prior to the announcement in 2025 of its planned merger with Marfrig to form MBRF.[29] It also includes JBS prior to its delisting from the B3 in June 2025, as part of a reorganization to create JBS NV, which will trade on the NYSE and have Brazilian Depositary Receipts (BDRs) on the B3.[30]

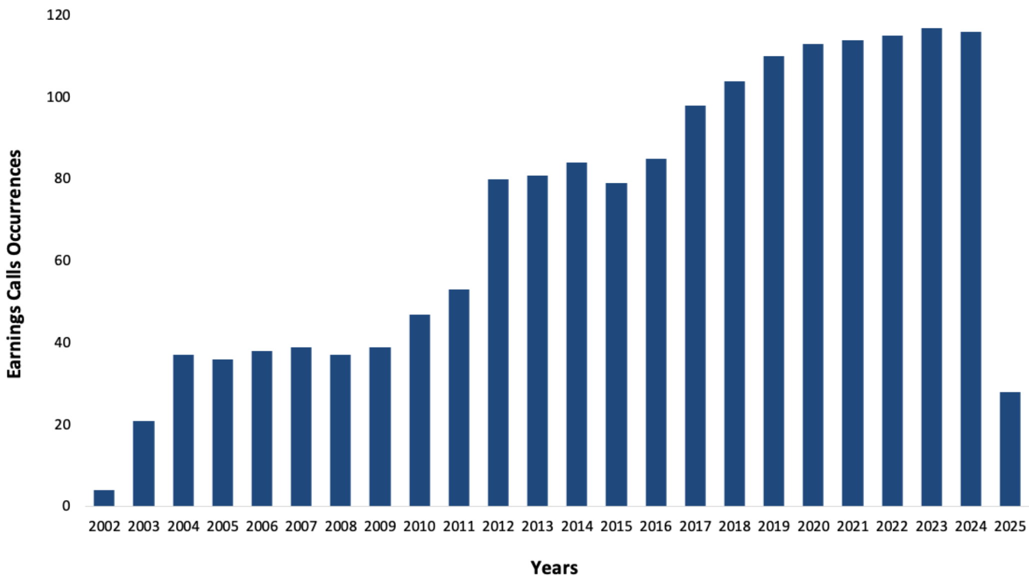
In contrast, if the company merely undergoes a rebranding or name change, we retain transcripts from both pre and post-change periods. This includes cases such as Companhia Vale do Rio Doce (CVRD), which in 2007 became Vale.[31] Another example is MPX, which became Eneva in 2013.[32] HRT Participações rebranded to PetroRio in 2015[33] and subsequently to PRIO in 2022.[34] Petrobras Distribuidora was renamed Vibra Energia following privatization in 2021.[35] Raia Drogasil adopted the name RD Saúde in 2024.[36] In cases of duplicate events, typically when earnings calls are recorded in two different events, one in English and the other in Portuguese with simultaneous translation, we retain only the first occurrence. This issue was observed, for example, in multiple Bradesco events between 2018 and 2023.

We also identified a small subset of transcripts in which the Q&A section was missing, either because it was not recorded in the database or because no questions were asked during the event. In total, 34 such cases were recorded. The issue was more frequent in some firms, with Sabesp (10 transcripts) and Ultrapar (6 transcripts) accounting for the largest counts, followed by Eneva (5) and Equatorial (4). In relative terms, Sabesp and Itaúsa exhibited the highest proportions of missing Q&A sessions, at 23% and 15% of their respective transcripts. When estimating regressions using the split-document setup, the sentiment for these missing Q&A sections was coded as not available (*NA*), indicating the absence of data for that component.

[22] More information available at oglobo.globo.com. Accessed April 2025.
[23] More information available at cnnbrasil.com.br. Accessed April 2025.
[24] More information available at economia.terra.com.br. Accessed August 2025.
[25] More information available at exame.com. Accessed April 2025.
[26] More information available at agenciabrasil.ebc.com.br. Accessed April 2025.
[27] More information available at infomoney.com.br. Accessed August 2025.
[28] More information available at ri.brf-global.com. Accessed April 2025.
[29] More information available at g1.globo.com. Accessed August 2025.
[30] More information available at research.ftserussell.com. Accessed August 2025.
[31] More information available at oglobo.globo.com. Accessed in August, 2025
[32] More information available at g1.globo.com. Accessed in April, 2025
[33] More information available at braziljournal.com. Accessed in August, 2025
[34] More information available at folhavitoria.com.br. Accessed in August, 2025
[35] More information available at g1.globo.com. Accessed in April, 2025
[36] More information available at braziljournal.com. Accessed in August, 2025

In terms of the temporal distribution of the calls in our sample, we observe a clear upward trend over the years in the total volume of ECTs. As shown in Figure A.12, the early 2000s were characterized by limited adoption, with earnings calls primarily conducted by a small group of prominent firms, particularly former state-owned enterprises such as Petrobras and Vale. These companies were among the first in Brazil to adopt earnings calls as a formal communication channel with the market. A marked expansion in the use of this practice began around 2012, reflecting the broader diffusion of investor relations standards across the corporate sector, driven by improved governance practices, growing transparency demands, and increased engagement from market participants.

Figure A.12: ECT Occurrences - Years



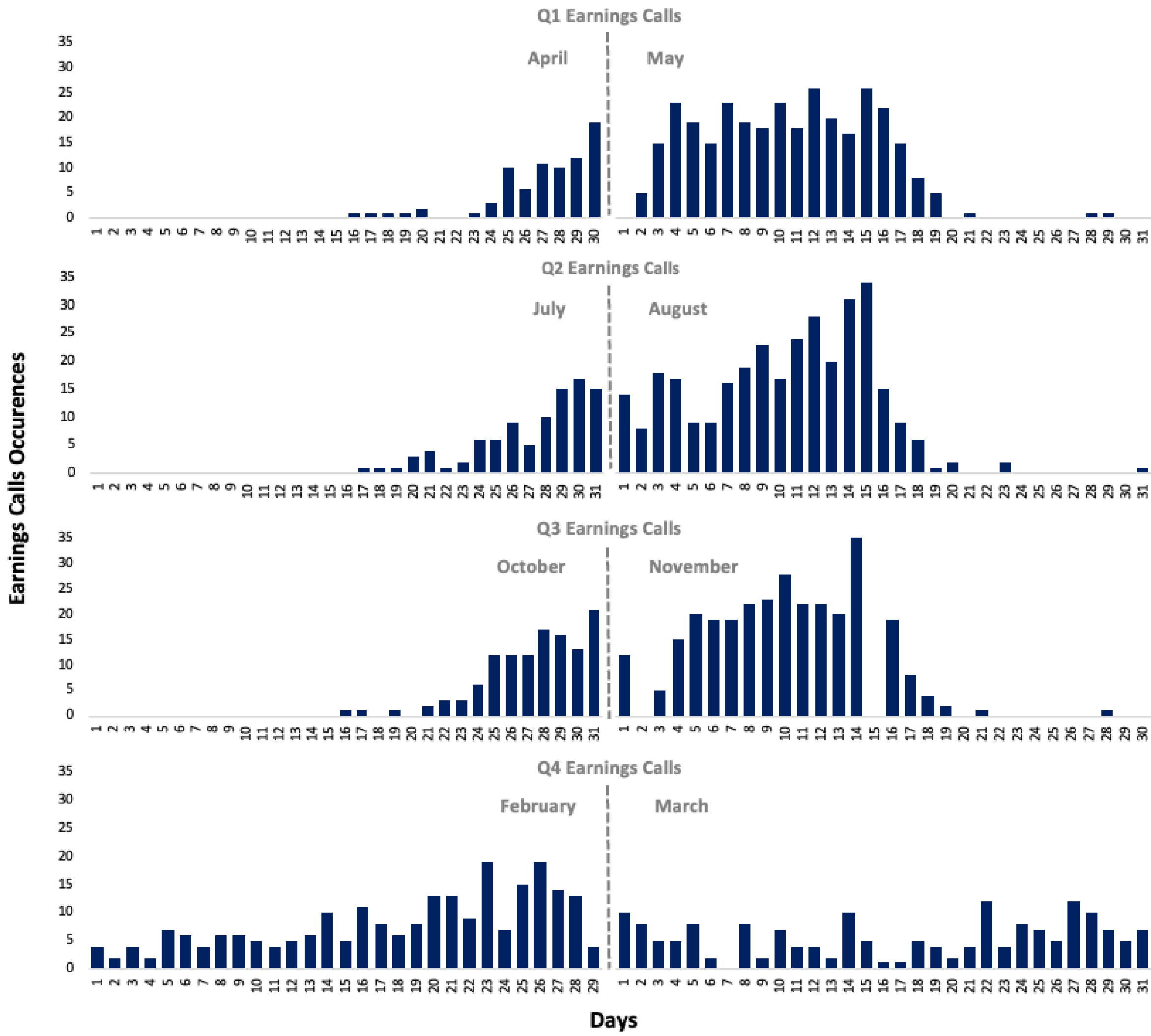
Notes: This figure shows the distribution of earnings call transcripts (ECTs) by the calendar year in which the call took place, rather than by the fiscal quarter being discussed. It is important to note that calls covering fourth-quarter and annual results are typically held in the early months of the following year, which creates a systematic shift between the call year and the year to which the reported results refer (see Figure A.13). The apparent decline in 2025 simply reflects the fact that our sample ends in 2024 Q4, whose corresponding calls occur in early 2025.

Turning to the timing of calls within each fiscal year, those discussing first-quarter results, typically held during April and May, exhibit a volume surge in the second half of April, which extends into the first half of May. This operational model, where calls about a given quarter's earnings peak around the turn of the month in the subsequent period, is replicated with remarkable fidelity for the results of the second quarter (discussed in July–August) and the third quarter (discussed in October– November), as clearly illustrated in Figure A.13. The cycle related to fourth-quarter results, occurring in February and March, exhibits a more diffuse pattern. The increased complexity of the annual report,[37] potentially combined with disruptions from the Carnival holiday, likely accounts for the more spread-out call distribution

[37] Under the regulatory framework of Brazil's securities commission (CVM), publicly traded companies include fourth-quarter results in their annual financial statements (DFP) rather than filing a separate quarterly report (ITR). The DFP deadline is significantly longer (up to three months after year-end compared to 45 days for the other quarters), due to the requirement for a full audit instead of the limited review applied to ITRs.

compared to other quarters.

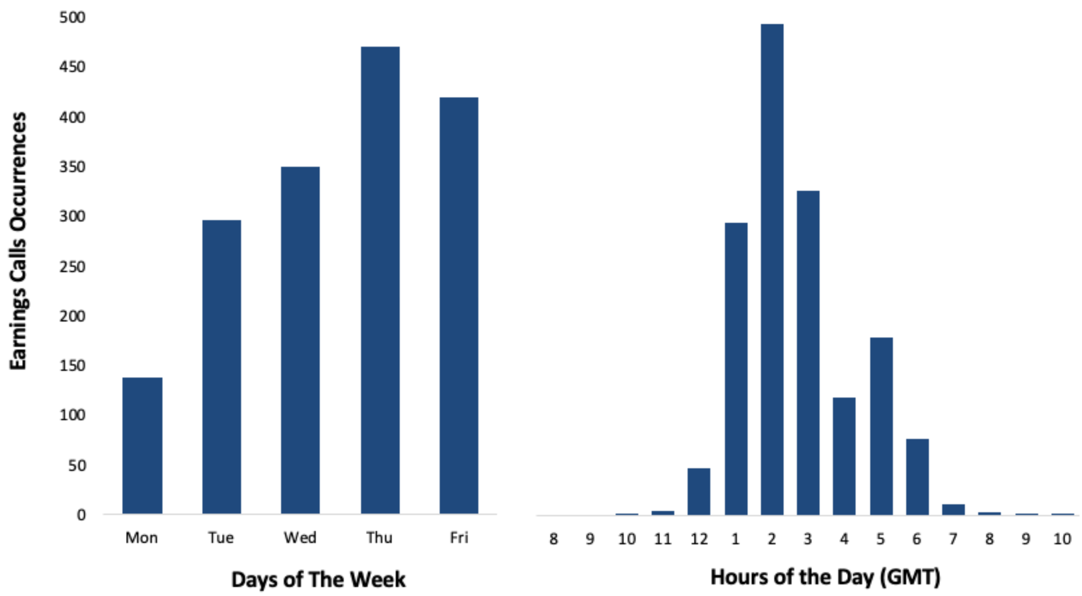
Figure A.13: ECT Occurrences - Months and Days of the Month



Notes: This figure displays the distribution of earnings call transcript (ECT) occurrences across calendar days in our sample, grouped by the month in which each call took place and organized by fiscal quarter. It highlights a consistent pattern across Q1 to Q3, where calls tend to cluster around the transition between months, specifically, from the second half of the first month into the first half of the following month (e.g., late April to mid-May for Q1 results). In contrast, calls for Q4 results (discussed in February–March) are more evenly dispersed. This dispersion is likely due to the extended deadline and higher scrutiny of annual financial statements (DFP), as well as potential scheduling disruptions around Carnival. Additionally, noticeable drops in call frequency on specific days, such as May 1st (Labor Day), November 2nd (Day of the Dead), and November 15th (Proclamation of the Republic), can be attributed to local holidays.

Weekly call volume exhibits a progressive intensification, starting low on Monday and building steadily to a peak on Thursday, while remaining substantial on Friday. When analyzing the intraday patterns in Brasília Time (BRT)[38], the calls can be conceptually grouped into a morning and an afternoon block. The analysis reveals a stark imbalance, with the vast majority of calls being disproportionately concentrated in the morning block. In contrast, the afternoon block shows significantly diminished volume with only minor, secondary peaks. This distribution suggests that the majority of critical calls for the day are concluded before the midday lunch period.These weekly and intraday patterns are visually summarized in Figure A.14.

Figure A.14: ECT Occurrences - Day of the Week and Hour of the Day



Notes: This figure shows the distribution of earnings calls in our sample, breaking them down by day of the week and hour of the day. The left panel demonstrates a clear weekly pattern where call volume increases from Monday to a peak on Thursday. The right panel shows the hourly frequency, with the crucial detail that times are listed in Greenwich Mean Time (GMT). To interpret this for the local time, one must convert to Brasília Time (BRT) by subtracting three hours (GMT-3) or to Brasília Summer Time (BRST) by subtracting two hours (GMT-2). This conversion reveals that call activity is highly concentrated in the late morning in Brazil, indicating the most intensive period for calls occurs just before lunch.

A.2 ECT Preprocessing

The primary objective of the preprocessing pipeline is to prepare and standardize the textual data from the earnings call transcripts while preserving the specific word forms found in the Loughran and McDonald (2011) financial lexicon. The process begins with contraction expansion, where terms such as "haven't" are converted to their full form, "have not," to ensure uniform tokenization. A comprehensive list of the contractions addressed is detailed in Table A.7. Following this, proper nouns are identified using part-of-speech (PoS) tagging with

[38]Brazil officially abolished Daylight Saving Time (DST) on February, 2019, resulting in Brasília Time (BRT) maintaining a consistent UTC-3 offset year-round. Prior to this date, during Brazil's summer months (typically from late-October/early-November to late-February), Daylight Saving Time was observed as Brasília Summer Time (BRST), with a UTC-2 offset. Outside these periods, the standard BRT (UTC-3) was in effect. For the purpose of this analysis, historical DST adjustments have not been factored into the time conversions.

the spaCy library and removed from the text. This step is intentionally performed before lowercasing to leverage capitalization for more accurate entity recognition.

After the initial steps, the text undergoes further normalization. All characters are converted to lowercase, and non-lexical elements, including punctuation and numbers, are stripped out to ensure consistent matching between the transcript tokens and the dictionary entries. The resulting text is then tokenized using the nltk library. A customized stopword list, detailed in Table A.8, is applied to filter out high-frequency words. This list is strategically adapted from the standard nltk collection to explicitly retain sentiment-bearing financial terms such as "up," "down," "not," and "no". As a final preparatory step, any resulting tokens with fewer than two characters are discarded. Critically, in direct alignment with the methodology of Hasani and Zaw (2024), stemming and lemmatization are not applied. This omission is essential because the Loughran and McDonald (2011) dictionary includes specific word inflections; reducing words to their root forms would create systematic mismatches and lead to an underestimation of sentiment-bearing terms.

Table A.7: List of Common English Language Contractions

Common Contractions			
ain't	aren't	can't	could've
couldn't	couldn't've	didn't	doesn't
don't	hadn't	hadn't've	hasn't
haven't	he'd	he'd've	he'll
he's	how'd	how'll	how's
I'd	I'd've	I'll	I'm
I've	isn't	it'd	it'd've
it'll	it's	let's	ma'am
mightn't	mightn't've	might've	mustn't
must've	needn't	not've	o'clock
oughtn't	'ow's'at	shan't	she'd
she'd've	she'll	she's	should've
shouldn't	shouldn't've	somebody'd	somebody'd've
somebody'll	somebody's	someone'd	someone'd've
someone'll	someone's	something'd	something'd've
something'll	something's	that'll	that's
there'd	there'd've	there're	there's
they'd	they'd've	they'll	they're
they've	'twas	wasn't	we'd
we'd've	we'll	we're	we've
weren't	what'll	what're	what's
what've	when's	where'd	where's
where've	who'd	who'd've	who'll
who're	who's	who've	why'll
why're	why's	won't	would've
wouldn't	wouldn't've	y'all	y'all'll
y'all'd've	you'd	you'd've	you'll
you're	you've		

Notes: This table presents a comprehensive list of common English contractions used in text preprocessing. The contractions are sourced from Hasani and Zaw (2024), who compiled this list from the GitHub repository available at github.com/J3RN. These contractions are were expanded to their full forms during text normalization procedures to ensure consistent analysis across different sentiment measurement approaches.

Table A.8: List of Stopwords

Stopwords							
a	about	after	again	against	ain	all	am
an	and	any	are	aren	aren't	as	at
be	because	been	before	being	between	both	but
by	can	couldn	couldn't	d	did	didn	didn't
do	does	doesn	doesn't	doing	don	don't	during
each	few	for	from	further	had	hadn	hadn't
has	hasn	hasn't	have	haven	haven't	having	he
he'd	he'll	he's	her	here	hers	herself	him
himself	his	how	i	i'd	i'll	i'm	i've
if	in	into	is	isn	isn't	it	it'd
it'll	it's	its	itself	just	ll	m	ma
me	mightn	mightn't	more	most	mustn	mustn't	my
myself	needn	needn't	nor	now	o	of	off
on	once	only	or	other	our	ours	ourselves
out	own	re	s	same	shan	shan't	she
she'd	she'll	she's	should	should've	shouldn	shouldn't	so
some	such	t	than	that	that'll	the	their
theirs	them	themselves	then	there	these	they	they'd
they'll	they're	they've	this	those	through	to	too
until	ve	very	was	wasn	wasn't	we	we'd
we'll	we're	we've	were	weren	weren't	what	when
where	which	while	who	whom	why	will	with
won	won't	wouldn	wouldn't	y	you	you'd	you'll
you're	you've	your	yours	yourself	yourselves		

Notes: This table presents the customized stopword list used in this study, adapted from Hasani and Zaw (2024). The stopword list is based on the standard `nltk` English stopwords collection, with strategic modifications for financial text analysis. Specifically, sentiment-relevant words such as "up," "down," "above," "below," "under," "over," "no," and "not" are explicitly retained to preserve crucial information for financial sentiment analysis. Additionally, the word "can" is added to the stopword list. This customization enhances the sensitivity of sentiment analysis in financial contexts while maintaining linguistic coherence.

A.3 Ternary Classification

Table A.9 highlights systematic differences in sentiment distribution across methods. The LM dictionary produces a relatively balanced classification, although the proportions depend on the chosen threshold. In comparison, transformer-based models such as FinBERT and GPT tend to assign most documents to the neutral or positive categories, with only a small share classified as negative. This pattern is consistent across our study and Hasani and Zaw (2024), despite differences in corpora: whereas they analyze earnings calls from Scandinavian firms, we focus on Brazilian firms. This recurring pattern across corpora suggests that transformer-based sentiment models share a tendency to classify very few observations as negative, regardless of the institutional or cultural setting. This tendency raises important methodological considerations, as the underrepresentation of negative classifications may attenuate the explanatory power. [39]

Table A.9: Comparison of Sentiment Distribution Classifications

Hasani and Zaw (2024)

Method	Threshold	Positive	Neutral	Negative
LM Dictionary	0.20	26%	47%	27%
FinBERT	–	26%	72%	3%
GPT-4 Turbo	–	76%	19%	5%

Ours

Method	Threshold	Positive	Neutral	Negative
LM Dictionary	0.05	46%	14%	40%
	0.10	39%	27%	34%
	0.20	27%	50%	23%
	0.33	16%	72%	12%
FinBERT	–	21%	78%	1%
GPT-4o Mini	–	83%	16%	2%

Notes: This table compares the distribution of sentiment classifications between our study and Hasani and Zaw (2024). For the LM Dictionary approach, different threshold values are used to classify sentiment: documents are classified as positive if the positive sentiment score exceeds the negative sentiment score by more than the threshold, negative if the negative score exceeds the positive score by more than the threshold, and neutral otherwise. FinBERT and GPT classifications are based on the models' direct sentiment predictions. The similarity in distributions across studies validates the robustness of our sentiment measurement approach and ensures comparability with existing literature.

Building on this distributional evidence, Table A.10 examines whether our regression results are sensitive to alternative threshold choices for the LM dictionary. The findings show that the qualitative interpretation remains stable across specifications: positive sentiment is always associated with higher CARs, while negative sentiment is linked to lower CARs. However, the magnitude of these effects varies with the cutoff. Positive coefficients become larger as the threshold increases, whereas the negative coefficient is consistently small and only reaches statistical significance at the 0.20 threshold. This indicates that while threshold selection influences the estimated strength of LM-based sentiment, it does not alter the overall direction or

[39] Because negative classifications are relatively rare, coefficients on the Neg_GPT variable may be more sensitive to minor output variability, which Atil et al. (2024) show may occur in LLMs even at temperature zero.

significance of the main results.

Table A.10: Sensitivity of Regression Results to LM Dictionary Threshold for Ternary Sentiment Classification

Dependent Variable: Immediate Cumulative Abnormal Returns (CAR)				
	T = 0.05	**T = 0.10**	**T = 0.20**	**T = 0.33**
Intercept	0.0439	0.0433	0.0432	0.0454*
	(0.0282)	(0.0277)	(0.0283)	(0.0256)
Pos_LM	0.0061**	0.0071***	0.0090***	0.0142***
	(0.0031)	(0.0021)	(0.0021)	(0.0046)
Neg_LM	-0.0052	-0.0050	-0.0096*	-0.0073
	(0.0045)	(0.0047)	(0.0051)	(0.0095)
Controls	Yes	Yes	Yes	Yes
Industry Dummies	Yes	Yes	Yes	Yes
Year Dummies	Yes	Yes	Yes	Yes
Quarter Dummies	Yes	Yes	Yes	Yes
N	1385	1385	1385	1385
R^2	0.0380	0.0376	0.0410	0.0398
Adjusted R^2	0.0086	0.0082	0.0117	0.0104
Sample Period	2002-2024	2002-2024	2002-2024	2002-2024

Notes: This table presents regression results examining the sensitivity of our findings to different threshold values (T) for the LM Dictionary ternary sentiment classification. Pos_LM and Neg_LM represent positive and negative sentiment indicators, respectively, derived from the Loughran and McDonald dictionary. Documents are classified as positive if the positive sentiment score exceeds the negative sentiment score by more than the threshold T, negative if the negative score exceeds the positive score by more than T, and neutral otherwise. Standard errors are reported in parentheses. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively. All regressions include firm-level controls, industry fixed effects, year fixed effects, and quarter fixed effects.

B More on Data Sources and Variables

B.1 LSEG Workspace

Most of the financial variables employed in this study are obtained from the *London Stock Exchange Group* (LSEG) *Workspace*, previously known as *Refinitiv Eikon*. LSEG is a global provider of financial market data, offering comprehensive information on equities, fixed income, derivatives, and economic indicators. Its datasets are widely used in academic and professional research due to their breadth, standardization, and reliability. The data and financials for each firm were retrieved during the first two weeks of August, 2025.

B.1.1 Market Capitalization

Market Capitalization (LSEG code: `TR.F.MKTCAP`) represents the total market value of a company's equity and is calculated by multiplying the closing price of the stock by the number of shares outstanding:

$$\text{Market Capitalization} = \text{Price per Share} \times \text{Shares Outstanding}. \tag{10}$$

B.1.2 Book-to-Market (B/M) Ratio

The Book-to-Market ratio is constructed as the inverse of the Price-to-Book ratio. The Price-to-Book ratio (LSEG code: `TR.F.PRICETOBOOKVALUEPERSHRISSUE`)[40] measures the relationship between the market price per share and the book value per share:

$$\text{Price-to-Book} = \frac{\text{Market Price per Share}}{\text{Book Value per Share}}, \tag{11}$$

$$\text{Book-to-Market} = \frac{1}{\text{Price-to-Book}}. \tag{12}$$

B.1.3 Return on Average Total Assets (ROA)

Return on Average Total Assets (LSEG code: `TR.F.RETURNAVGTOTASSETSPCTTTM`) measures a company's ability to generate income relative to its asset base. ROA is calculated as the ratio of income before discontinued operations and extraordinary items to the average total assets, expressed as a percentage:

$$\text{ROA} = \left(\frac{\text{Income Before Discontinued Operations and Extraordinary Items}}{\text{Average Total Assets}}\right) \times 100. \tag{13}$$

B.1.4 Stock Turnover

Share Turnover quantifies the proportion of a company's outstanding shares that were traded over a given fiscal quarter. It is constructed using trading volume data and the number of common shares outstanding, both obtained from LSEG. Specifically, we retrieve:

- **Volume** (LSEG code: `TR.Volume`): Total number of shares traded during the fiscal quarter;
- **Common Shares Outstanding** (LSEG code: `TR.F.COMSHROUTSTOT`): Number of shares outstanding at the end of the fiscal quarter.

The variable is computed as:

$$\text{Share Turnover (\%)} = \left(\frac{\text{Quarterly Trading Volume}}{\text{Common Shares Outstanding at Quarter-End}}\right) \times 100. \tag{14}$$

[40] For Rumo (RAIL3), the Book-to-Market data were unavailable on the LSEG Workspace platform. The missing values were retrieved from Bloomberg using the `PX_TO_BOOK_RATIO` ticker.

B.1.5 Analyst Coverage

We construct a variable that captures the number of analysts participating in each earnings call by extracting this information directly from the document header section of the transcripts. As illustrated in the example provided in Figure B.15, the header includes a list of all analysts identified by name and affiliation.

Figure B.15: List of Earnings Call (Non-Corporate) Participants as found in *LSEG Workspace* Transcripts

```
================================================================================
Conference Call Participants
================================================================================

 * Mario Pierry
   BofA Merrill Lynch - Analyst
 * Philip Finch
   UBS - Analyst
 * Tito Labarta
   Deutsche Bank - Analyst
 * Marcelo Telles
   Credit Suisse - Analyst
 * Boris Molina
   Santander - Analyst
 * Saul Martinez
   JPMorgan - Analyst
 * Thiago Batista
   Itau BBA - Analyst
 * Patricia Medina
   ING - Analyst
 * Carlos Macedo
   Goldman Sachs - Analyst
 * Jorge Kuri
   Morgan Stanley - Analyst
```

Notes: This figure shows an excerpt from the conference call participants list typically found in the header section of earnings call transcripts from *LSEG Workspace*. The list includes the names and affiliated institutions of all external participants, such as sell-side analysts. In this example, the analyst coverage variable would be assigned a value of 10, corresponding to the total number of distinct non-corporate participants. This count is used as a proxy for the level of external market attention and informational demand associated with the earnings call.

This count serves as a proxy for the degree of analyst coverage associated with both the firm and the specific earnings call. By using the number of participating analysts as a continuous variable, we aim to capture variations in external market attention, informational demand, and potential scrutiny that the firm faces from financial intermediaries.[41]

[41] In some cases, the analyst coverage variable displayed consecutive zero values (for instance, for Petrobras (PETR4) from 2020 Q1 to 2023 Q3 and for Eneva (ENEV3) from 2021 Q2 to 2023 Q2). This pattern often arose during the COVID-19 pandemic, when questions were read by an Investor Relations (IR) representative rather than posed directly by analysts. In addition, some isolated zero values resulted from a recurring typographical error ("Conference Call Participiants") in the transcripts, which hindered correct identification via regular expressions and prevented the automated count of participants. As it was initially unclear whether these zeros reflected an actual absence of participation or a recording omission, we revisited the original transcripts and, within the Q&A section, manually counted the number of distinct individuals asking questions. This procedure enabled the reconstruction of the series for most missing entries. When no questions and no external participants were present, the zero value was retained. Given our use of logarithmic transformations in the empirical analysis, we applied the $\ln(1 + \text{Coverage})$ transformation to accommodate true zero values.

B.1.6 Daily Stock Return

Stock Return is computed as the percentage change in the daily closing price (LSEG code: `TR.PriceClose`) of each stock. The closing price represents the last reported transaction price during the trading day. Stock returns are calculated according to the following formula:

$$\text{Stock Return}_t = \frac{\text{Price Close}_t - \text{Price Close}_{t-1}}{\text{Price Close}_{t-1}}, \tag{15}$$

where Price Close_t denotes the closing price of the stock at day t, and Price Close_{t-1} denotes the closing price at day $t-1$.

The use of daily stock returns is essential for calculating cumulative abnormal returns (CAR), as it allows for the construction of the time series needed to estimate expected returns around the event window.

B.2 NEFIN

The Brazilian Center for Research in Financial Economics (NEFIN) is a research group associated with the University of São Paulo (USP) that provides widely used financial datasets for academic research. Among its contributions, NEFIN offers daily series of risk factors[42] constructed according to rigorous eligibility criteria for the Brazilian stock market. The NEFIN risk factor series used in this study were downloaded in the beginning of August, 2025.

B.2.1 Market Factor

The market factor (R_m) is defined as the difference between the value-weighted daily return of the market portfolio and the daily risk-free rate. The market portfolio is constructed using all eligible stocks traded on Bovespa.

A stock is considered eligible for year t if it satisfies the following conditions:

(i) It must be the most traded stock of the firm, identified by the highest trading volume during the previous year.

(ii) It must have been traded on more than 80% of the trading days in year $t-1$, with an average daily trading volume exceeding R$ 500,000.00. For stocks listed during year $t-1$, the evaluation period covers from the listing date to the end of the year.

(iii) It must have been initially listed before December of year $t-1$.

The daily return on the market portfolio is calculated based on these eligible stocks and is value-weighted by market capitalization.

[42] Available at nefin.com.br. Accessed in March, 2025.

B.2.2 Risk-Free Rate

The daily risk-free rate (R_f) used in the construction of the market factor is derived from the 30-day DI Swap rate. This rate is commonly used in the Brazilian financial system as a proxy for the risk-free asset.

B.2.3 Small Minus Big (SMB)

The Small Minus Big factor (SMB) captures the size premium by measuring the return differential between portfolios formed on firm size. Each January, firms are sorted into three size-based quantiles according to their market capitalization at the end of December of year $t-1$. The SMB factor is computed as the equal-weighted average return of the small-size portfolio minus the return of the big-size portfolio.

B.2.4 High Minus Low (HML)

The High Minus Low factor (HML) captures the value premium by measuring the return differential between firms based on their book-to-market ratios. Firms are sorted annually, every January, into three quantiles according to the book-to-market ratio measured in June of year $t-1$. The HML factor is defined as the equal-weighted return of the high book-to-market portfolio minus the return of the low book-to-market portfolio.

B.2.5 Winners Minus Losers (WML)

The Winners Minus Losers factor (WML) captures the momentum effect by considering past stock returns. Firms are ranked annually by their cumulative returns between months $t-12$ and $t-2$, and sorted into three quantiles. The WML factor is computed as the equal-weighted return of the portfolio of past winners minus the return of the portfolio of past losers.

B.3 B3

Brasil, Bolsa, Balcão (B3) is the main financial market infrastructure provider in Brazil, operating the country's official stock exchange. It offers services related to the trading of equities, fixed income securities, derivatives, and other financial assets. B3 also provides standardized classifications for listed companies, including detailed sectoral categorizations based on the nature of their products and services. These classifications facilitate academic research and market analysis by enabling consistent identification and grouping of firms according to economic activity. The sector classification data used in this study were downloaded in early February, 2025.

B.3.1 Sector Dummies

The industry classification used in this study is obtained from B3 website [43], the main stock exchange in Brazil. B3 provides a structured classification system based on the type of products and services developed by companies, aiming to:

- Provide a clear identification of companies' sectors from the first level of the structure;
- Group companies that, despite engaging in different activities, belong to the same production chain or produce related goods and services, thus exhibiting similar economic behavior;
- Facilitate the identification of companies' primary areas of activity;
- Align the classification with international standards adopted by major financial institutions globally.

Following B3's sector classification, each firm in the sample is assigned to a corresponding economic sector. Based on this assignment, we construct sector dummies that are used as control variables in the empirical analysis. Each dummy variable equals one if the firm belongs to the specified economic sector and zero otherwise.

C More on Empirical Evidence

C.1 Control Sensibility Test

To assess the robustness of our main findings and ensure that the documented sentiment effects are not artifacts of model specification choices, we conduct a series of control sensitivity tests. These tests systematically examine how the inclusion or exclusion of specific control variables affects our core sentiment coefficient estimates. The following tables present these sensitivity analyses for each of our three sentiment measurement approaches (LM Dictionary, FinBERT, and GPT), applied both to the unified and divided earnings call transcripts.

[43] Available at www.b3.com.br. Accessed in February, 2025.

Table C.11: Control Sensibility Test - LM Sentiment using Unified Documents (Full Sample: 2002 - 2024)

Dependent Variable: Immediate Cumulative Abnormal Returns (CAR)						
	(1)	(2)	(3)	(4)	(5)	(6)
Intercept	0.0433	-0.0225**	0.0610**	0.0420	0.0419	0.0377
	(0.0277)	(0.0106)	(0.0271)	(0.0278)	(0.0262)	(0.0247)
Pos_LM	0.0071***	0.0066***	0.0073***	0.0066***	0.0072***	0.0074***
	(0.0021)	(0.0023)	(0.0020)	(0.0020)	(0.0021)	(0.0021)
Neg_LM	-0.0050	-0.0070	-0.0047	-0.0051	-0.0050	-0.0049
	(0.0047)	(0.0044)	(0.0046)	(0.0045)	(0.0047)	(0.0047)
ln(Market Cap)	-0.0068***		-0.0068***	-0.0065***	-0.0069***	-0.0062***
	(0.0026)		(0.0025)	(0.0024)	(0.0027)	(0.0023)
ln(B/M)	0.0009	0.0030		0.0006	0.0008	0.0006
	(0.0023)	(0.0024)		(0.0022)	(0.0021)	(0.0021)
ROA	0.0001	0.0000	0.0001		0.0001	0.0001
	(0.0001)	(0.0001)	(0.0001)		(0.0001)	(0.0001)
ln(1 + Coverage)	-0.0013	-0.0051	-0.0008	-0.0015		-0.0010
	(0.0036)	(0.0046)	(0.0032)	(0.0035)		(0.0035)
ln(Turnover)	-0.0031	-0.0013	-0.0021	-0.0026	-0.0030	
	(0.0024)	(0.0017)	(0.0017)	(0.0025)	(0.0024)	
Industry Dummies	Yes	Yes	Yes	Yes	Yes	Yes
Year Dummies	Yes	Yes	Yes	Yes	Yes	Yes
Quarter Dummies	Yes	Yes	Yes	Yes	Yes	Yes
N	1385	1400	1457	1412	1385	1385
R^2	0.0376	0.0275	0.0372	0.0355	0.0375	0.0364
Adjusted R^2	0.0082	-0.0011	0.0100	0.0074	0.0089	0.0077
Sample Period	2002-2024	2002-2024	2002-2024	2002-2024	2002-2024	2002-2024

Notes: This table presents results from control sensitivity tests examining the robustness of LM Dictionary sentiment measures applied to unified earnings call transcripts. Each transcript is analyzed as a single document rather than being segmented by section. The dependent variable is the cumulative abnormal return (CAR), measuring stock price reactions around the earnings call date. Column (1) presents the baseline specification with the complete set of firm-level control variables (market capitalization, book-to-market ratio, return on assets, analyst coverage, and stock turnover). Columns (2) through (6) systematically exclude one control variable at a time to evaluate the stability and robustness of the sentiment coefficient estimates across different model specifications. This approach helps ensure that the documented sentiment effects are not driven by correlations with specific control variables. Standard errors are reported in parentheses and are clustered at the firm and time level. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

Table C.12: Control Sensibility Test - FinBERT Sentiment using Unified Documents (Full Sample: 2002 - 2024)

Dependent Variable: Immediate Cumulative Abnormal Returns (CAR)						
	(1)	(2)	(3)	(4)	(5)	(6)
Intercept	0.0493*	-0.0230**	0.0626**	0.0474*	0.0463*	0.0432*
	(0.0280)	(0.0100)	(0.0278)	(0.0277)	(0.0265)	(0.0251)
Pos_FinBERT	0.0015	0.0020	0.0018	0.0020	0.0022	0.0017
	(0.0031)	(0.0033)	(0.0030)	(0.0031)	(0.0029)	(0.0031)
Neg_FinBERT	-0.0113	-0.0085	-0.0113	-0.0106	-0.0100	-0.0108
	(0.0120)	(0.0122)	(0.0118)	(0.0120)	(0.0122)	(0.0116)
ln(Market Cap)	-0.0073***		-0.0072***	-0.0070***	-0.0074***	-0.0067***
	(0.0025)		(0.0025)	(0.0024)	(0.0027)	(0.0023)
ln(B/M)	0.0004	0.0026		0.0001	0.0001	0.0001
	(0.0024)	(0.0025)		(0.0023)	(0.0022)	(0.0021)
ROA	0.0001	0.0000	0.0001		0.0001	0.0001
	(0.0001)	(0.0001)	(0.0001)		(0.0001)	(0.0001)
ln(1 + Coverage)	-0.0024	-0.0066	-0.0020	-0.0024		-0.0020
	(0.0036)	(0.0048)	(0.0033)	(0.0036)		(0.0036)
ln(Turnover)	-0.0033	-0.0014	-0.0023	-0.0029	-0.0032	
	(0.0024)	(0.0019)	(0.0017)	(0.0025)	(0.0025)	
Industry Dummies	Yes	Yes	Yes	Yes	Yes	Yes
Year Dummies	Yes	Yes	Yes	Yes	Yes	Yes
Quarter Dummies	Yes	Yes	Yes	Yes	Yes	Yes
N	1385	1400	1457	1412	1385	1385
R^2	0.0329	0.0210	0.0323	0.0311	0.0326	0.0315
Adjusted R^2	0.0034	-0.0078	0.0050	0.0028	0.0030	0.0026
Sample Period	2002-2024	2002-2024	2002-2024	2002-2024	2002-2024	2002-2024

Notes: This table presents results from control sensitivity tests examining the robustness of FinBERT sentiment measures applied to unified earnings call transcripts. Each transcript is analyzed as a single document rather than being segmented by section. The dependent variable is the cumulative abnormal return (CAR), measuring stock price reactions around the earnings call date. Column (1) presents the baseline specification with the complete set of firm-level control variables (market capitalization, book-to-market ratio, return on assets, analyst coverage, and stock turnover). Columns (2) through (6) systematically exclude one control variable at a time to evaluate the stability and robustness of the sentiment coefficient estimates across different model specifications. This approach helps ensure that the documented sentiment effects are not driven by correlations with specific control variables. Standard errors are reported in parentheses and are clustered at the firm and time level. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

Table C.13: Control Sensibility Test - GPT Sentiment using Unified Documents (Full Sample: 2002 - 2024)

Dependent Variable: Immediate Cumulative Abnormal Returns (CAR)						
	(1)	(2)	(3)	(4)	(5)	(6)
Intercept	0.0371	-0.0305***	0.0567**	0.0353	0.0345	0.0316
	(0.0296)	(0.0099)	(0.0284)	(0.0295)	(0.0289)	(0.0260)
Pos_GPT	0.0097*	0.0107**	0.0094*	0.0100*	0.0096*	0.0101*
	(0.0055)	(0.0050)	(0.0052)	(0.0052)	(0.0055)	(0.0053)
Neg_GPT	-0.0193***	-0.0216***	-0.0188***	-0.0196***	-0.0193***	-0.0190***
	(0.0053)	(0.0059)	(0.0060)	(0.0055)	(0.0053)	(0.0053)
ln(Market Cap)	-0.0068***		-0.0069***	-0.0065***	-0.0071***	-0.0063***
	(0.0026)		(0.0026)	(0.0024)	(0.0027)	(0.0023)
ln(B/M)	0.0013	0.0033		0.0010	0.0009	0.0009
	(0.0022)	(0.0023)		(0.0022)	(0.0021)	(0.0020)
ROA	0.0001	0.0000	0.0001		0.0001	0.0001
	(0.0001)	(0.0001)	(0.0001)		(0.0001)	(0.0001)
ln(1 + Coverage)	-0.0026	-0.0067	-0.0020	-0.0028		-0.0024
	(0.0035)	(0.0044)	(0.0032)	(0.0035)		(0.0035)
ln(Turnover)	-0.0030	-0.0012	-0.0020	-0.0026	-0.0029	
	(0.0026)	(0.0019)	(0.0018)	(0.0026)	(0.0026)	
Industry Dummies	Yes	Yes	Yes	Yes	Yes	Yes
Year Dummies	Yes	Yes	Yes	Yes	Yes	Yes
Quarter Dummies	Yes	Yes	Yes	Yes	Yes	Yes
N	1385	1400	1457	1412	1385	1385
R^2	0.0393	0.0292	0.0385	0.0378	0.0389	0.0381
Adjusted R^2	0.0100	0.0006	0.0113	0.0098	0.0103	0.0095
Sample Period	2002-2024	2002-2024	2002-2024	2002-2024	2002-2024	2002-2024

Notes: This table presents results from control sensitivity tests examining the robustness of GPT sentiment measures applied to unified earnings call transcripts. Each transcript is analyzed as a single document rather than being segmented by section. The dependent variable is the cumulative abnormal return (CAR), measuring stock price reactions around the earnings call date. Column (1) presents the baseline specification with the complete set of firm-level control variables (market capitalization, book-to-market ratio, return on assets, analyst coverage, and stock turnover). Columns (2) through (6) systematically exclude one control variable at a time to evaluate the stability and robustness of the sentiment coefficient estimates across different model specifications. This approach helps ensure that the documented sentiment effects are not driven by correlations with specific control variables. Standard errors are reported in parentheses and are clustered at the firm and time level. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

Table C.14: Control Sensibility Test - LM Sentiment using the Unified Document (Subsample: 2012-2024)

Dependent Variable: Immediate Cumulative Abnormal Returns (CAR)						
	(1)	(2)	(3)	(4)	(5)	(6)
Intercept	0.0570**	-0.0090	0.0576**	0.0586**	0.0568**	0.0566**
	(0.0286)	(0.0112)	(0.0292)	(0.0279)	(0.0277)	(0.0284)
Pos_LM	0.0070***	0.0068**	0.0072***	0.0069***	0.0071***	0.0072***
	(0.0025)	(0.0026)	(0.0024)	(0.0024)	(0.0024)	(0.0025)
Neg_LM	-0.0047	-0.0062	-0.0049	-0.0048	-0.0048	-0.0046
	(0.0052)	(0.0051)	(0.0053)	(0.0050)	(0.0052)	(0.0053)
ln(Market Cap)	-0.0076***		-0.0077***	-0.0074***	-0.0077***	-0.0068***
	(0.0028)		(0.0028)	(0.0027)	(0.0030)	(0.0025)
ln(B/M)	0.0006	0.0035		0.0002	0.0005	0.0004
	(0.0024)	(0.0027)		(0.0024)	(0.0022)	(0.0023)
ROA	0.0001	0.0001	0.0001		0.0001	0.0001
	(0.0001)	(0.0001)	(0.0001)		(0.0001)	(0.0001)
ln(1 + Coverage)	-0.0008	-0.0056	-0.0005	-0.0011		-0.0007
	(0.0038)	(0.0049)	(0.0035)	(0.0038)		(0.0038)
ln(Turnover)	-0.0015	-0.0011	-0.0035	-0.0036	-0.0035	
	(0.0027)	(0.0019)	(0.0025)	(0.0029)	(0.0027)	
Industry Dummies	Yes	Yes	Yes	Yes	Yes	Yes
Year Dummies	Yes	Yes	Yes	Yes	Yes	Yes
Quarter Dummies	Yes	Yes	Yes	Yes	Yes	Yes
N	1226	1241	1242	1231	1226	1226
R^2	0.0355	0.0295	0.0359	0.0355	0.0354	0.0341
Adjusted R^2	0.0104	0.0006	0.0125	0.0097	0.0112	0.0098
Sample Period	2012-2024	2012-2024	2012-2024	2012-2024	2012-2024	2012-2024

Notes: This table presents results from control sensitivity tests examining the robustness of LM Dictionary sentiment measures applied to unified earnings call transcripts for the subsample period 2012-2024. Each transcript is analyzed as a single document rather than being segmented by section. The dependent variable is the cumulative abnormal return (CAR), measuring stock price reactions around the earnings call date. Column (1) presents the baseline specification with the complete set of firm-level control variables (market capitalization, book-to-market ratio, return on assets, analyst coverage, and stock turnover). Columns (2) through (6) systematically exclude one control variable at a time to evaluate the stability and robustness of the sentiment coefficient estimates across different model specifications. This approach helps ensure that the documented sentiment effects are not driven by correlations with specific control variables. Standard errors are reported in parentheses and are clustered at the firm and time level. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

Table C.15: Control Sensibility Test - FinBERT Sentiment using the Unified Document (Subsample: 2012-2024)

Dependent Variable: Immediate Cumulative Abnormal Returns (CAR)						
	(1)	(2)	(3)	(4)	(5)	(6)
Intercept	0.0633**	-0.0071	0.0627**	0.0642**	0.0623**	0.0629**
	(0.0283)	(0.0109)	(0.0304)	(0.0279)	(0.0279)	(0.0290)
Pos_FinBERT	0.0006	0.0010	0.0012	0.0012	0.0013	0.0008
	(0.0034)	(0.0036)	(0.0034)	(0.0033)	(0.0031)	(0.0034)
Neg_FinBERT	-0.0122	-0.0093	-0.0122	-0.0118**	-0.0110	-0.0115
	(0.0122)	(0.0123)	(0.0122)	(0.0123)	(0.0123)	(0.0118)
ln(Market Cap)	-0.0080***		-0.0080***	-0.0078***	-0.0083***	-0.0072***
	(0.0028)		(0.0030)	(0.0027)	(0.0030)	(0.0024)
ln(B/M)	0.0002	0.0032		-0.0003	-0.0001	-0.0000
	(0.0025)	(0.0028)		(0.0024)	(0.0023)	(0.0024)
ROA	0.0001	0.0001	0.0001		0.0001	0.0001
	(0.0001)	(0.0001)	(0.0001)		(0.0001)	(0.0001)
ln(1 + Coverage)	-0.0022	-0.0072	-0.0020	-0.0023		-0.0019
	(0.0040)	(0.0052)	(0.0036)	(0.0039)		(0.0040)
ln(Turnover)	-0.0036	-0.0022	-0.0037	-0.0037	-0.0036	
	(0.0028)	(0.0021)	(0.0029)	(0.0030)	(0.0028)	
Industry Dummies	Yes	Yes	Yes	Yes	Yes	Yes
Year Dummies	Yes	Yes	Yes	Yes	Yes	Yes
Quarter Dummies	Yes	Yes	Yes	Yes	Yes	Yes
N	1226	1241	1242	1231	1226	1226
R^2	0.0313	0.0193	0.0319	0.0297	0.0311	0.0298
Adjusted R^2	0.0061	-0.0051	0.0080	0.0036	0.0067	0.0054
Sample Period	2012-2024	2012-2024	2012-2024	2012-2024	2012-2024	2012-2024

Notes: This table presents results from control sensitivity tests examining the robustness of FinBERT sentiment measures applied to unified earnings call transcripts for the subsample period 2012-2024. Each transcript is analyzed as a single document rather than being segmented by section. The dependent variable is the cumulative abnormal return (CAR), measuring stock price reactions around the earnings call date. Column (1) presents the baseline specification with the complete set of firm-level control variables (market capitalization, book-to-market ratio, return on assets, analyst coverage, and stock turnover). Columns (2) through (6) systematically exclude one control variable at a time to evaluate the stability and robustness of the sentiment coefficient estimates across different model specifications. This approach helps ensure that the documented sentiment effects are not driven by correlations with specific control variables. Standard errors are reported in parentheses and are clustered at the firm and time level. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

Table C.16: Control Sensibility Test - GPT Sentiment using the Unified Document (Subsample: 2012-2024)

Dependent Variable: Immediate Cumulative Abnormal Returns (CAR)						
	(1)	(2)	(3)	(4)	(5)	(6)
Intercept	0.0531*	-0.0133	0.0539*	0.0541*	0.0527*	0.0528*
	(0.0297)	(0.0111)	(0.0313)	(0.0292)	(0.0295)	(0.0302)
Pos_GPT	0.0101*	0.0100*	0.0101*	0.0105*	0.0101*	0.0102*
	(0.0060)	(0.0059)	(0.0059)	(0.0058)	(0.0060)	(0.0060)
Neg_GPT	-0.0176***	-0.0208***	-0.0174***	-0.0166***	-0.0166***	-0.0166***
	(0.0058)	(0.0066)	(0.0066)	(0.0059)	(0.0058)	(0.0059)
ln(Market Cap)	-0.0076***		-0.0078***	-0.0075***	-0.0079***	-0.0069***
	(0.0028)		(0.0030)	(0.0027)	(0.0029)	(0.0024)
ln(B/M)	0.0008	0.0037		0.0004	0.0005	0.0006
	(0.0024)	(0.0026)		(0.0023)	(0.0022)	(0.0023)
ROA	0.0001	0.0000	0.0001		0.0001	0.0001
	(0.0001)	(0.0001)	(0.0001)		(0.0001)	(0.0001)
ln(1 + Coverage)	-0.0020	-0.0070	-0.0011	-0.0023		-0.0019
	(0.0038)	(0.0047)	(0.0035)	(0.0038)		(0.0038)
ln(Turnover)	-0.0035	-0.0011	-0.0036	-0.0035	-0.0034	
	(0.0029)	(0.0020)	(0.0027)	(0.0030)	(0.0029)	
Industry Dummies	Yes	Yes	Yes	Yes	Yes	Yes
Year Dummies	Yes	Yes	Yes	Yes	Yes	Yes
Quarter Dummies	Yes	Yes	Yes	Yes	Yes	Yes
N	1226	1241	1242	1231	1226	1226
R^2	0.0371	0.0264	0.0376	0.0359	0.0369	0.0357
Adjusted R^2	0.0121	0.0023	0.0144	0.0103	0.0127	0.0115
Sample Period	2012-2024	2012-2024	2012-2024	2012-2024	2012-2024	2012-2024

Notes: This table presents results from control sensitivity tests examining the robustness of GPT sentiment measures applied to unified earnings call transcripts for the subsample period 2012-2024. Each transcript is analyzed as a single document rather than being segmented by section. The dependent variable is the cumulative abnormal return (CAR), measuring stock price reactions around the earnings call date. Column (1) presents the baseline specification with the complete set of firm-level control variables (market capitalization, book-to-market ratio, return on assets, analyst coverage, and stock turnover). Columns (2) through (6) systematically exclude one control variable at a time to evaluate the stability and robustness of the sentiment coefficient estimates across different model specifications. This approach helps ensure that the documented sentiment effects are not driven by correlations with specific control variables. Standard errors are reported in parentheses and are clustered at the firm and time level. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

Table C.17: Control Sensibility Test - LM Sentiment using Divided Documents

Dependent Variable: Immediate Cumulative Abnormal Returns (CAR)						
	(1)	(2)	(3)	(4)	(5)	(6)
Intercept	0.0541	-0.0191	0.0669	0.0519	0.0529	0.0477
	(0.0500)	(0.0146)	(0.0441)	(0.0492)	(0.0495)	(0.0455)
Pos_pres_LM	-0.0096	-0.0070	-0.0088	-0.0088	-0.0100	-0.0093
	(0.0087)	(0.0085)	(0.0086)	(0.0084)	(0.0088)	(0.0086)
Neg_pres_LM	-0.0254	-0.0268	-0.0236	-0.0245	-0.0256	-0.0230
	(0.0185)	(0.0198)	(0.0171)	(0.0183)	(0.0182)	(0.0165)
Pos_qa_LM	0.0094***	0.0094***	0.0083**	0.0092***	0.0094***	0.0093***
	(0.0035)	(0.0032)	(0.0036)	(0.0034)	(0.0035)	(0.0036)
Neg_qa_LM	-0.0200**	-0.0201*	-0.0199**	-0.0198**	-0.0202*	-0.0201*
	(0.0100)	(0.0103)	(0.0099)	(0.0100)	(0.0100)	(0.0101)
ln(Market Cap)	-0.0072*		-0.0072*	-0.0069*	-0.0074*	-0.0066*
	(0.0042)		(0.0040)	(0.0041)	(0.0043)	(0.0038)
ln(B/M)	0.0009	0.0032		0.0006	0.0007	0.0005
	(0.0024)	(0.0029)		(0.0022)	(0.0022)	(0.0023)
ROA	0.0001	0.0000	0.0001		0.0001	0.0001
	(0.0001)	(0.0001)	(0.0001)		(0.0001)	(0.0001)
ln(1 + Coverage)	-0.0016	-0.0062	-0.0012	-0.0019		-0.0014
	(0.0035)	(0.0045)	(0.0033)	(0.0035)		(0.0035)
ln(Turnover)	-0.0035	-0.0016	-0.0023	-0.0029	-0.0034	
	(0.0034)	(0.0026)	(0.0023)	(0.0032)	(0.0034)	
Industry Dummies	Yes	Yes	Yes	Yes	Yes	Yes
Year Dummies	Yes	Yes	Yes	Yes	Yes	Yes
Quarter Dummies	Yes	Yes	Yes	Yes	Yes	Yes
N	1385	1400	1457	1412	1385	1385
R^2	0.0418	0.0298	0.0400	0.0397	0.0417	0.0403
Adjusted R^2	0.0111	-0.0002	0.0115	0.0103	0.0117	0.0102
Sample Period	2002-2024	2002-2024	2002-2024	2002-2024	2002-2024	2002-2024

Notes: This table reports the results from panel regressions examining LM Dictionary sentiment measures separately for Presentation and Q&A segments of earnings calls. The dependent variable is the cumulative abnormal return (CAR). Pos_pres_LM and Neg_pres_LM capture positive and negative sentiment in presentations, while Pos_qa_LM and Neg_qa_LM measure sentiment in Q&A segments. Column (1) presents the baseline specification with all control variables (market capitalization, book-to-market ratio, return on assets, analyst coverage, and stock turnover). Columns (2) through (6) systematically exclude one control variable at a time to evaluate robustness. Standard errors are reported in parentheses and are clustered at the firm and time level. ***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively.

Table C.18: Control Sensibility Test - FinBERT Sentiment using Divided Documents

Dependent Variable: Immediate Cumulative Abnormal Returns (CAR)						
	(1)	(2)	(3)	(4)	(5)	(6)
Intercept	0.0472	-0.0252*	0.0588	0.0457	0.0449	0.0415
	(0.0435)	(0.0134)	(0.0411)	(0.0429)	(0.0420)	(0.0392)
Pos_pres_FinBERT	0.0066*	0.0065*	0.0080**	0.0066*	0.0067*	0.0068*
	(0.0036)	(0.0039)	(0.0035)	(0.0035)	(0.0035)	(0.0036)
Neg_pres_FinBERT	0.0019	0.0022	0.0021	0.0019	0.0019	0.0016
	(0.0088)	(0.0086)	(0.0077)	(0.0087)	(0.0088)	(0.0088)
Pos_qa_FinBERT	-0.0002	0.0027	-0.0003	-0.0001	0.0004	0.0003
	(0.0058)	(0.0056)	(0.0059)	(0.0056)	(0.0053)	(0.0056)
Neg_qa_FinBERT	-0.0474***	-0.0455***	-0.0459***	-0.0470***	-0.0459***	-0.0454***
	(0.0122)	(0.0143)	(0.0119)	(0.0129)	(0.0121)	(0.0110)
ln(Market Cap)	-0.0073*		-0.0072*	-0.0070*	-0.0075*	-0.0067*
	(0.0041)		(0.0040)	(0.0040)	(0.0042)	(0.0037)
ln(B/M)	0.0004	0.0025		0.0001	0.0001	0.0000
	(0.0024)	(0.0031)		(0.0023)	(0.0022)	(0.0023)
ROA	0.0001	0.0000	0.0001		0.0001	0.0001
	(0.0001)	(0.0001)	(0.0001)		(0.0001)	(0.0001)
ln(1 + Coverage)	-0.0022	-0.0064	-0.0018	-0.0024		-0.0018
	(0.0039)	(0.0050)	(0.0034)	(0.0039)		(0.0038)
ln(Turnover)	-0.0032	-0.0013	-0.0023	-0.0028	-0.0032	
	(0.0034)	(0.0026)	(0.0022)	(0.0032)	(0.0033)	
Industry Dummies	Yes	Yes	Yes	Yes	Yes	Yes
Year Dummies	Yes	Yes	Yes	Yes	Yes	Yes
Quarter Dummies	Yes	Yes	Yes	Yes	Yes	Yes
N	1385	1400	1457	1412	1385	1385
R^2	0.0353	0.0235	0.0357	0.0334	0.0350	0.0339
Adjusted R^2	0.0043	-0.0068	0.0071	0.0038	0.0048	0.0037
Sample Period	2002-2024	2002-2024	2002-2024	2002-2024	2002-2024	2002-2024

Notes: This table reports the results from panel regressions examining FinBERT sentiment measures separately for Presentation and Q&A segments of earnings calls. The dependent variable is the cumulative abnormal return (CAR). Pos_pres_FinBERT and Neg_pres_FinBERT capture positive and negative sentiment in presentations, while Pos_qa_FinBERT and Neg_qa_FinBERT measure sentiment in Q&A segments. Column (1) presents the baseline specification with all control variables (market capitalization, book-to-market ratio, return on assets, analyst coverage, and stock turnover). Columns (2) through (6) systematically exclude one control variable at a time to evaluate robustness. Standard errors are reported in parentheses and are clustered at the firm and time level. ***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively.

Table C.19: Control Sensibility Test - GPT Sentiment using Divided Documents

Dependent Variable: Immediate Cumulative Abnormal Returns (CAR)						
	(1)	(2)	(3)	(4)	(5)	(6)
Intercept	0.0230	-0.0397**	0.0504	0.0213	0.0202	0.0170
	(0.0420)	(0.0156)	(0.0396)	(0.0414)	(0.0413)	(0.0376)
Pos_pres_GPT	0.0071	0.0078	0.0075	0.0073	0.0070	0.0075
	(0.0073)	(0.0071)	(0.0070)	(0.0071)	(0.0074)	(0.0070)
Neg_pres_GPT	-0.0226*	-0.0241*	-0.0208	-0.0229*	-0.0226*	-0.0220
	(0.0136)	(0.0143)	(0.0133)	(0.0136)	(0.0137)	(0.0136)
Pos_qa_GPT	0.0169***	0.0176***	0.0156***	0.0168***	0.0169***	0.0167***
	(0.0054)	(0.0056)	(0.0053)	(0.0053)	(0.0054)	(0.0053)
Neg_qa_GPT	-0.0027	-0.0021	-0.0030	-0.0026	-0.0029	-0.0027
	(0.0078)	(0.0076)	(0.0075)	(0.0076)	(0.0081)	(0.0080)
ln(Market Cap)	-0.0062		-0.0064	-0.0058	-0.0065	-0.0056
	(0.0040)		(0.0040)	(0.0039)	(0.0041)	(0.0036)
ln(B/M)	0.0019	0.0037		0.0016	0.0015	0.0015
	(0.0025)	(0.0031)		(0.0023)	(0.0022)	(0.0023)
ROA	0.0001	0.0000	0.0000		0.0001	0.0001
	(0.0001)	(0.0001)	(0.0001)		(0.0001)	(0.0001)
ln(1 + Coverage)	-0.0029	-0.0063	-0.0021	-0.0031		-0.0026
	(0.0036)	(0.0045)	(0.0031)	(0.0036)		(0.0035)
ln(Turnover)	-0.0034	-0.0017	-0.0023	-0.0030	-0.0033	
	(0.0035)	(0.0026)	(0.0023)	(0.0033)	(0.0034)	
Industry Dummies	Yes	Yes	Yes	Yes	Yes	Yes
Year Dummies	Yes	Yes	Yes	Yes	Yes	Yes
Quarter Dummies	Yes	Yes	Yes	Yes	Yes	Yes
N	1385	1400	1457	1412	1385	1385
R^2	0.0560	0.0474	0.0535	0.0547	0.0556	0.0545
Adjusted R^2	0.0257	0.0180	0.0254	0.0257	0.0260	0.0249
Sample Period	2002-2024	2002-2024	2002-2024	2002-2024	2002-2024	2002-2024

Notes: This table reports the results from panel regressions examining GPT sentiment measures separately for Presentation and Q&A segments of earnings calls. The dependent variable is the cumulative abnormal return (CAR). Pos_pres_GPT and Neg_pres_GPT capture positive and negative sentiment in presentations, while Pos_qa_GPT and Neg_qa_GPT measure sentiment in Q&A segments. Column (1) presents the baseline specification with all control variables (market capitalization, book-to-market ratio, return on assets, analyst coverage, and stock turnover). Columns (2) through (6) systematically exclude one control variable at a time to evaluate robustness. Standard errors are reported in parentheses and are clustered at the firm and time level. ***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively.

Table C.20: Control Sensibility Test - LM Sentiment using Q&A Segment Only

Dependent Variable: Immediate Cumulative Abnormal Returns (CAR)						
	(1)	(2)	(3)	(4)	(5)	(6)
Intercept	0.0440	-0.0250**	0.0604**	0.0426	0.0416	0.0382
	(0.0287)	(0.0098)	(0.0286)	(0.0287)	(0.0277)	(0.0259)
Pos_qa_LM	0.0091***	0.0092***	0.0081**	0.0089***	0.0090***	0.0090***
	(0.0033)	(0.0031)	(0.0037)	(0.0032)	(0.0034)	(0.0034)
Neg_qa_LM	-0.0179***	-0.0184***	-0.0179***	-0.0179***	-0.0181***	-0.0181***
	(0.0068)	(0.0068)	(0.0068)	(0.0068)	(0.0068)	(0.0069)
ln(Market Cap)	-0.0070***		-0.0071***	-0.0067***	-0.0073***	-0.0064***
	(0.0026)		(0.0025)	(0.0024)	(0.0027)	(0.0023)
ln(B/M)	0.0011	0.0032		0.0008	0.0008	0.0007
	(0.0024)	(0.0024)		(0.0023)	(0.0022)	(0.0022)
ROA	0.0001	0.0000	0.0001		0.0001	0.0001
	(0.0001)	(0.0001)	(0.0001)		(0.0001)	(0.0001)
ln(1 + Coverage)	-0.0023	-0.0066	-0.0018	-0.0025		-0.0020
	(0.0035)	(0.0045)	(0.0032)	(0.0035)		(0.0035)
ln(Turnover)	-0.0033	-0.0015	-0.0022	-0.0028	-0.0032	
	(0.0024)	(0.0018)	(0.0017)	(0.0024)	(0.0024)	
Industry Dummies	Yes	Yes	Yes	Yes	Yes	Yes
Year Dummies	Yes	Yes	Yes	Yes	Yes	Yes
Quarter Dummies	Yes	Yes	Yes	Yes	Yes	Yes
N	1385	1400	1457	1412	1385	1385
R^2	0.0400	0.0286	0.0385	0.0381	0.0398	0.0386
Adjusted R^2	0.0107	0.0001	0.0114	0.0101	0.0112	0.0100
Sample Period	2002-2024	2002-2024	2002-2024	2002-2024	2002-2024	2002-2024

Notes: This table reports the results from panel regressions examining LM Dictionary sentiment measures using only the Q&A segments of earnings calls. The dependent variable is the cumulative abnormal return (CAR). Pos_qa_LM and Neg_qa_LM measure positive and negative sentiment in Q&A segments, respectively. Column (1) presents the baseline specification with all control variables (market capitalization, book-to-market ratio, return on assets, analyst coverage, and stock turnover). Columns (2) through (6) systematically exclude one control variable at a time to evaluate robustness. Standard errors are reported in parentheses and are clustered at the firm and time level. ***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively.

Table C.21: Control Sensibility Test - FinBERT Sentiment using Q&A Segment Only

Dependent Variable: Immediate Cumulative Abnormal Returns (CAR)						
	(1)	(2)	(3)	(4)	(5)	(6)
Intercept	0.0495*	-0.0223**	0.0626**	0.0481*	0.0467*	0.0434*
	(0.0269)	(0.0107)	(0.0265)	(0.0267)	(0.0258)	(0.0241)
Pos_qa_FinBERT	0.0009	0.0037	0.0010	0.0008	0.0016	0.0013
	(0.0068)	(0.0065)	(0.0067)	(0.0063)	(0.0066)	(0.0067)
Neg_qa_FinBERT	-0.0474***	-0.0455***	-0.0459***	-0.0470***	-0.0456***	-0.0455***
	(0.0112)	(0.0126)	(0.0116)	(0.0118)	(0.0109)	(0.0104)
ln(Market Cap)	-0.0072***		-0.0071***	-0.0070***	-0.0075***	-0.0067***
	(0.0025)		(0.0025)	(0.0023)	(0.0026)	(0.0022)
ln(B/M)	0.0004	0.0026		0.0001	0.0002	0.0000
	(0.0023)	(0.0024)		(0.0022)	(0.0021)	(0.0021)
ROA	0.0001	0.0000	0.0001		0.0001	0.0001
	(0.0001)	(0.0001)	(0.0001)		(0.0001)	(0.0001)
ln(1 + Coverage)	-0.0026	-0.0068	-0.0022	-0.0027		-0.0022
	(0.0036)	(0.0046)	(0.0034)	(0.0036)		(0.0036)
ln(Turnover)	-0.0034	-0.0015	-0.0023	-0.0029	-0.0033	
	(0.0024)	(0.0019)	(0.0017)	(0.0025)	(0.0024)	
Industry Dummies	Yes	Yes	Yes	Yes	Yes	Yes
Year Dummies	Yes	Yes	Yes	Yes	Yes	Yes
Quarter Dummies	Yes	Yes	Yes	Yes	Yes	Yes
N	1385	1400	1457	1412	1385	1385
R^2	0.0332	0.0215	0.0325	0.0313	0.0329	0.0317
Adjusted R^2	0.0037	-0.0073	0.0052	0.0031	0.0041	0.0029
Sample Period	2002-2024	2002-2024	2002-2024	2002-2024	2002-2024	2002-2024

Notes: This table reports the results from panel regressions examining FinBERT sentiment measures using only the Q&A segments of earnings calls. The dependent variable is the cumulative abnormal return (CAR). Pos_qa_FinBERT and Neg_qa_FinBERT measure positive and negative sentiment in Q&A segments, respectively. Column (1) presents the baseline specification with all control variables (market capitalization, book-to-market ratio, return on assets, analyst coverage, and stock turnover). Columns (2) through (6) systematically exclude one control variable at a time to evaluate robustness. Standard errors are reported in parentheses and are clustered at the firm and time level. ***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively.

Table C.22: Control Sensibility Test - GPT Sentiment using Q&A Segment Only

Dependent Variable: Immediate Cumulative Abnormal Returns (CAR)						
	(1)	(2)	(3)	(4)	(5)	(6)
Intercept	0.0301	-0.0347***	0.0535**	0.0285	0.0275	0.0240
	(0.0222)	(0.0104)	(0.0251)	(0.0220)	(0.0208)	(0.0193)
Pos_qa_GPT	0.0177***	0.0186***	0.0166***	0.0177***	0.0177***	0.0176***
	(0.0047)	(0.0052)	(0.0046)	(0.0046)	(0.0047)	(0.0048)
Neg_qa_GPT	-0.0074*	-0.0071*	-0.0081*	-0.0074*	-0.0074*	-0.0077*
	(0.0043)	(0.0039)	(0.0041)	(0.0042)	(0.0045)	(0.0040)
ln(Market Cap)	-0.0064***		-0.0065***	-0.0061***	-0.0067***	-0.0058***
	(0.0022)		(0.0023)	(0.0020)	(0.0023)	(0.0019)
ln(B/M)	0.0013	0.0032		0.0010	0.0009	0.0009
	(0.0024)	(0.0025)		(0.0021)	(0.0022)	(0.0022)
ROA	0.0001	0.0000	0.0001		0.0001	0.0001
	(0.0001)	(0.0001)	(0.0001)		(0.0001)	(0.0001)
ln(1 + Coverage)	-0.0025	-0.0061	-0.0019	-0.0028		-0.0022
	(0.0035)	(0.0042)	(0.0031)	(0.0034)		(0.0035)
ln(Turnover)	-0.0035	-0.0017	-0.0024	-0.0031	-0.0034	
	(0.0023)	(0.0020)	(0.0017)	(0.0024)	(0.0023)	
Industry Dummies	Yes	Yes	Yes	Yes	Yes	Yes
Year Dummies	Yes	Yes	Yes	Yes	Yes	Yes
Quarter Dummies	Yes	Yes	Yes	Yes	Yes	Yes
N	1385	1400	1457	1412	1385	1385
R^2	0.0511	0.0418	0.0487	0.0496	0.0508	0.0495
Adjusted R^2	0.0222	0.0136	0.0218	0.0219	0.0225	0.0212
Sample Period	2002-2024	2002-2024	2002-2024	2002-2024	2002-2024	2002-2024

Notes: This table reports the results from panel regressions examining GPT sentiment measures using only the Q&A segments of earnings calls. The dependent variable is the cumulative abnormal return (CAR). Pos_qa_GPT and Neg_qa_GPT measure positive and negative sentiment in Q&A segments, respectively. Column (1) presents the baseline specification with all control variables (market capitalization, book-to-market ratio, return on assets, analyst coverage, and stock turnover). Columns (2) through (6) systematically exclude one control variable at a time to evaluate robustness. Standard errors are reported in parentheses and are clustered at the firm and time level. ***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively.