Is Talk Cheap? The Predictive Power of Online Investor Sentiment on Post-Earnings Announcement Drift

Matthew McCrea

Submitted for MA Economics and Mathematics (Hons)

Supervisor: Dr. Alex Kostylev

Word Count: 9,882

Abstract:

This paper explores how online investor sentiment, as captured through social media analysis, influences the Post Earnings Announcement Drift (PEAD). We contribute to the literature by uncovering, for the first time, compelling evidence that sentiment holds valuable predictive information on this market anomoly. We utilise event study methodology and traditional panel data methods to investigate the presence of PEAD and the significance of investor sentiment alongside attention, earnings surprise, and past returns. We find that social media sentiment does contain persistent predictive information on PEAD across a range of measurements and model specifications which constitute a market inefficiency with the potential for arbitrage.

Acknowledgements

I would like to thank Dr. Alex Kostylev for his patience throughout this process, especially in discussing model specifications with me. I am grateful for the invaluable support of my friends and family, who endured my "thinking out loud", and for my brother Jude, who inspired the following research.

Table of contents

1 Introduction

Efficient markets are the cornerstone of modern economies. In theory, asset prices reflect all available information, ensuring capital flows to the most productive uses (Fama, 1970) and is reallocated in response to economic changes (Malkiel, 2003). However, persistent pricing anomalies challenge this ideal to the detriment of the wider economic production and efficiency.

Information efficiency plays a key role in this resource allocation as inefficient absorption distorts price signals and misdirects capital. Anomalies in price behaviour challenge the Efficient Market Hypothesis (EMH) and have been a major focus of financial economics (Yalçın, 2010).

This paper investigates the most enduring market anomaly, the Post-Earnings Announcement Drift (PEAD), which describes how stock prices continue to move in the direction of the earnings surprise long after the announcement. PEAD was first identified by Ball and Brown (1968) and violates the EMH by providing reproducible and predictable returns without new information.

The research question of this paper is "Does the sentiment of online investors hold information on PEAD?" We contribute to the literature by examining investor sentiment as a novel predictor of PEAD. To the authors knowledge there is no similar research bridging the gap between PEAD research and sentiment analysis.

We examine whether sentiment measured via finance-focused social media platform StockTwits, contains information on PEAD.

The power of traditional predictors of PEAD, which include earnings surprise and institutional ownership (Bernard & Thomas, 1989; Hirshleifer et al., 2009), has diminished over time as traders exploited the anomaly, which reduces its magnitude (Chordia et al., 2009). In this way the study of PEAD enhances market efficiency, motivating this study in order to increase market efficiency. Novel data sources are now required to capture the remaining PEAD, and online investor sentiment may be one such source. For the first time in literature, we utilise this dataset through advanced data analysis tailored to the online context.

Outside of earnings call periods, sentiment can cause price movements even in the absence of new information, violating the EMH (Chang et al., 2016) and may drive bubbles, crashes, or feedback loops. Sentiment can predict abnormal returns both for specific stocks and composite indices (Bollen et al., 2011; Renault, 2020; Danieli & Denenzis, 2024).

We pay particular attention to retail investor sentiment throughout this study. Retail investors often exhibit behavioural biases during earnings events, trading aggressively and emotionally and losing large sums (Barber & Odean, 2000; JPMorgan Chase Institute, 2022). Their emotional buying amplifies price changes and contributes concretely to PEAD (So, 2022; Friedman & Zeng, 2022). This finding further motivates study to understand the impacts of this behaviour to increase market efficiency and reduce irrational, wealth eroding trading. Secondly, StockTwits largely caters to retail investors, and we believe we have a high incidence of them in the sample, making it an appropriate data source for studying this group (Tan, 2021).

To assess online sentiment's predictive power we tested three hypotheses:

Our first sub hypothesis (SH_1) is that more extreme sentiment is associated with reduced PEAD.

This hypothesis stems from literature on retail investor behaviour (Barber & Odean, 2000; So, 2022) which shows that emotional and momentum-driven trading offset market underreaction and reduce drift.

Our empirical results found that measures of average sentiment are statistically significant predictors of PEAD at least 20 days post-announcement. More positive sentiment is correlated

with reduced PEAD, and this finding was robust across model specifications displaying strong support for SH_1.

Our second sub hypothesis (SH_2) is that greater variation in sentiment, as measured by the interquartile range, predicts more negative PEAD irrespective of the earnings surprise.

We hypothesise that after a positive earnings surprise, increased disagreement might prevent continued buying which diminishes PEAD, while a negative earnings surprise likely leads to a greater variation in sentiment which is also associated with negative PEAD.

We found that the range of sentiment has consistent predictive power on PEAD. Dictionary-based sentiment scoring performed best in univariate regressions, while Natural Language Processing methods proved more effective in multivariate settings. These results strongly support SH_2.

Our third sub hypothesis (SH_3) is that a higher percentage of negative posts correlates with more negative PEAD.

Positive online sentiment skew might indicate that when a relatively high percentage of posts are negative there is concern or pessimism which might fuel dynamics similar to the motivation for SH_2. Literature supports that negative sentiment has a stronger price impact than positive sentiment (Danieli & Denenzis, 2024).

Our results found that increased negativity in online commentary held information on PEAD. The percentage of negative posts was not consistently significant on its own but gained predictive power when used in combination with other sentiment measures in support of SH_3.

Our research question was resolutely answered through these three sub hypotheses, which explored distinct parts of investor sentiment, finding strong evidence that social media sentiment holds information on PEAD.

Contrary to much of the literature, we find no consistent evidence that the earnings result itself holds information on PEAD. In fact, only measures of online investor attention and the pre-earnings announcement drift showed persistent explanatory power outside of sentiment variables.

The structure of this paper is as follows: Section 2 reviews the relevant literature and theoretical background. Section 3 describes the dataset and the construction of variables. Section 4 outlines our empirical methodology. Section 5 presents and interprets the empirical results. Section 6 summarises our conclusions and discusses implications for future research.

2 Literature

The following section discusses the literature and theoretical explanations for PEAD, its evolution through time, literature on using social media sentiment as a predictor of market returns, and methodologies in literature for the construction of relevant metrics.

2.1 Post Earnings Adjustment Drift

The Post Earnings Announcement Drift (PEAD) was first identified by Ball and Brown (1968). They found that markets consistently underreact to earnings surprises as stock prices continued drifting in the direction of the earnings surprise for over a month. This finding sparked sustained inquiry into the informational frictions behind this slow adjustment.

PEAD has since been confirmed globally, across developed and emerging markets, including the US, Europe, Africa, India, and South Korea (Fink, 2020). PEAD is distinct from other anomalies such as price momentum and the accrual anomaly (Bohl et al., 2016; Louis & Sun, 2011). Daniel et al. (2020) finds that controlling for PEAD renders most short-term anomalies insignificant.

2.1.1 Explanations for the Post Earnings Adjustment Drift

The two leading explanations for PEAD are behavioural underreaction due to limited attention, and market frictions that delay arbitrage (Fink, 2020).

The behavioural view argues that investors are slow to process earnings news. Announcements on busy days, Fridays, or during sports events all show stronger PEAD. Inattention is magnified in complex firms and during down markets (Fink, 2020).

Retail investors are particularly susceptible. Battalio and Mendenhall (2005) show that less sophisticated traders rely on random walk expectations rather than analyst forecasts and misjudge surprises. These inaccurate expectations could accentuate/diminish PEAD.

Digital metrics of investor attention, such as SEC database traffic and social media engagement have been shown to attenuate PEAD (Li et al., 2019). Increased Google searches can have an effect in either direction depending on the profile of the investor searching (Chi and Shanthikumar, 2016). Wu (2019) finds that abnormal social media attention can even offset the negative effects of earnings misses.

Another explanation argues that market frictions like short-selling constraints, wider bid-ask spreads, and post-call liquidity shocks limit arbitrage opportunities (Fink, 2020). These frictions are higher for small and mid-cap stocks with less liquidity.

2.1.2 Decline in Post Earnings Adjustment Drift over time

Evidence suggests that its magnitude of PEAD has declined, particularly in large-cap stocks. Martineau (2021) finds that earnings responsiveness in the S&P 1500 increased 15 times from 1984–1988 to 2011–2015, suggesting increased market efficiency to earnings information. Martineau attributes this to the widespread use of limit orders and increased access to data.

Numerous firm characteristics influence PEAD. Larger firms, those with more institutional attention, and high-reputation firms face reduced PEAD (Chan et al., 1996; Pfarrer et al., 2010). Firms with ‘celebrity’ or ‘high reputation’, face asymmetric reactions, receiving outsized reward for positive surprises and lesser punishment for negative ones. Son et al. (2018) show that firms with historically strong PEAD continue to do so, suggesting firm specific effects.

2.2 Social Media Sentiment as Price Movement Predictor

Social sentiment is a relatively new tool in financial research. Bollen et al. (2011) showed Twitter sentiment could predict the Dow Jones Index, unlocking a new source of market information. Danieli and Denenzis (2024) found that social media sentiment predicted returns in European equities and that negative sentiment could predict multi-day returns. Renault (2017) showed that reaction to sentiment signals are muted during days with macroeconomic news days. Trading the S&P 500 using sentiment in just the final 30 minutes of the trading day yielded a 4.55% annualised return. Bollen et al. also reported up to 87% predictive accuracy using Twitter sentiment on index values.

As Twitter grew API access became more limited (Davidson, 2023), and finance-specific platforms gained popularity for sentiment studies, most notably StockTwits (Di Wu, 2019; Renault, 2020) but also Yahoo Finance forums (Nguyen et al., 2015).

StockTwits is now a key source for online sentiment as it contains focused, finance specific content with minimal noise (Divernois, 2024), and its predictive ability is at least comparable to Twitter (Renault, 2021). Renault (2020) and Divernois (2024) confirm that StockTwits sentiment scores correlate with short-term price movements, though they do not investigate PEAD.

Di Wu (2019) showed that post volume on StockTwits held predictive power on PEAD, highlighting attention effects. Our study adds to this finding by introducing sentiment measures to that method.

While StockTwits do not disclose user demographics, the CEO attributes their growth to the rise of 'meme stocks', indicating a large retail investor base (Tan, 2021).

We use StockTwits data for its low noise, finance-specific content, predictive strength, and high incidence of retail investors which are ideal for exploring sentiment's influence on PEAD.

2.3 Variable Constructions in Literature

Researchers typically use Ordinary Least Squares (OLS) or Fixed Effects (FE) regressions to estimate the effect of variables on Cumulative Abnormal Returns (CAR) after earnings calls. Historically, pre-call CAR and earnings surprise measures were the most consistent predictors, though these have lost information as PEAD has declined. Recent significant variables include social media investor attention (Di Wu, 2019). The literature often combines novel and control regressors in multivariate models to assess predictive power.

2.3.1 Earnings surprise calculation

Two main methods to determine earning surprise are used. Either the difference between reported EPS and either the analyst consensus estimate, or a time-series based forecast. Both methods are consistently significant, however, analyst estimates are more effective for large companies (Bradshaw et al., 2012). We adopt the analyst estimate measure.

We calculate the earnings surprise measured in dollars, and the percentage earnings surprise. Observations without consistent estimates (66) or with \$0.00 estimates (2) were dropped, leaving 1,172 events. The constructions are as follows where j represents a company and t is the earnings report date:

$$Surprise_{j,t} = Reported\ Earnings_{j,t} - Analyst\ Estimate_{j,t}$$

$$\%\ Surprise_{j,t} = \frac{Reported\ Earnings_{j,t} - Analyst\ Estimate_{j,t}}{|Analyst\ Estimate_{j,t}|} \times 100$$

The absolute value in the denominator ensures the directionality is preserved.

2.3.2 Abnormal Return

Abnormal returns are calculated as the difference between realised and expected returns. The expected return used is typically a market benchmark or an OLS estimate. In the absence of consensus and because our sample comes from the S&P 500, we use a simple market benchmark method. No firm is large enough to significantly impact the index, avoiding estimation bias.

We calculate Cumulative Abnormal Returns by summing the abnormal returns for each day in a window. The cumulative abnormal return in the pre-announcement period is a historically significant regressor for PEAD, potentially capturing investor anticipation, insider trading, or firm-specific momentum effects (Bernard & Thomas, 1990; Ke & Petroni, 2004). These pre-period CARs are included as regressors in our models to test their ongoing informational relevance.

2.3.3 Earnings Surprise Direction

The literature finds that the direction of the earnings surprise determines the direction of PEAD. Cox (2020) notes PEAD being more pronounced for negative results. Cox proposes that investors are more likely to take a passive strategy in response to a negative surprise, extending its impact. The asymmetry has been consistent since 1980 at minimum (Zhang et al., 2024). We construct surprise direction dummies as follows:

$$\text{Positive Dummy}_{j,t} = \begin{cases} 1, & \text{if Surprise}_{j,t} > 0 \\ 0, & \text{otherwise} \end{cases}$$

$$\text{Negative Dummy}_{j,t} = \begin{cases} 1, & \text{if Surprise}_{j,t} < 0 \\ 0, & \text{otherwise} \end{cases}$$

Where j is the company and t is the earnings call date.

2.3.4 Investor Attention

The only study to the authors knowledge of social media measures as a predictor of PEAD is “Does Social Media Get Your Attention” by Di Wu (2019) which measured the culminative mentions about a company on Twitter and StockTwits as a proxy for investor attention. It was found that increased posts after an earnings call increased CAR regardless of the direction of earnings surprise. This effect increased positive PEAD and diminished negative drift. The more frequently discussed companies displayed less sensitivity to the effect. This is a relevant finding for our study as our sample contains some of the most talked about companies worldwide.

2.3.5 Investor Sentiment

The methods for studying sentiment are split between two camps. The first is machine learning-based natural language processing (NLP), which classifies text based on patterns learned from labelled data. Pre-processing is important for NLP as many machine learning models struggle with emojis, bijections, or double negatives and depends heavily on the choice of model, training data, and scoring method. The choice between binary classification or scaled sentiment impacts on both accuracy and the numerical transformations possible to construct sentiment statistics.

The second method is to use dictionaries of positive/negative words to create a sentiment score. These dictionaries are constructed by experts with focus on the most mentioned words in the data set. The specific dictionary used impacts on the scoring accuracy, with dictionaries specially constructed for stock market discussion providing superior accuracy than generalised dictionaries (Renault, 2017).

Both methods are highly effective, Renault (2017) showed that topic specific dictionaries were similarly accurate to ML methods for sentiment scoring and variable construction for index price prediction, achieving an accuracy in NLP classification of over 70%.

In this study, we apply both techniques. We construct the following sentiment variables: the ratio of positive to negative posts, average sentiment score, and range of sentiment scores on the day of the earnings call, using both methods to compare methods and increase consistency.

2.4 Event Study and Window

We apply event study methodology in this work. Event study focuses on the effects of a specific event and company, as opposed to time series which aggregates companies that announce within a period and creates a portfolio to study them.

The event study methodology used in financial economics to assess how quickly and accurately markets incorporate new information into asset prices by tracking the daily effect of the event and thereby observing discrepancies from the EMH in real time (MacKinlay, 1996).

In our case the events are earnings calls, and companies are 119 randomly selected companies from the S&P 500. Abnormal returns are calculated as the difference between a stock's cumulative return and that of the S&P 500. We define a window that runs from 20 days before the call to 20 days after. A combination of increased speed with which the market adjusts to earnings news motivates this window size (Martineau, 2021). The speed with which PEAD becomes insignificant has become very short and we believe that 20 trading days (one month) should cover the effect well. Furthermore, we are investigating the impact of online sentiment and believe that any effects after a month are unlikely to be related to the sentiment of online investors. The attention of the internet moves fast generally and many of the online commentators will move from one earnings call to the next, giving little consideration to events from a month ago.

We further split this window into smaller timeframes of 5 and 10 days to measure the persistence of information.

We define 3 periods:

The "pre-event window" from X days before to 1 day before the earnings call, denoted by [-X, -1].

The "event window" from one day before the call to one day afterwards, denoted by [-1, 1].

The "post-event window", from 1 to X days after the call, denoted by [1, X].

Where we define X as an integer taking the values $X \in \{5, 10, 20\}$ such that we have 7 periods of study in total.

Abnormal returns are calculated at market open the day after the earnings call for all announcement times. This choice is motivated by the short window of PEAD and that price movements occur in overnight trading.

Calculating measures from social media occurs on the day of the call, because the online reaction to earnings call news is rapid but subsides quickly and helps us to avoid spillover effects.

Our testing and results framework are standard OLS and Fixed Effects regressions, with the aim of observing the effect of investor sentiment during earnings calls on PEAD as well as checking for any anticipatory effects. We will predominantly use the FE regressions due to large firm specific, time invariant characteristics and refer to the OLS for robustness checks.

3 Data

To evaluate the potential impact of investor sentiment on PEAD, we compiled a dataset covering 119 randomly selected companies from the S&P 500. The condition that the companies belong to the S&P 500 was chosen to ensure sufficient data existed on in terms of online commentary and accessible financial information. The original sample of 125 was reduced after excluding companies with missing earnings estimates data or no StockTwits posts during the event period.

3.1 Collection

Price and earnings data was collected from DoltHub, an open-source data repository (DoltHub, 2024). Price data includes the OHLCV values for each company under study from November 2024 to April 2021. Earnings data consists of the earnings call date, consensus analyst EPS estimate, reported EPS, and the call timing in relation to market close/open. We note that the EPS reported is the company's preferred metric and is not always GAAP compliant. Changes from GAAP EPS are usually made to better reflect recurring costs and as such may present a better estimate of future cashflows, making them suitable for our study (PwC, 2024).

Social media data was collected from StockTwits (StockTwits, 2025). Each company in the dataset has its own dedicated page on the website, which includes all posts made containing that company's unique cashtag identifier. The data collected is all the text posts available from this page and consisted of $\approx$ 6Mb of text data (including emojis, excluding other media) for each company. This limit creates an unbalanced panel in which more popular companies have shorter coverage. Despite this limit, the higher PEAD among smaller, less-discussed firms should allow us to make significant inference. In total we collected 689,903 total text posts across 111,033 unique company-date pairs.

3.2 Constructing Social Media Measures

Constructing our social media measures largely follow the methods in literature. We contribute to the methodology by applying more advanced machine learning and optimisation methods. What follows is process for the construction of sentiment and attention metrics from social media.

3.2.1 Pre-Processing

To extract sentiment from our data we must pre-process it. We follow the methodology of Renault (2017). The processing was as follows:

1. Standardise dates and time format for ease and accuracy. We change the format to "YYYY-MM-DD 00:00:00" removing month names and difficult AM/PM formatting. E.g. "Oct 15, 2024 5:42 PM" becomes "2024-10-15 17:42:00"
2. Make all text lower case, this is done as the language we use throughout this project (Python) is case specific and will read the same word as different if capitalised.
3. Replace all individual cashtags with a non company specific marker "cashtag" to avoid biases associated with specific companies skewing messages that contain multiple cashtags e.g. "bought $pton, $c and added $jd and $bidu."
4. Replace all numbers with "numbertag" to avoid coincidental sentiment scoring for unrelated numbers.
5. Append words that appear after common negative terms with the prefix "negtag_". This is done to recognise common negations that would otherwise be classified as opposite their intended meaning e.g. "never buy", "not good", "don't sell".
6. Replace the most common emojis with positive and negative tags "emojipos" and "emojineg". Those without a dominant implied sentiment were left as is. This was done by hand,

extracting the 232 emojis in our sample and classifying them, cross referencing with random sampling to ensure accuracy.

This processing allows dictionary and Machine Learning methods to better read and evaluate the sentiment of the text. A sample of raw text and its processed form is shown in Table 3.

3.2.2 Sentiment Metrics

We will create a variety of sentiment metrics using a combination of dictionary and machine learning methods from the text data.

3.2.3 Dictionary scoring

We use the expert dictionary L1 from Renault (2017) to score our processed posts. This was shown to be highly accurate both in 2017 and 2020 (Renault, 2020), it contains the 8000 most common words found in a large sample from StockTwits. Each word was scored by the formula

$$Score = \frac{n_{pos} - n_{neg}}{n_{pos} + n_{neg}}$$

Where n_{pos} and n_{neg} are the total number of appearances in positive/negative posts producing scores bounded between -1 and 1. The total sentiment of a post is taken as the average sentiment of each word within it.

Spot checks on posts revealed minor inconsistencies. The presence of posts containing neutral statements had scores close to zero. To mitigate against any bias posts with a sentiment score $< |0.10|$ were classified as neutral to try and improve accuracy.

3.2.4 Machine Learning Natural Language Processing

We apply machine learning (ML) to sentiment analysis through Natural Language Processing (NLP), a set of methods that allow computers to interpret human language. StockTwits provides an ideal dataset for this through its built-in sentiment flag feature. Users can tag posts as "Bullish" or "Bearish," creating a valuable supervised learning dataset. Out of 689,903 total posts, 138,601 were explicitly labelled by the authors, giving us a robust training set with a clear objective function.

Advances in ML and off-the-shelf NLP tools have made sentiment classification increasingly accessible and accurate. A 65–75% accuracy rate is standard for in the literature (Renault, 2020), which we aim to exceed. After experimentation with Multinomial Naive Bayes and Support Vector Machine, we selected the Maximum Entropy classifier for performance and computational efficiency. This model makes no assumptions about word distributions, unlike other classifier models.

To train the model, we removed the sentiment tags from 138,601 scored posts and labelled them +1 ("Bullish") or -1 ("Bearish"). We vectorized the text into numerical form, enabling the model to learn the contribution of each word toward our objective of sentiment classification. This method allows us to extract consistent, replicable sentiment scores for all posts, which we then apply across the full dataset.

Running the ML model with its default parameters provided a classification output shown in Table 3

Precision measures how often the model is correct when it predicts a given sentiment, while recall indicates how well it identifies all actual instances of that sentiment. The F1-score, a

harmonic mean of precision and recall, balances both. Accuracy reflects the model's overall success rate, while the support is the number of posts in each set.

We used 80% of our labelled posts for training, leaving 27,710 posts in the test set. The "macro average" treats both sentiment classes equally, while the "weighted average" adjusts for the class imbalance, giving more influence to the more frequent positive class. The disparity in recall and f1-scores between the classifications is likely due to the disparity in sample size, with bad performance in classifying negatives.

To improve model accuracy, we adjusted class weights using iterative numerical optimisation based on the Newton-Raphson method. We tested relative weightings for negative versus positive classifications across a range of $x \in [0.1, 5]$. We aim to maximise the f1-score for the negative classification. This objective was chosen because initial results showed the positive classification performed well regardless of weighting. We found a relative weighting of 2.3:1 provided the best model for maximising f1-score in the negative category without reducing performance. The classification report is in Table 3.

This model is accurate and optimises for negative classification. The overall accuracy of 86% is above the literature standard and the 70% performance accuracy in negative classification is sufficiently high. The 90% accuracy for positive classification is higher than anything the author has seen in literature.

We use the confidence level of the NLP classifier as the sentiment score for that post, with a positive sign when positive and negative when negative. Posts with a confidence level under 0.65 are classified as neutral.

3.3 Sentiment Constructions

To increase accuracy further we processed our sample, comparing the NLP and dictionary scores for each message and classifying posts that had opposite signs as neutral with a sentiment of 0. If it was neutral in either method, it was marked as neutral. This was undertaken to remove inconsistencies between methods, as some messages which had extreme values in the dictionary scoring method not in line with their content. Very few posts were in this category on earnings call days with a total of 14 date-company pairs having any, making this issue of little consequence.

We constructed sentiment metrics from the database of agreed sentiment posts from the day of each earnings call. They are constructed using both methods of scoring for each day and company pair. We calculate:

1. The mean score for the company day pair.

$$Mean\ Score_{j,t,m} = \frac{1}{n}\sum_{k=1}^{n} Sentiment_{j,t,k}$$

2. The percentage of negative posts for the company day pair.

$$\%Negative_{j,t,m} = \frac{\sum_{k=1}^{n} \mathbb{1}(Sentiment_{j,t,k} < 0)}{n} \times 100$$

3. The interquartile range of sentiment scores with each scoring method.

$$IQR_{j,t,m} = Q3_{j,t} - Q1_{j,t}$$

Where j represents a specific company, t is the earnings call date, m is the method of scoring, n is the number of posts about that company that day, and $\mathbb{1}$ is an indicator function that takes

the value 1 if the sentiment of a post is negative ($Sentiment_{j,t,k} < 0$), and 0 otherwise. $Q3_{j,t}$ is the third quartile (75th percentile), and $Q1_{j,t}$ is the first quartile (25th percentile) of sentiment scores.

3.4 Attention Metrics

We calculate attention metrics by replicating the method of Di Wu (2019). These metrics are the total posts on a given day and the log of total posts. These measure attention through the quantitate of online content generated.
We construct our variables as follows:

$$Total\ Posts_{j,t} = n_{j,t}$$

$$Log\ Total\ Posts_{j,t} = \log(n_{j,t})$$

Where n is the total number of posts for company j on date t and log is the natural logarithm function.

The summary statistics and distributions for these variables are available in Table 1 and Figure 2 respectively.

The distribution of total posts is heavily skewed whereas the log transforms the distribution to near normal.

The mean total posts exceeds the median (47.7 vs 30) highlighting that the extreme values at the higher end of the distribution are very impactful and that our sample consists of many events with relatively few posts.

3.5 Summary Statistics

The summary statistics on the social media variables on the day of earnings calls are shown in Table 1 and their histogram distributions are in Figure 2.

These statistics agree with existing literature in observing the optimistic bias in online posts, the mean sentiment for each method is positive and the distribution of negative posts is left skewed. The interquartile range has an uptick of points in the 0 to 0.02 range which are events with few posts and thus homogenous sentiment.

Figure 1 shows the distribution of the earnings results metrics defined earlier and their means. In total we observed 1,304 earnings calls from 2024-11-08 to 2021-04-20 with a mean of 11 observations per company. After removing entries missing key earnings data we have 1,174 observations.

Summary statistics for EPS result, surprise, and percentage surprise are found in Table 1. Each has a positive mean and median, suggesting positive skew. The distribution of the earnings surprise indicates that analysts consistently underestimate EPS for our sample through its positive mean and heavy positive tail. Percentage earnings surprise is distributed similarly, though more leptokurtic. Our sample is part of the S&P 500, ensuring they are extremely successful, companies which do not consistently overperform are unlikely to become part of this sample, likely creating our positive bias.

To calculate abnormal returns, we use the return of a firm relative to the S&P 500 index. We calculate returns from market open on day d to market open on day $d + 1$, thereby capturing overnight trading.

The daily abnormal return (AR) is defined as:

$$AR_{j,t,d} = r_{j,t,d} - r_{m,t,d}$$

where $r_{j,t,d}$ is the daily return of stock j on day d from earnings call date t, and $r_{m,t,d}$ is the market return of the S&P 500 on the same day.

We construct Cumulative Abnormal Returns (CAR) over three key windows:

The Pre-Event Period

$$CAR^{-}_{j,t,d} = \sum_{k=-d}^{-2} AR_{j,t,k}$$

Where $CAR^{-}_{j,t,d}$ is the sum of abnormal returns from market close on day d before the call, to market open 1 day before.

The Event Period

$$CAR^{E}_{j,t,d} = \sum_{k=-1}^{1} AR_{j,t,k}$$

Where $CAR^{E}_{j,t,d}$ is the sum of abnormal returns from market open the day before the call, to market close the day afterwards.

The Post-Event Period

$$CAR^{+}_{j,t,d} = \sum_{k=1}^{d} AR_{j,t,k}$$

Where $CAR^{+}_{j,t,d}$ is the sum of abnormal returns for company from market close the day after the call to market close on day d.

The daily abnormal returns allow us to conduct a daily event study to investigate in finer detail the impact of our variables on abnormal return through time to establish the persistence of information contained within.

Histograms of CAR for the 5, 10, and 20 days before/after an earnings call, and the earnings period are in Figure 1. The distribution is relatively symmetric for all time frames and smoothness increases over time due to regression to the mean. The mean and median are very close to 0 for all periods, implying returns are on average in line with the market.

4 Methodology

This study will use event study methodology to investigate the impact of investor sentiment on abnormal returns at the daily level for 20 days before and after an earnings call. We utilise two event study methods. We make use of coefficient plots on the daily abnormal returns in the period to gain a daily view of sentiment effects. We then replicate the literature by using standard OLS and FE regressions to determine sentiment effects on the cumulative abnormal returns for 5, 10, and 20 days after an earnings call.

4.1 Event study and Model Construction

To investigate our research question "Does the sentiment of online investors hold information on PEAD?" we test our sentiment variables for significance and magnitude against the abnormal return in the event windows. We test against the null hypothesis H_0: The abnormal return of a stock has no response to the online sentiment of retail investors.

The variables are split into the literature and novel regressors and given shorthand names, given below:

- Literature Regressors:
  - Pre-Event Abnormal Return (PEAR)
  - Earnings Surprise (Dollars) (ESD)
  - Percentage Earnings Surprise (PES)
  - Surprise Direction (DIR)
  - Total Posts (TP)
  - Log Total Posts (LTP)
- Novel Regressors:
  - Avg. Sentiment Score (Dictionary) (ASD)
  - Avg. Sentiment Score (NLP) (ASN)
  - Percentage Negative Posts (PNP)
  - Interquartile Range (Dictionary) (IQRD)
  - Interquartile Range (NLP) (IQRN)

We employ coefficient plotting, derived from estimating daily abnormal returns with the following Model 1:

$$\begin{aligned} AR_{j,t,d} = \alpha + \sum_{k=-K}^{K} \left(\beta_k \cdot PES_{j,t} \cdot Sentiment_{j,t} \cdot D_{d=k}\right) \\ + \sum_{k=-K}^{K} \gamma_k \cdot PES_{j,t} \cdot D_{d=k} + \sum_{k=-K}^{K} \delta_k \cdot D_{d=k} + \epsilon_{j,t,d} \end{aligned} \tag{1}$$

Where each regressor is referred to by its shortname, *Sentiment* is a placeholder for our sentiment variable under study, the subscript j is the company, t is the date of the earnings call, d is the day relative to the earnings call, $k \in [-20, 20]$, $\{\alpha, \beta, \gamma, \delta\}$ are estimated values, and ϵ is the error term for each observation.

Our interest lies in examining the β coefficients of the interaction terms between sentiment variables and earnings surprise magnitude. By plotting these coefficients across the event period, we can assess how sentiment metrics, adjusted for the magnitude of earnings surprises, dynamically influence abnormal returns. We also plot the coefficients of the earnings surprise alone, γ, to compare as a baseline.

We provide this plot for 3 metrics, chosen for their relevance to our hypotheses. The variables are the average NLP sentiment score, the interquartile range of the NLP sentiment score, and the percentage of negative posts.

Our second method is to run univariate regressions of CAR on individual metrics are to identify standalone effects. We then use multivariate models then evaluate joint significance and control for literature variables, allowing us to compare our novel sentiment measures to those grounded in prior literature.

Our univariate model (Model 2) takes the following form:

$$CAR_{j,t,d} = \alpha + \beta_1 Regressor_{j,t} + \epsilon_{j,t,d} \tag{2}$$

$$d \in [-20,-1],[-10,-1],[-5,-1],[-1,1],[1,5],[1,10],[1,20]$$

where *Regressor* represents one of our predefined variables, the subscript j is the company, t is the date of the earnings call, d is the period under study measured relative to the day of the call, and ϵ is the error term.

Alongside simple OLS, we employ firm-level Fixed Effects (FE) regressions to control for unobserved, time-invariant characteristics specific to each company that influence PEAD such as management quality, business model, or sector-specific norms (Fink, 2021). These firm-specific traits may confound the relationship between investor sentiment and PEAD if left uncontrolled. By using firm fixed effects, we isolate the variation within firms over time, allowing us to better identify the impact of changes in sentiment on changes in abnormal returns following earnings announcements.

We have a relatively small sample of companies and observations, with some firms having as few as one observation. This restricts our ability to estimate fixed effects reliably, as the model requires sufficient within-firm variation to produce meaningful results. To address this, we reduce our sample to companies with at least 10 observations, ensuring that firm-specific effects can be estimated robustly.

We do not employ time fixed effects due to the small number of observations per company even with this filtering. Time fixed effects risk absorbing meaningful variation related to sentiment, earnings, or pre-event CAR. Product releases or continuous company improvements may influence both sentiment and earnings outcomes and with a small sample, these real effects could be mistakenly explained away by time fixed effects. This additional control could be utilised to greater effect with an even more restricted dataset, an exercise we omit for future study.

We conduct multivariate regressions of CAR in all periods using 6 distinct model formulations to evaluate joint significance and introduce control variables.

The first formulation will regress each periods' CAR on all available variables, to investigate the significance and magnitude of all regressors and mitigate against suppression effects. For instance, a mean sentiment score might not be informative without considering the range. Further, our two scoring methodologies might capture different aspects of sentiment, as NLP dynamically learns firm specific language and conditional statements. In contrast, dictionary scoring is rigid and

might better capture technical information, as the lexicon was specifically created for financial terminology.

To avoid perfect collinearity from our dummy variables we define the base case as a positive earnings surprise, and regress only on the negative dummy. We refer to this large model as Model 3 and construct it as follows:

$$\begin{aligned} CAR_{j,t,d} = \beta_1 ESD_{j,t} + \beta_2 PES_{j,t} + \beta_3 DIR_{j,t} + \beta_4 TP_{j,t} + \beta_5 LTP_{j,t} \\ + \beta_6 PEAR_{j,t}^{(5)} + \beta_7 PEAR_{j,t}^{(10)} + \beta_8 PEAR_{j,t}^{(20)} \\ + \beta_9 ASD_{j,t} + \beta_{10} ASN_{j,t} + \beta_{11} PNP_{j,t} \\ + \beta_{12} IQRD_{j,t} + \beta_{13} IQRN_{j,t} + \epsilon_{j,t} \end{aligned} \tag{3}$$

for k is in the pre event windows $K \in [-20,-1],[-10,-1],[-5,-1]$ and $PEAR_{j,t,k}$ applies only when d is in the event or post event window. The subscript j is the company, t is the date of the earnings call, and d is the period under study measured relative to the day of the call.

To mitigate against multicollinearity between sentiment variables and evaluate their independent informational value more robustly we also conduct multivariate regressions using only one sentiment scoring method at a time, alongside literature-based variables. We refer to these constructions as Model 4 and construct them as follows:

NLP Metrics:

$$\begin{aligned} CAR_{j,t,d} = \beta_1 ESD_{j,t} + \beta_2 PES_{j,t} + \beta_3 DIR_{j,t} + \beta_4 TP_{j,t} + \beta_5 LTP_{j,t} \\ + \beta_6 PEAR_{j,t}^{(5)} + \beta_7 PEAR_{j,t}^{(10)} + \beta_8 PEAR_{j,t}^{(20)} \\ + \beta_9 ASN_{j,t} + \beta_{10} IQRN_{j,t} + \epsilon_{j,t} \end{aligned}$$

Dictionary Metrics:

$$\begin{aligned} CAR_{j,t,d} = \beta_1 ESD_{j,t} + \beta_2 PES_{j,t} + \beta_3 DIR_{j,t} + \beta_4 TP_{j,t} + \beta_5 LTP_{j,t} \\ + \beta_6 PEAR_{j,t}^{(5)} + \beta_7 PEAR_{j,t}^{(10)} + \beta_8 PEAR_{j,t}^{(20)} \\ + \beta_9 ASD_{j,t} + \beta_{10} IQRD_{j,t} + \epsilon_{j,t} \end{aligned} \tag{4}$$

Negative Posts:

$$\begin{aligned} CAR_{j,t,d} = \beta_1 ESD_{j,t} + \beta_2 PES_{j,t} + \beta_3 DIR_{j,t} + \beta_4 TP_{j,t} + \beta_5 LTP_{j,t} \\ + \beta_6 PEAR_{j,t}^{(5)} + \beta_7 PEAR_{j,t}^{(10)} + \beta_8 PEAR_{j,t}^{(20)} \\ + \beta_9 PNP_{j,t} + \epsilon_{j,t} \end{aligned}$$

Where the subscript j is the company, t is the date of the earnings call, d is the period under study measured relative to the day of the call, and ϵ is the error term.

Additionally, we replicate the model from Di Wu (2019) using only established variables to benchmark their explanatory power. These regressions are referred to as Model 5 and takes the following form:

$$\begin{aligned} CAR_{j,t,d} = \beta_1 ESD_{j,t} + \beta_2 PES_{j,t} + \beta_3 DIR_{j,t} + \beta_4 TP_{j,t} + \beta_5 LTP_{j,t} \\ + \beta_6 PEAR_{j,t}^{(5)} + \beta_7 PEAR_{j,t}^{(10)} + \beta_8 PEAR_{j,t}^{(20)} + \epsilon_{j,t} \end{aligned} \tag{5}$$

Where the subscript j is the company, t is the date of the earnings call, d is the period under study measured relative to the day of the call, and ϵ is the error term.

Finally, we will conduct a multivariate regression on only our new sentiment variables to assess how they perform as a standalone methodology and compare this performance to that of the literature variables. This formulation is called Model 6 and is constructed:

$$\begin{aligned} CAR_{j,t,d} = \beta_1 ASD_{j,t} + \beta_2 ASN_{j,t} + \beta_3 PNP_{j,t} \\ + \beta_4 IQRD_{j,t} + \beta_5 IQRN_{j,t} + \epsilon_{j,t} \end{aligned} \tag{6}$$

Where the subscript j is the company, t is the date of the earnings call, d is the period under study measured relative to the day of the call, and ϵ is the error term.

To assess the potential multicollinearity introduced by overlapping CAR variables, we re-estimate all model specifications excluding these controls. The results remain broadly consistent, with only minor shifts in coefficient magnitudes and negligible changes in statistical significance. This robustness suggests that multicollinearity exerts limited influence on the estimation of sentiment effects and does not materially affect inference. Due to the constraints placed on this paper we include these results only when explicitly referred to.

4.1.1 Diagnostic Tests

To test for heteroskedasticity and serial correlation in our models we conducted a Breusch Pagan test and for serial correlation we conducted a Durbin Watson test to ensure consistent and unbiased estimates. The result of these tests revealed heteroskedasticity on a number of the regressions, but no statistically significant autocorrelation. We use clustered standard errors across estimates, thereby reducing the risk of falsely rejecting our null hypothesis.

5 Results

This section presents and interprets our empirical results for all event periods. We begin by testing whether sentiment impacts the earnings surprise effect to augment daily abnormal returns using event study coefficient plotting. We then assess the direct effect through univariate regressions on the cumulative abnormal return. Finally, we introduce controls through multivariate regressions that test for information when accounting for other variable under increasingly rigorous conditions.

5.1 Coefficient Plotting Results

To determine if sentiment augments the earnings surprise effect on daily abnormal returns we plot the coefficients specified by Model 1 across the event window in Figure 3 in blue, along with the baseline percentage earnings surprise coefficients in orange.

Our analysis reveals no evidence of statistically significant relationships in any of these regressors. The coefficients for the average NLP sentiment score, the interquartile range of NLP sentiment, and the percentage of negative posts have confidence intervals which include zero across the period. The coefficients and confidence intervals of the interaction term over the post event period are shown in Table 6, along with the summary of the model. As we can see there is very little explanatory power in this model. The un-interacted earnings surprise coefficients are not statistically different from zero, implying that the earnings surprise effect contains little information on daily abnormal returns in our dataset.

The lack of significant interaction effects across all three sentiment measures suggests that sentiment does not systematically amplify or attenuate the earnings surprise response meaningfully. Consequently, our analysis does not support any of our hypotheses, as no clear directional relationship emerges between sentiment and abnormal returns. Although minor fluctuations in the coefficients are observable, they are neither statistically nor economically significant. However, the absence of a daily effect does not rule out the possibility that sentiment influences investor behaviour over a longer horizon. We therefore shift focus to cumulative abnormal returns to investigate whether sentiment contains predictive information over the full event window.

5.2 Univariate Regression Results

To explore the direct predictive power of sentiment we estimate univariate regressions between each variable and cumulative abnormal returns across event windows. Table 7 presents univariate regression results for variables significant under Model 2 at the 10% level under both Ordinary Least Squares (OLS) and fixed effects (FE) methods. These results highlight several relationships of interest across the event, pre-event, and post-event windows. The notation for significance level is defined with ($p < x$) where x is the percentage significance level.

During the event window, the three pre-event CARs, and the interquartile range (IQR) of the NLP sentiment score show significance. The 5-, 10-, and 20-day pre-event CARs are all positively associated with event-day returns, with coefficients of 0.28, 0.18, and 0.11 respectively ($p < 0.01$), consistent with momentum effects reported in the literature. The significance and negative coefficient of the interquartile range of NLP sentiment (–6.25, $p < 0.01$) in the 5-day post-event window suggests that greater divergence in online sentiment around earnings events is associated with more negative abnormal returns, supporting the theoretical justification for SH_2. Surprisingly, the actual earnings result and its magnitude are not significant, further suggesting that traditional PEAD indicators may no longer carry persistent informational value. This could also stem from the sample bias toward strong performers, where positive earnings

surprises are expected and thus fail to elicit strong price reactions unless they significantly exceed expectations.

In the pre-event window, relatively few variables emerge as significant. The average NLP sentiment score during the event is positively associated with 5- and 10-day pre-event CARs, with coefficients of 1.00 ($p < 0.05$) and 1.79 ($p < 0.01$), respectively, under the FE model. This suggests that more positive sentiment often follows stronger prior returns. The percentage of negative posts is negatively associated with the 10-day pre-event CAR (–4.15, $p < 0.05$). One explanation could be anticipatory trading by informed market participants, where sentiment and CAR are both influenced by the forthcoming earnings result. However, an alternative explanation is that event period sentiment reflects, in part, past performance. Companies with weak/strong pre-event returns elicit more pessimistic/optimistic commentary on the day of the call, regardless of earnings results. This reverse relationship would invalidate the insider trading inference and is further supported by the lack of significant correlation between actual earnings results and pre-event CARs.

In the post-event window, we observe more consistent significance among the pre-event CAR variables. The 10- and 20-day pre-event CARs are positively associated with 5- and 10-day post-event CARs, consistent with literature on the persistence of abnormal returns following momentum (Fink, 2021). The negative coefficient on the 5-day pre-event CAR for the 20-day post-event window is unexpected and could result from profit-taking behaviour by traders. This explanation necessitates an average holding period between 10 and 20 days on profitable trades, which is not empirically verified.

Post-event regressions show that both total and log total posts are significantly and negatively associated with returns across models, in the 20-day post-event window, total posts have a coefficient of –0.025 ($p < 0.01$), while log total posts have a coefficient of –1.15 ($p < 0.01$). A potential explanation lies in sample bias, as our sample generally exceeds expectations, with an average earnings surprise of 5%. Heightened online commentary may reflect disappointment when firms underperform, increasing activity around negative surprises. The FE specification accounts for time-invariant firm characteristics mitigating larger firms from biased total posts and IQR metrics, as increased commentary likely increases the range of sentiment independent of other factors, strengthening the proposal that increased post volume is related to underperformance. The IQR of NLP sentiment remains significant and negative in both the 5- and 10-day post-event periods with coefficients of –6.25 and –7.20, respectively ($p < 0.01$), reinforcing its potential as a predictor of short-term price corrections and supporting SH_2.

In summary, the univariate results provide partial support for SH_2, the IQR of sentiment appears to carry persistent informational value. However, there is no supporting evidence for SH_1 or SH_3. For our overall research question these findings provide limited evidence that online sentiment, as constructed, contains meaningful incremental information beyond traditional indicators. We suspect sentiment is influenced by prior performance which raises issues in interpretation due to endogeneity. We address this in the following section by including pre-event CARs in multivariate models to better isolate the role of sentiment variables and identify any persistent informational content.

5.3 Multivariate Analysis

We turn to the multivariate regressions to assess the presence of incremental information in sentiment on PEAD beyond known predictors. This method allows us to evaluate what impact sentiment has when accounting for other effects.

The regression results for the full Model 3 of literature and novel variables using FE are found in Table 8. The OLS results for the same specification are found in Table 9. The restricted

regressions from Model 4 under OLS and FE are found in Table 12, Table 13, and Table 14. Finally, the strictly literature/sentiment variables of Model 5 and 6 are found in Table 11 and Table 10 respectively.

For conciseness, unless stated otherwise, all coefficient values referenced below are from the full Fixed Effects (FE) specification (Model 3), as it most effectively controls for unobserved firm-level heterogeneity. Coefficients from alternative specifications are discussed only where they meaningfully diverge from this baseline.

5.3.1 Pre-event periods

We test the extent to which sentiment reflects expectations formed before the earnings call through pre-event period regressions. Comparing the results of OLS and FE methods reveals similar results, although the FE method enhances the statistical significance of average sentiment metrics (Table 8 and Table 9). We will focus on the fixed effects results as stated prior.

In the pre-event periods, the average NLP sentiment score and average dictionary sentiment score are statistically significant at the 1% level in most windows, with effect sizes increasing over longer periods. In the 20-day pre-event window, the coefficients are 10.15 for NLP sentiment and 17.57 for dictionary sentiment, both significant at the 1% level. These consistent results suggest that event-day sentiment is partially shaped by prior performance. This introduces a potential issue violating the endogeneity condition, where sentiment reflects pre-event price movements rather than predict future ones. This is especially true of retail investors who more readily base their analysis on time series observations rather than analyst estimates (Battalio & Mendenhall, 2005). As a result, their apparent predictive power could be overstated. To mitigate this, our event and post-event regressions include pre-event CAR controls.

We investigate this relationship further through the restricted regressions in Table 12, Table 13, and Table 10.

Estimating Model 6 shows the statistical significance of these average measures remains robust across periods (Table 10). Conversely, when using sentiment measures from only one method as in Model 4 the average NLP sentiment score has diminished significance, while the average dictionary sentiment score loses significance entirely (Table 12 and Table 13).

This provides further evidence that sentiment scores are partially determined by pre-event CAR and suggests that the dictionary method's apparent relevance is largely due to correlation with other sentiment variables, rather than independent informational value. This weaker correlation with pre-event CAR indicates some advantage in terms of exogeneity.

Estimations using Model 4 show an increase in the significance the IQR variables (Table 12 and Table 13). The interquartile range of the dictionary score becomes statistically significant at the 10% level for all pre-event periods with a negative sign and coefficients increasing with the size of the window to a maximum of -4.86. The interquartile range of the NLP scores gains significance in only the 20-day pre-period. Positive pre-event CAR is associated with higher agreement among commenters, while negative CAR is linked to increased disagreement, reflecting a general positivity bias in online commentary.

The other variables significant under Model 3 are the percentage of negative posts in the 5 day pre period, the log of total posts in the 10-day pre period, and the interquartile range of the NLP score in the 20-day pre period, all at the 5% significance level. These findings are inconsistent across periods and models, indicating that these metrics do not necessarily reflect stable structural relationships with pre-event CAR.

The consistent correlation of average sentiment scores with prior returns suggests that sentiment is at least partly reactive. This motivates our decision to control for pre-event CARs in all

subsequent regressions to isolate sentiment's forward predictive power.

5.3.2 Event period

Event-day abnormal returns reflect investors' rapid incorporation of new information disclosed during earnings calls, alongside contemporaneous sentiment from social media discussions. We recognize that the sentiment metrics' magnitude and direction are largely driven by the earnings announcement and the immediate market response, and we expect severe multicollinearity. Inference about the impact of sentiment on contemporaneous CAR is therefore unclear due to the likely presence of feedback loops. Nevertheless, we can observe which regressors are significant and interpret some results.

In the full regression model, the cumulative abnormal returns of 5 and 20 days before the call are statistically significant at the 1 and 5% level respectively, though the 10-day pre-event CAR is not (Table 8). In the fixed effects model, this reflects demeaned abnormal returns and thus captures firm-specific deviations from typical pre-announcement drift. Part of the event-period return can plausibly be attributed to pre-earnings announcement drift or anticipatory trades from market participants with inside information which are present in the pre-event return as well, which would guarantee significance. The 5-day pre-event CAR coefficient of 0.18 implies that a 1% increase in pre-event abnormal returns is associated with a 0.18% increase in event-day CAR.

Among the literature-based variables, earnings difference in USD and log of total posts are significant at the 10% level. As expected, the earnings difference coefficient is positive (0.43), indicating that the direction of CAR follows the earnings surprise. The log of total posts has a negative coefficient of –0.24 ($p < 0.10$), suggesting that greater online activity on earnings day is associated with lower abnormal returns. This finding contradicts prior literature but is consistent with our earlier results, suggesting that unusually high volumes of online commentary are driven by earnings misses.

Among sentiment variables, the interquartile range of dictionary sentiment is significant at the 1% level (–2.94), and the average dictionary sentiment score is significant at the 5% level (–3.35). IQR has a negative coefficient, indicating that greater variation in sentiment on the day of the earnings call is correlated with more negative abnormal returns.

Interestingly, both the average dictionary and NLP sentiment scores have negative coefficients at -3.35 and -1.65 respectively, though the NLP measure is not significant. This suggests that sentiment, which is shaped contemporaneously, is often more positive than average even when returns are negative. These coefficients are likely influenced by multicollinearity and suppression effects caused by inclusion of pre-event CAR. These pre-event variables absorb much of the explanatory power for positive returns, leaving the sentiment metrics to capture residual negative variation. Both log of total posts and average dictionary sentiment score lose their statistical significance in restricted regressions that omit the pre-event CAR variables (Table 15) and the sign of the insignificant average NLP score turns positive.

It is difficult to identify the causal effects of sentiment on event period CAR or vice versa. The IQR of dictionary scoring appears correlated with negative reactions, indicating some contemporaneous dynamics. We move onward to the post event period without major conclusions on the relevance of sentiment to event period CAR.

5.3.3 Post Event Periods

Finally, we assess the most critical period for our hypotheses, the post-event periods when the Post-Earnings Announcement Drift actually occurs. This section evaluates which variables

continue to hold predictive value, how sentiment measures behave, and what these patterns reveal about our research question and hypotheses.

In the full regression model (Table 8), the log of total posts shows a consistently negative and increasingly significant relationship with post-event returns, with coefficients ranging from –0.41 ($p < 0.05$) in the 5-day window to –1.43 ($p < 0.01$) in the 20-day window. This consistent pattern is in line with our univariate findings which attribute increased post volume with underperformance and contain actionable information on PEAD. The total number of posts remains insignificant, which implies logarithmic scaling better captures the predictive relationship between post activity and returns. This finding is robust across models, retaining significance in all specifications using FE and losing significance substantially in simple OLS models.

The earnings surprise percentage only becomes significant in the 20-day post window, with a negative sign. This result could reflect a market correction after an initial overreaction or a reversal of PEAD. However, it lacks significance in all other model specifications, suggesting it contains little information for predicting PEAD.

The pre-period abnormal return variables are highly informative across all models and windows. The 5-day pre-event CAR is significant across all post-event windows, the 10-day CAR is significant in the 5- and 10-day post periods, and the 20-day CAR is significant only in the 5-day post period. The progressive drop-off implies that more recent performance is more relevant for shaping post-event expectations. This is consistent with investors reliance on shorter-term performance, and our univariate results.

All pre-event CAR coefficients are negative in all FE models, but are positive in OLS, except for 5-day pre-event CAR (-0.32, $p < 0.01$ in the 20 day period). This difference demonstrates that firms with dramatic pre-announcement drift also have large post-announcement drift. An explanation for the negative sign under FE is that the excess pre-event CAR, combined with the earnings announcement outcome leads to temporary over/undervaluation and triggers corrective trades. Alternatively, investors holding positions taken pre-event might close them, normalising prices. The correction effect observed in FE points to increased market efficiency in processing information and limiting overreaction.

The variables mentioned above are the only significant ones from the literature, the absence of any significant variables related to the actual earnings result demonstrates the reduction in PEAD and the extent to which markets have increased efficiency.

We can now address the sentiment metrics which are our novel contribution to the study of PEAD. We compare them across models to assess the presence of persistent information and check the signs and significance of variables to test our sub hypotheses.

Average NLP Sentiment is significant at the 1% level and negative across all post-event periods in the full model under fixed effects (Table 8). The magnitude increases over longer horizons, with the coefficient reaching -13.10 in the 20-day post-event window ($p < 0.01$). This implies that a 0.1 increase in sentiment above the mean corresponds to a 1.3% decline in post-event abnormal returns. Significance is maintained in the absence of literature regressors (Table 10) but lost without dictionary-based measures (Table 12), indicating multicollinearity with other sentiment measures in the combined specification.

Under OLS, the average NLP sentiment score reverses sign and loses some significance (Table 9), highlighting firm-level sentiment biases and reinforcing the importance of fixed effects in accurately capturing the sentiment–PEAD relationship.

The Average Dictionary Sentiment Score is significant at the 1% level and negative across all post-event windows in the full model (Table 8), implying more positive sentiment reduces subsequent abnormal returns. The magnitude of the coefficient grows from -5.76 in the 5 day

period to -14.46 in the 20 day. An abnormal increase in sentiment of 0.1 results in a mean decline in abnormal return of 1.4% over 20 days, even larger than the NLP method. When using Model 4 (Table 13), this significance weakens and in Model 6 (Table 10), the 5- and 10-day effects are no longer significant, indicating dictionary sentiment's predictive value may be more context-dependent than that of NLP measures. Under OLS, coefficients remain negative but are smaller and less robust, except in Model 4 (Table 13). This likely reflects both multicollinearity with other sentiment variables and numerical influence of a narrower distribution of scores. Average dictionary sentiment captures post-event dynamics less consistently than NLP-based measures and again highlights the importance of controlling for firm-level biases.

These results strongly support SH_1, suggesting that higher average sentiment on the day of an earnings call attenuates PEAD. While significance is not uniform, the consistent results across several models and periods indicate that average sentiment contains informational value. We reject the null hypothesis that average sentiment holds no predictive information; it does have predictive power. Our initial hypothesis that this effect results from an outsized market reaction during the event is not supported by event-period results.

Alternative explanations for the significance of average sentiment measures rely on the interaction between emotional investor behaviour and the classic explanation of PEAD as market underreaction. Strong sentiment may prompt retail traders to enter options/futures positions during the event which must be fulfilled later. Another possibility is that upon seeing outsized sentiment during the event window institutional investors take up positions to capitalise on any trading which moves the price in the opposite direction after the call and diminishes PEAD. Alternatively, extreme sentiment could discourage entry, as investors believe they have missed the opportunity. Further study is needed to establish likely causes.

Coefficients on the IQR of NLP Sentiment are negative and significant at the 1% level across all FE models at 5 and 10-day horizons with values of -6.55 and -7.17 respectively. This is consistent with our univariate results and the hypothesis that disagreement in sentiment predicts price corrections. The less significant 20-day horizon implies diminishing predictive power over time. The OLS results are mixed, with weaker significance and a single positive coefficient in Model 4 (Table 12) which may point again to recurring multicollinearity.

The interquartile range of the dictionary sentiment scores were generally not significant, highlighting the difference between methods which were relatively similar in the significance of their mean scores.

These results provide support for SH_2, as we reject the null hypothesis that event-period sentiment range lacks predictive power for PEAD. The interquartile range (IQR) of the NLP sentiment score is consistently significant across specifications, with greater disagreement linked to more negative post-event returns. Poor earnings performances prompt more divided online reactions and greater price declines, the metrics provide more predictive power than the earnings metrics themselves and could provide arbitrage opportunities.

The percentage of negative posts shows increasing negative significance over time in Model 3 (Table 8), with coefficients increasing in magnitude from –7.52 ($p < 0.10$) in the 5-day period to –19.43 ($p < 0.01$) in the 20-day window. The result is similar under Model 6 (Table 10) but becomes insignificant under Model 4 (Table 14), indicating this variable captures unique sentiment dimensions only in conjunction with other metrics. An increase in the incidence of negative posts of 10% correlates with a 1.9% decrease in abnormal returns over 20 days. This supports the idea that negative sentiment exerts downward pressure on prices, offering some predictive information. Under OLS, the sign reverses and becomes significantly positive, highlighting firm-level sentiment bias once again (Table 9).

These results support SH_3, that increases in negative sentiment during the event period predict

more negative PEAD. The consistent sign, significance, and growing magnitude of coefficients across FE models suggest that online negativity holds predictive power. Unusually pessimistic investor sentiment can outperform traditional indicators like earnings surprise in forecasting the direction and magnitude of PEAD. We reject the null hypothesis that the percentage of negative posts holds no information.

The results from the post-event windows allow us to answer our research question, "Does the sentiment of online investors hold information on PEAD?" Given the consistent significance of various sentiment variables after the event, and their nontrivial magnitude, we reject the null hypothesis that sentiment variables hold no information on PEAD.

Our findings suggest that sentiment contains information not immediately internalised by the market during the earnings call. The lack of significance in traditional variables, paired with the significance of sentiment-based predictors post earnings call, implies some market inefficiency.

Average sentiment metrics consistently predict negative returns under fixed effects models, supporting an overreaction-based interpretation of PEAD. The significance of sentiment range metrics indicates that the degree of disagreement between investors holds information on PEAD. The percentage of negative sentiment posts also demonstrates strong predictive power towards negative movements. In unison these findings strongly imply that online investor sentiment has predictive power on PEAD.

These observations strongly confirm our sub hypotheses, with average sentiment (SH_1), sentiment range (SH_2), and negative sentiment (SH_3) each finding degrees of support. The robust negative relationship found in most sentiment measures post-event, especially under fixed effects specifications, strongly suggests sentiment's role in prompting subsequent corrections or reversals as investors incorporate emotional or speculative biases back into market equilibrium.

The confirmation of our hypotheses implies there is an opportunity for arbitrage utilising these more advanced data collection and processing techniques, increasing market efficiency as sentiment data offers actionable insights beyond traditional financial indicators. As such, our findings support the growing relevance of alternative data in asset pricing and investment strategies. Over time, the use of sentiment-based trading signals may contribute to reducing PEAD as markets become more adept at immediately incorporating this previously underutilised information. While unconfirmed, this is likely already the case as StockTwits provide enhanced access to their content and in house sentiment tracking for a large fee which is likely aimed at technical investors and proprietary trading firms.

5.3.4 Limitations

This study has several limitations that future research could address. The dataset is relatively small with 1,172 observations, and is biased toward large, successful, and widely followed firms from the S&P 500 which influenced the earnings metrics. There is also a strong skew in social media activity across firms, with the mean number of posts per event at 47.7, a median of 30, and the bottom quartile at just 18 posts which limited accuracy in sentiment metrics for unpopular companies. Methodologically, multicollinearity posed challenges due to overlap in earnings surprise metrics, the inclusion of both total and log total posts, and sentiment variables capturing similar dynamics through different methods. Although this was partially addressed through restricted model specifications, it remains a constraint. Furthermore, the dictionary-based sentiment measure may suffer from lower accuracy, and while the NLP method shows improvement over traditional techniques, there is still room for enhancement in sentiment extraction and interpretation. Finally, we restricted our window to 20 days pre and post event. After observing increased significance over longer periods of NLP measures this choice may have limited our findings and could be extended in the future.

Future studies could expand the dataset beyond the S&P 500 to include smaller firms which historically display more PEAD. Enhancing sentiment analysis through more advanced NLP techniques or real-time data could refine measurement. Additionally, testing trading strategies based on sentiment signals could assess their true potential.

6 Summary and conclusion

This study set out to answer a clear research question: Does online investor sentiment contain predictive information about the Post-Earnings Announcement Drift (PEAD)? Motivated by the decline in PEAD and the explanatory power its predictors, and the rising influence of social media in financial markets. We examined sentiment derived from StockTwits, using advanced machine learning techniques, to test a new data source for predictive power.

Our findings contribute to the literature by providing clear and consistent evidence that measures of sentiment on the day of earnings announcements contain statistically significant predictive information about PEAD for the first time. We conclude that sentiment is not fully incorporated into prices at the time of the earnings call and provides an exploitable inefficiency. In contrast, the traditional predictors of PEAD did not contain statistically significant information. Only pre-event abnormal returns and measures of investor attention retained consistent significance and imply a historical increase in market efficiency by pricing in these factors. The accuracy of our sentiment scoring was well in excess of similar literature, contributing further to the field of sentiment analysis within financial economics.

Our findings on the effect of increased online attention on PEAD stand in direct contrast to those of Di Wu (2019), who finds that heightened attention tends to reinforce post-earnings drift. We attribute this divergence not to a contradiction of Wu's conclusions, but to structural differences in our sample. Specifically, our dataset is skewed toward successful, large-cap firms where earnings outperformance is the norm. In such cases, increased attention may reflect unmet expectations or investor disappointment despite positive surprises, reversing the expected relationship. As such, we interpret our results as context-dependent, not as a refutation of prior literature.

We tested three sub-hypotheses. First, we examined whether average sentiment influences the direction of PEAD, across multiple model specifications. We found that higher average sentiment is associated with diminished PEAD for at least 20 days. This is consistent with a market correction interpretation that strong sentiment responses reduce the room for continued drift and support SH_1. The average online sentiment on the day of an earnings call holds predictive information on PEAD in excess of traditional predictors.

Second, we tested whether the dispersion in sentiment, as captured by the interquartile range, carries information on PEAD. The results showed that larger dispersion is consistently correlated with more negative PEAD, suggesting that disagreement among online investors can be used as a predictor, offering broad support for SH_2. Dictionary and NLP measures were significant in different models and further study could improve the effectiveness of this measure.

Third, we assessed whether the percentage of negative posts on the day of an earnings call predicts PEAD. This measure had minimal standalone power but became significant in combination with other sentiment metrics. We interpret this as supporting SH_3, though the effect is dependent on context and interaction with other sentiment measures.

Our findings show that online investor sentiment contains persistent predictive information on PEAD, in excess of the traditional predictors of this market inefficiency, over at least a 20 day window. This supports the broader view that behavioural signals derived from alternative data can help explain persistent deviations from the efficient market hypothesis and are important in improving informational efficiency in financial markets.

7 References

Ball, R., & Brown, P. (1968). An empirical evaluation of accounting income numbers. Journal of Accounting Research, 6(2), 159. https://doi.org/10.2307/2490232

Barber, B. M., & Odean, T. (2000). Trading is hazardous to your wealth: The common stock investment performance of individual investors. The Journal of Finance, 55(2), 773–806.

Bernard, V. L., & Thomas, J. K. (1989). Post-earnings-announcement drift: Delayed price response or risk premium? Journal of Accounting Research, 27, 1–34. https://doi.org/10.2307/2491062

Bollen, J., Mao, H., & Zeng, X. (2011). Twitter mood predicts the stock market. Journal of Computational Science, 2(1), 1–8. https://doi.org/10.1016/j.jocs.2010.12.007

Bradshaw, M. T., Drake, M. S., Myers, J. N., & Myers, L. A. (2012). A re-examination of analysts' superiority over time-series forecasts of annual earnings. Review of Accounting Studies, 17(4), 944–968. https://doi.org/10.1007/s11142-012-9185-8

Chan, L. K., Jegadeesh, N., & Lakonishok, J. (1996). Momentum strategies. The Journal of Finance, 51(5), 1681. https://doi.org/10.2307/2329534

Chang, A. (Chun-Chia), Yu, S. (Carol), Reinstein, A., & Churyk, N. T. (2016). An overview of investor sentiment in stock market. Journal of Contemporary Business Issues, 22(1), 1–14. https://www.wiu.edu/cbt/jcbi/documents/NAASFeb2016/SpecialNAASIssueFeb2016-InvestorSentiment.pdf

Chi, S. S., & Shanthikumar, D. M. (2016). Local bias in google search and the market response around earnings announcements. The Accounting Review, 92(4), 115–143. https://doi.org/10.2308/accr-51632

Chordia, T., Roll, R., & Subrahmanyam, A. (2009). Does information risk matter? Journal of Finance, 63(1), 1–34.

Daniel, K., Hirshleifer, D., & Sun, L. (2020). Short-and long-horizon behavioral factors. The Review of Financial Studies, 33(4), 1673–1736.

Danieli, L., & Denenzis, T. (2024). Social media sentiment: Influence on EU equity prices. https://www.esma.europa.eu/sites/default/files/2024-04/ESMA50-524821-3157_Risk_Article_Social_Media_sentiment_influence_on_EU_equity_prices.pdf.

Davidson, B., Wischerath, D., Racek, D., Parry, D. A., Godwin, E., Hinds, J., Linden, D. van der, Roscoe, J. F., Ayravainen, L., & Cork, A. (2023). Platform-controlled social media APIs threaten open science. Nature Human Behaviour. https://doi.org/10.1038/s41562-023-01750-2

Divernois, M. A., & Filipović, D. (2024). StockTwits classified sentiment and stock returns. Digital Finance, 6, 249–281. https://doi.org/10.1007/s42521-023-00102-z

DoltHub, post. (2024). Stocks. https://www.dolthub.com/repositories/post-no-preference/stocks

Earnings expectations, investor trade size, and anomalous returns around earnings announcements. (2005). Journal of Financial Economics, 77(2), 289–319. https://doi.org/https://doi.org/10.1016/j.jfineco.2004.08.002

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

Fink, J. (2020). A review of the post-earnings-announcement drift. Journal of Behavioral and Experimental Finance.

Friedman, M., & Zeng, Y. (2022). Retail investors and earnings announcement drift. Journal of Financial Economics.

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

JPMorgan Chase Institute. (2022). Retail investors and market volatility.

Ke, B., & Petroni, K. (2004). How informed are actively trading institutional investors? Evidence

from their trading behavior before a break in a string of consecutive earnings increases. Journal of Accounting Research, 42(5), 895–927. https://doi.org/10.1111/j.1475-679X.2004.00160.x

Li, Y., Nekrasov, A., & Teoh, S. H. (2020). Opportunity knocks but once: Delayed disclosure of financial items in earnings announcements and neglect of earnings news. Review of Accounting Studies, 25, 159–200.

Louis, H., & Sun, A. X. (2011). Earnings management and the post earnings announcement drift. Financial Management, 40(3), 591–621. https://doi.org/10.1111/j.1755-053x.2011.01154.x

MacKinlay, A. C. (1997). Event studies in economics and finance. Journal of Economic Literature, 35(1), 13–39. http://www.jstor.org/stable/2729691

Malkiel, B. G. (2003). The efficient market hypothesis and its critics. Journal of Economic Perspectives, 17(1), 59–82.

Martineau, C. (2021). Rest in peace post-earnings announcement drift. Critical Finance Review. https://doi.org/10.2139/ssrn.3111607

Nguyen, T. H., Shirai, K., & Velcin, J. (2015). Sentiment analysis on social media for stock movement prediction. Expert Systems with Applications, 42(24), 9603–9611. https://doi.org/10.1016/j.eswa.2015.07.052

Pfarrer, M. D., Pollock, T. G., & Rindova, V. P. (2010). A tale of two assets: The effects of firm reputation and celebrity on earnings surprises and investors' reactions. Academy of Management Journal, 53(5), 1131–1152. https://doi.org/10.5465/amj.2010.54533222

Post earnings announcement drift: A simple earnings surprise measure, the medium effect of investor attention and investing strategy. (2024). International Review of Financial Analysis, 95, 103460. https://doi.org/https://doi.org/10.1016/j.irfa.2024.103460

PwC. (2024). Earnings with a twist: 2024 update on SEC staff non-GAAP comment trends. https://viewpoint.pwc.com/dt/us/en/pwc/in_depths/2024/id2024/Iid202408.html.

Renault, T. (2017). Intraday online investor sentiment and return patterns in the u.s. Stock market. Journal of Banking & Finance, 84, 25–40. https://doi.org/10.1016/j.jbankfin.2017.07.002

Renault, T. (2020). Sentiment analysis and machine learning in finance: A comparison of methods and models on one million messages. Digital Finance, 2, 1–13. https://doi.org/10.1007/s42521-019-00014-x

So, E. (2022). Earnings calls and retail investor losses [Working Paper]. Harvard Business School.

Son, D. H., Palmon, D., & Yezegel, A. (2018). The persistence of firm-specific post-earnings announcement returns. Investment Analysts Journal, 47(1), 31–47. https://doi.org/10.1080/10293523.2017.1413151

StockTwits. (n.d.). StockTwits. https://stocktwits.com/

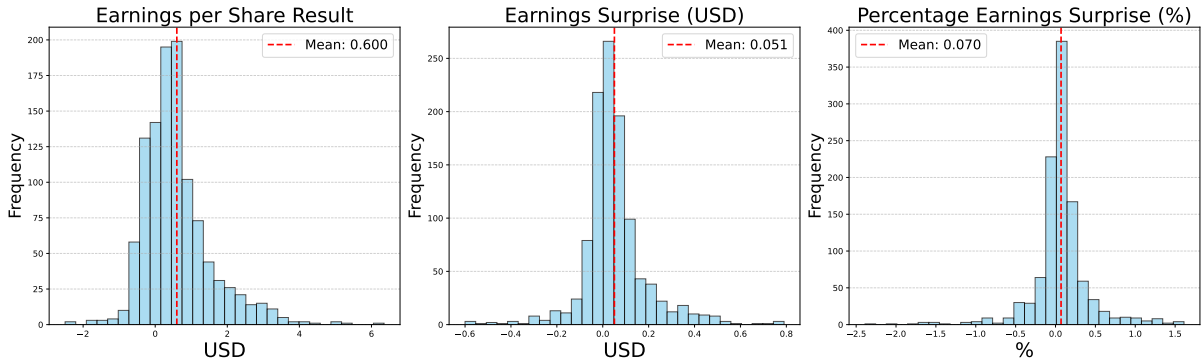
Tan, G. (2021). Social-media platform stocktwits nabs $210 million valuation. https://www.bloomberg.com/news/articles/2021-12-16/social-media-platform-stocktwits-nabs-210-million-valuation.

Wu, D. (2019). Does social media get your attention? Journal of Behavioral Finance, 20(2), 213–226. https://doi.org/10.1080/15427560.2018.1505729

Yalçin, K. C. (2010). Market rationality: Efficient market hypothesis versus market anomalies. European Journal of Economic and Political Studies, 3(2), 23–37. https://arastirmax.com/en/system/files/dergiler/25821/makaleler/3/2/arastirmax-market-rationality-efficient-market-hypothesis-versus-market-anomalies.pdf

8 Figures, Tables, and Diagrams

Earnings Result Distributions



CAR Distributions in all event windows

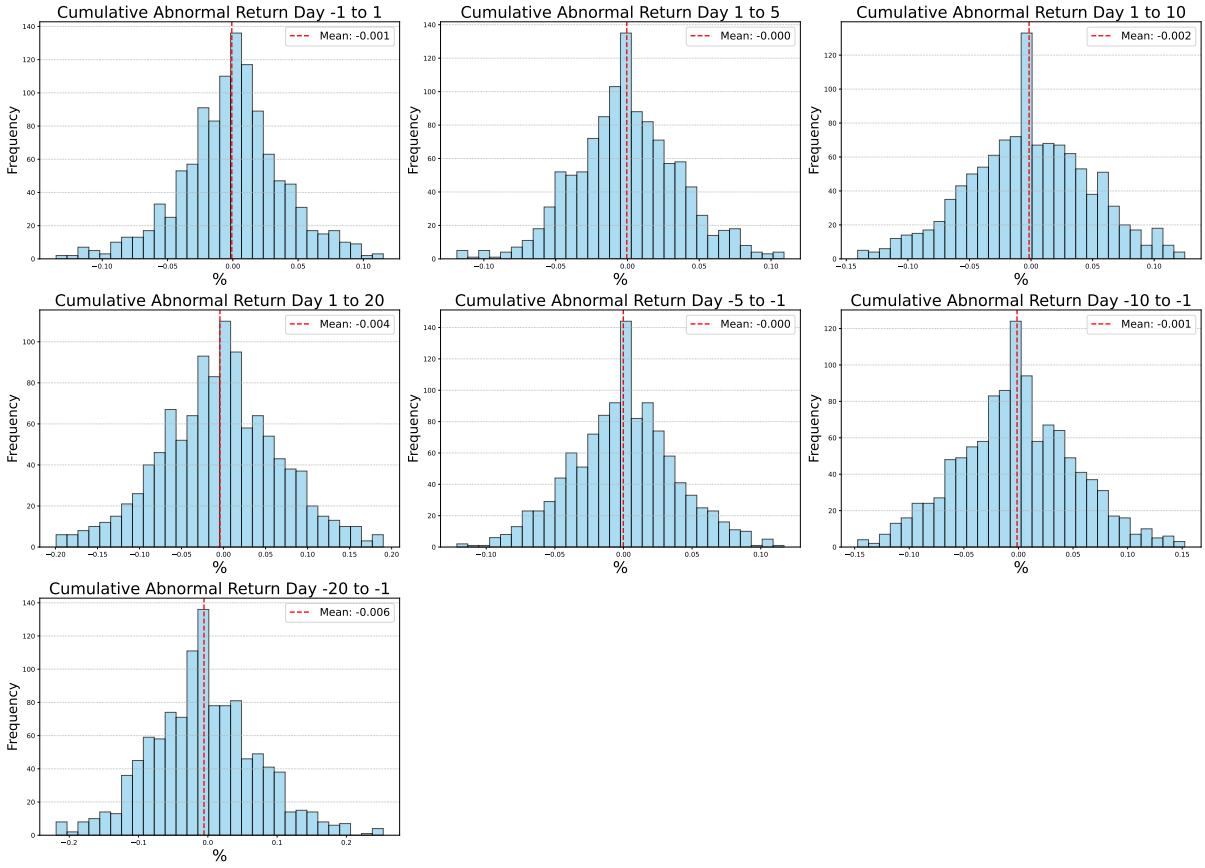


Fig. 1: Histograms of financial variables used in the study with 30 bars. Each subplot shows the distribution of a different key variable across the sample, including earnings measures, and Cumulative Abnormal Returns over all event windows measured from the day of the event (Day 0). The x axis is labelled with its measurement, either USD or percentage, and the y axis is the frequency with which observations have values in the range define by the bar. Vertical dashed lines indicate sample means and their values are given in each plot.

Tab. 1: Summary statistics of all variables used in the study including earnings metrics, cumulative abnormal returns (CAR) over all windows used measured from the event day, and all social media metrics with their number of observations, mean, standard deviation (std), minimum, all 3 quartiles, and maximum reported.

	count	mean	std	min	25%	50%	75%	max
Earnings per Share (USD)	1240	0.583	1.411	-10.88	-0.013	0.37	0.91	10.04
Earnings Surprise (USD)	1174	0.044	0.382	-7.68	-0.01	0.03	0.1	4.65
Percentage Earnings Surprise (%)	1172	0.053	2.208	-33.391	-0.023	0.048	0.179	58.5
CAR Day - 20 to -1	1304	0.042	0.572	-1.617	-0.059	-0.006	0.044	11.418
CAR Day -10 to -1	1304	0.034	0.516	-1.195	-0.039	0	0.036	11.299
CAR Day -5 to -1	1304	0.023	0.425	-0.373	-0.026	0	0.024	11.352
CAR Day -1 to 1	1303	0.001	0.085	-0.786	-0.022	0.001	0.022	1.937
CAR Day 1 to 5	1304	-0.003	0.053	-0.859	-0.026	0	0.023	0.186
CAR Day 1 to 10	1304	-0.002	0.078	-0.817	-0.036	-0.001	0.034	0.883
CAR Day 1 to 20	1304	-0.002	0.11	-0.826	-0.056	-0.002	0.045	1.038
Total Posts	1304	47.705	99.072	1	18	30	47	1771
Log of Total Posts	1304	3.288	1.056	0	2.89	3.401	3.85	7.479
Average NLP Sentiment Score	1296	0.697	0.249	-0.835	0.629	0.775	0.867	1
IQR of NLP Sentiment	1304	0.172	0.069	0	0.125	0.168	0.219	0.391
Average Dictionary Score	1296	0.146	0.084	-0.296	0.108	0.156	0.195	0.585
IQR of Dictionary Sentiment	1296	0.183	0.081	0	0.134	0.179	0.225	0.674
Percentage of Negative Posts	1304	0.112	0.127	0	0.022	0.078	0.154	1
Percentage of Neutral Posts	1304	0	0.002	0	0	0	0	0.048

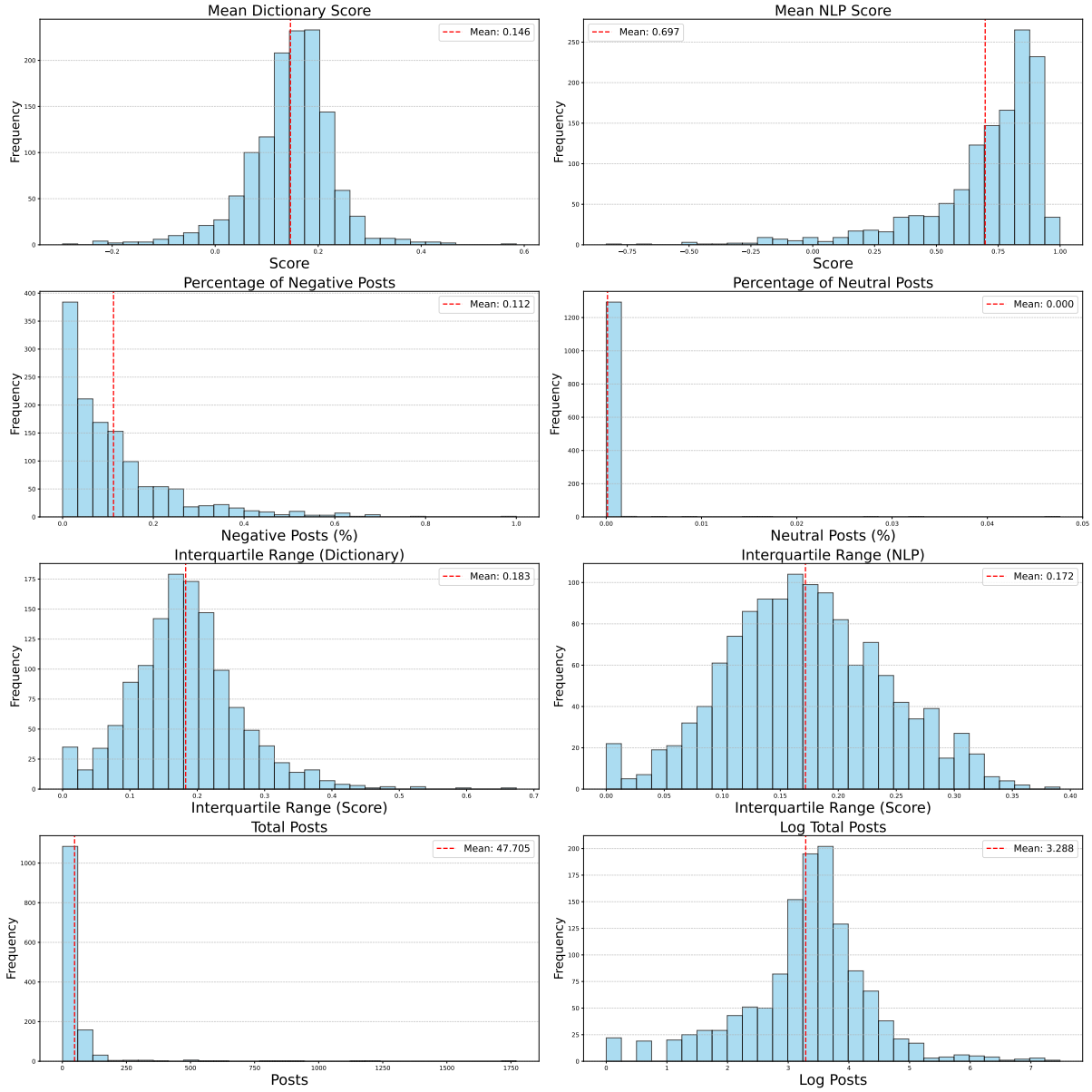


Fig. 2: Histograms of the social media variables on the earnings call date constructed for the study with 30 bars. Each subplot shows the distribution of a variable including the mean sentiment and NLP score using NLP and dictionary methods, the percentage of negative and neutral posts, and the number of total posts and the log of total posts. The x axis is the value of the metric and the y axis is the frequency with which observations are within the range defined by the bar. Vertical dashed lines indicate sample means and their values are given in each plot.

Tab. 3: Content relevant to sentiment scoring. The first item is a sample text before and after pre processing. The second is the machine learning classification results for a negative weighting of 1 using an 80:20 train/test split.The third is the machine learning classification results for a negative weighting of 2.3 using an 80:20 train/test split which was used to create the NLP scores.

Original Content	Date	Transformed Content	Date
$c market manipulation move today. Pulled the trigger on Goldman at $538 when I started to see them moving down. Looks like small caps are breaking out.	Oct 15, 2024 5:42 PM	cashtag market manipulation move today. pulled the trigger on goldman at $numbertag when i started to see them moving down. looks like small caps are breaking out.	2024-10-15 17:42:00
$c on sale today. 👍 bullish	Oct 15, 2024 5:38 PM	cashtag on sale today. emojipos bullish	2024-10-15 17:38:00
$c anyone down to start a class action against Wall Street for market manipulation? This should not have happened today.	Oct 15, 2024 5:32 PM	cashtag anyone down to start a class action against wall street for market manipulation? this should not negtag_have negtag_happened negtag_today.	2024-10-15 17:32:00
Bought $PTON, $C and added $JD and $BIDU.	Oct 15, 2024 5:27 PM	bought cashtag, cashtag and added cashtag and cashtag.	2024-10-15 17:27:00
$c sold at open for around breakeven. Dud of a day, like I said in a previous post, a lot was baked in last Friday. **Bearish**	Oct 15, 2024 5:23 PM	cashtag sold at open for around breakeven. dud of a day, like i said in a previous post, a lot was baked in last friday. **bearish**	2024-10-15 17:23:00

Classification Report for Negative Weight 1

Class	Precision	Recall	F1-Score	Support
-1	0.82	0.54	0.65	6,381
1	0.88	0.97	0.92	21,329
Accuracy			0.87	27,710
Macro Avg	0.85	0.75	0.79	27,710
Weighted Avg	0.86	0.87	0.86	27,710

Classification Report for Negative Weight 2.3

Class	Precision	Recall	F1-Score	Support
-1	0.68	0.73	0.70	6,381
1	0.91	0.91	0.91	21,329
Accuracy			0.86	27,710
Macro Avg	0.80	0.81	0.81	27,710
Weighted Avg	0.86	0.86	0.86	27,710

Tab. 6: Overall Model Statistics, Coefficients, and Confidence Intervals for Post-Event Period for the Mean NLP Score, Interquartile Range of NLP Score, and the Percentage of Negative Posts for Coefficient Plotting. None of these coefficients are statistically significant from zero.

[htbp]

	Avg_NLP_Score_Agree	IQR_NLP_Score	perc_Negative_Agreements
Summary Statistics			
Observations	44,706	44,963	44,963
R-squared	0.003	0.003	0.003
Adj. R-squared	0.000	0.000	0.000
F-statistic	1.069	1.137	1.076
Prob (F)	0.288	0.147	0.272
Days Since Event	Coefficent and CI	Coefficent and CI	Coefficent and CI
1	−0.0009 [−0.083, 0.081]	−0.0738 [−0.623, 0.476]	−0.0134 [−0.434, 0.408]
2	0.0180 [−0.064, 0.100]	0.0128 [−0.537, 0.563]	−0.0849 [−0.506, 0.336]
3	−0.0345 [−0.116, 0.047]	0.3191 [−0.231, 0.869]	0.1348 [−0.286, 0.556]
4	0.0619 [−0.020, 0.144]	−0.3954 [−0.953, 0.162]	−0.4488 [−0.872, −0.026]
5	0.0034 [−0.079, 0.085]	−0.0665 [−0.617, 0.484]	0.1282 [−0.296, 0.552]
6	0.0009 [−0.081, 0.083]	0.0901 [−0.461, 0.641]	0.0270 [−0.394, 0.448]
7	0.0261 [−0.056, 0.108]	0.0688 [−0.491, 0.629]	−0.0652 [−0.490, 0.359]
8	−0.0033 [−0.085, 0.079]	0.1606 [−0.389, 0.710]	0.0095 [−0.412, 0.431]
9	−0.0025 [−0.084, 0.079]	0.0987 [−0.452, 0.649]	−0.0159 [−0.437, 0.406]
10	−0.0579 [−0.140, 0.024]	0.5255 [−0.028, 1.079]	0.2394 [−0.182, 0.661]
11	0.0200 [−0.063, 0.103]	0.1498 [−0.401, 0.700]	−0.0901 [−0.516, 0.336]
12	−0.0356 [−0.121, 0.050]	0.2587 [−0.299, 0.817]	0.2282 [−0.225, 0.681]
13	−0.0702 [−0.152, 0.012]	0.3871 [−0.164, 0.939]	0.3997 [−0.026, 0.825]
14	−0.0305 [−0.113, 0.051]	0.2730 [−0.278, 0.824]	0.1153 [−0.309, 0.539]
15	0.0076 [−0.074, 0.090]	−0.1792 [−0.730, 0.372]	0.0256 [−0.396, 0.447]
16	0.0569 [−0.026, 0.139]	−0.4080 [−0.970, 0.154]	−0.2593 [−0.684, 0.166]
17	−0.0230 [−0.105, 0.059]	0.0467 [−0.503, 0.597]	0.1053 [−0.316, 0.527]
18	−0.0047 [−0.087, 0.077]	0.1722 [−0.378, 0.723]	−0.0151 [−0.437, 0.407]
19	−0.0109 [−0.093, 0.071]	−0.0425 [−0.593, 0.508]	0.0320 [−0.390, 0.454]
20	−0.0036 [−0.086, 0.078]	0.0634 [−0.487, 0.614]	−0.0142 [−0.436, 0.408]

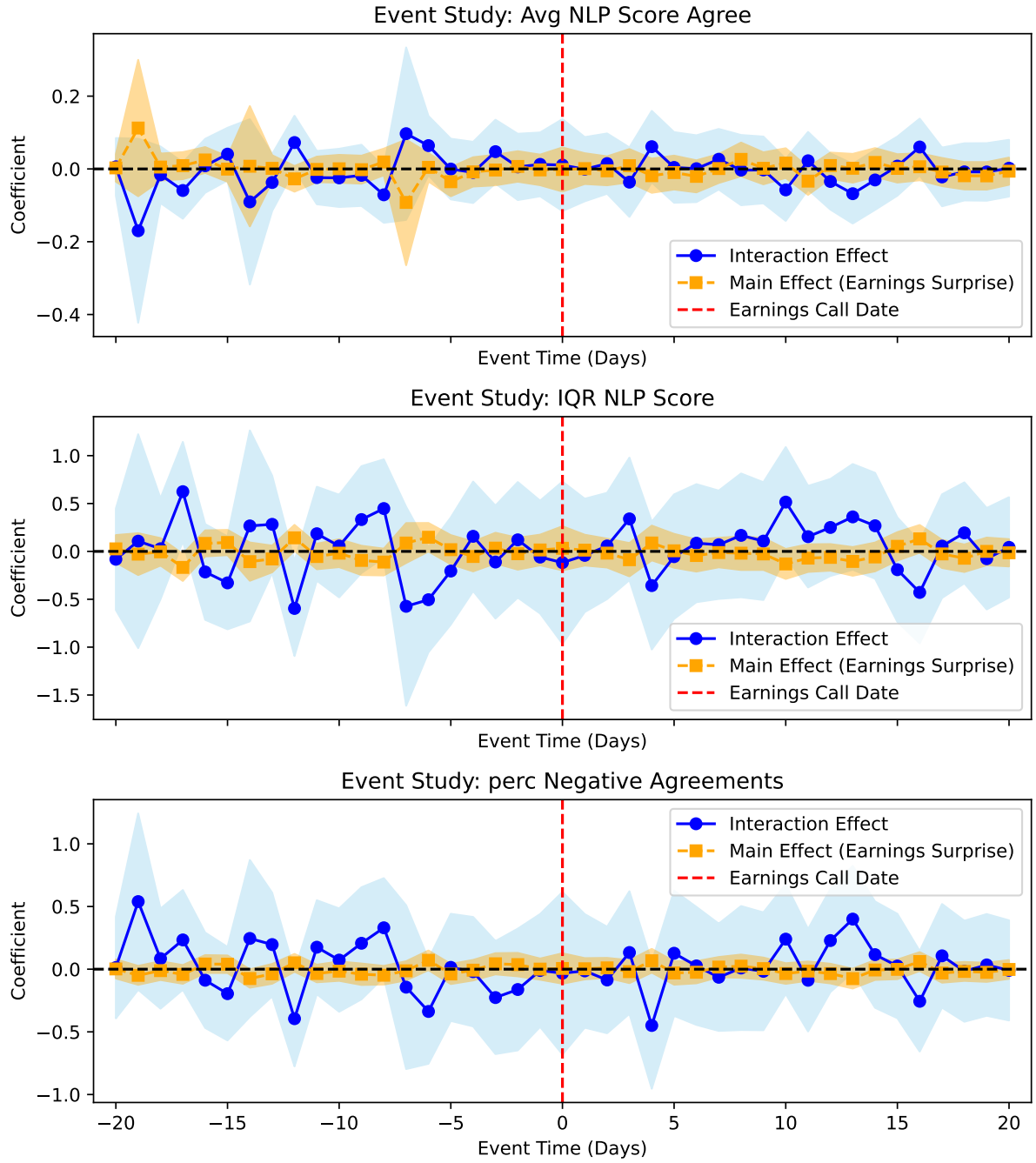


Fig. 3: Coefficient plots of the estimated percentage earnings surprise (orange) and interaction effects between earnings surprise and three sentiment metrics (blue): the average NLP sentiment score, the interquartile range of NLP sentiment, and the percentage of negative posts. These interactions are estimated from regressions of daily abnormal returns over a 20-day window before and after earnings calls. The shaded areas represent the 95% confidence interval, the x-axis shows event time, and the y-axis indicates the value of the estimated interaction coefficient.

Tab. 7: Univariate regression results significant at at least the 10% level sorted by event window and if the regression used simple OLS or Fixed Effects. Robust entity clustered standard errors were used for all regressions. *** indicates significance at the 1%, ** indicates significance at the 5% level, and * indicates significance at the 10% level.

Independent Variable	Coefficient	StdErr	t_value	p_value	Significance
Method: FE, Period: Post 20					
Total Posts	-0.0245	0.0065	-3.7873	0.0002	***
Log of Total Posts	-1.1468	0.2203	-5.2052	0.0000	***
CAR 5 Days Pre Event	-0.2204	0.0601	-3.6639	0.0003	***
Method: FE, Period: Post 10					
IQR of NLP Score	-7.1966	2.5320	-2.8423	0.0046	***
Total Posts	-0.0145	0.0050	-2.8691	0.0042	***
Log of Total Posts	-0.8090	0.1718	-4.7100	0.0000	***
CAR 10 Days Pre Event	0.1312	0.0325	4.0371	0.0001	***
CAR 20 Days Pre Event	0.0766	0.0220	3.4836	0.0005	***
Method: FE, Period: Post 5					
IQR of NLP Score	-6.2482	1.7438	-3.5830	0.0004	***
Log of Total Posts	-0.3354	0.1193	-2.8105	0.0050	***
CAR 10 Days Pre Event	0.0801	0.0225	3.5656	0.0004	***
CAR 20 Days Pre Event	0.0618	0.0152	4.0815	0.0000	***
Method: FE, Event Period					
IQR of Dictionary Score	-3.2167	0.9450	-3.4041	0.0007	***
Log of Total Posts	-0.1771	0.0880	-2.0130	0.0444	**
CAR 10 Days Pre Event	0.1828	0.0157	11.6723	0.0000	***
CAR 20 Days Pre Event	0.1142	0.0107	10.6980	0.0000	***
CAR 5 Days Pre Event	0.2795	0.0223	12.5154	0.0000	***
Method: FE, Period: Pre 5					
Average NLP Score	0.9999	0.4557	2.1941	0.0284	**
Method: FE, Period: Pre 10					
Average NLP Score	1.7884	0.6545	2.7327	0.0064	***
Percentage of Negative Posts	-4.1506	1.6368	-2.5358	0.0114	**
Method: OLS, Period: Post 20					
IQR of NLP Score	-5.5370	2.4739	-2.2382	0.0254	**
Total Posts	-0.0184	0.0053	-3.4889	0.0005	***
Log of Total Posts	-0.8397	0.1838	-4.5681	0.0000	***
CAR 10 Days Pre Event	-0.0849	0.0404	-2.1018	0.0358	**
CAR 5 Days Pre Event	-0.2311	0.0575	-4.0186	0.0001	***
Method: OLS, Period: Post 10					
IQR of NLP Score	-5.9433	1.9113	-3.1096	0.0019	***
Total Posts	-0.0119	0.0041	-2.9154	0.0036	***
Log of Total Posts	-0.6433	0.1423	-4.5198	0.0000	***
CAR 10 Days Pre Event	0.1261	0.0311	4.0528	0.0001	***
CAR 20 Days Pre Event	0.0731	0.0212	3.4437	0.0006	***
Method: OLS, Period: Post 5					
IQR of NLP Score	-4.6392	1.3272	-3.4955	0.0005	***
Log of Total Posts	-0.2402	0.0996	-2.4126	0.0160	**
CAR 10 Days Pre Event	0.0767	0.0217	3.5427	0.0004	***
CAR 20 Days Pre Event	0.0614	0.0147	4.1732	0.0000	***
Method: OLS, Event Period					
IQR of Dictionary Score	-2.7443	0.8060	-3.4050	0.0007	***
Log of Total Posts	-0.1989	0.0735	-2.7063	0.0069	***
CAR 10 Days Pre Event	0.1808	0.0152	11.9063	0.0000	***
CAR 20 Days Pre Event	0.1159	0.0104	11.1403	0.0000	***
CAR 5 Days Pre Event	0.2747	0.0215	12.7451	0.0000	***
Method: OLS, Period: Pre 20					
Total Posts	-0.0131	0.0057	-2.3113	0.0210	**

Tab. 8: Panel regression results using fixed effects (FE) for all explanatory variables across all event windows. Robust entity clustered standard errors were used for all regressions. *** indicates significance at the 1%, ** indicates significance at the 5% level, and * indicates significance at the 10% level.

Variable	20-Day Pre	10-Day Pre	5-Day Pre	Event Period	5-Day Post	10-Day Post	20-Day Post
Log of Total Posts	0.6740	0.7587**	0.2408	-0.2359*	-0.4091**	-1.2229***	-1.4327***
	(0.4368)	(0.2998)	(0.1802)	(0.1240)	(0.2033)	(0.2646)	(0.3405)
Total Posts	-0.0202	-0.0144	-0.0063	0.0048	0.0064	0.0110	0.0024
	(0.0157)	(0.0099)	(0.0043)	(0.0029)	(0.0064)	(0.0096)	(0.0115)
Percentage Earnings Surprise	-0.0187	-0.0224	0.0047	-0.0136	-0.0079	-0.0510	-0.1077**
	(0.0539)	(0.0327)	(0.0188)	(0.0148)	(0.0283)	(0.0419)	(0.0456)
Earnings Surprise	-0.2757	-0.2517	0.1955	0.4287*	0.4943	0.3768	0.2782
	(0.5287)	(0.2884)	(0.2944)	(0.2301)	(0.3064)	(0.4199)	(0.6485)
Negative Surprise Dummy	0.2414	-0.0899	0.1280	-0.1097	0.2758	-0.0964	0.1850
	(0.5229)	(0.3406)	(0.2317)	(0.1858)	(0.2345)	(0.3280)	(0.4478)
CAR 5 Days Pre Event				0.1836***	-0.2114***	-0.1816***	-0.2970***
				(0.0345)	(0.0519)	(0.0642)	(0.0887)
CAR 10 Days Pre Event				0.0444	0.1109***	0.1930***	0.0654
				(0.0370)	(0.0407)	(0.0673)	(0.0864)
CAR 20 Days Pre Event				0.0488**	0.0753***	0.0504	0.0057
				(0.0218)	(0.0250)	(0.0392)	(0.0487)
Average NLP Score	10.1522***	7.5551***	4.9709***	-1.6534	-4.9677***	-10.4276***	-13.0952***
	(3.2867)	(2.5099)	(1.6857)	(1.1520)	(1.7460)	(2.7846)	(3.3463)
IQR of Dictionary Score	-0.0718	-0.0515	-1.2019	-2.9369***	-2.1359	0.6866	-0.5203
	(2.8750)	(2.1433)	(1.5691)	(1.0756)	(1.4303)	(2.0510)	(3.0598)
IQR of NLP Score	8.6131**	2.3300	0.3770	-0.9893	-6.5485***	-7.1677***	-0.8731
	(3.5602)	(2.4951)	(1.8620)	(1.2855)	(1.9833)	(2.3481)	(2.9496)
Average Dictionary Score	17.5734***	10.9682***	4.3977**	-3.3455**	-5.7553***	-10.6458***	-14.4640***
	(4.1724)	(2.5615)	(1.8226)	(1.4205)	(1.9968)	(3.0123)	(3.9265)
Percentage of Negative Posts	9.0927	6.2571	7.7446*	-0.6119	-7.5208*	-16.5423***	-19.4315***
	(7.4221)	(5.5506)	(4.1005)	(2.6662)	(3.8981)	(5.9760)	(7.3904)
R2	0.0258	0.0284	0.0148	0.1684	0.0670	0.0669	0.0581
Adj_R2	0.0164	0.0190	0.0053	0.1579	0.0552	0.0552	0.0463
F_stat	2.5144	2.7775	1.4289	14.7512	5.2284	5.2269	4.4958
F_pval	0.005497	0.002176	0.1622	1.11e-16	4.302e-09	4.336e-09	1.766e-07
N	1046	1046	1046	1046	1046	1046	1046

Tab. 9: Panel regression results using standard OLS for all explanatory variables across all event windows. Robust entity clustered standard errors were used for all regressions. *** indicates significance at the 1%, ** indicates significance at the 5% level, and * indicates significance at the 10% level.

Variable	20-Day Pre	10-Day Pre	5-Day Pre	Event Period	5-Day Post	10-Day Post	20-Day Post
Log of Total Posts	0.1958	0.4727*	0.0426	-0.1640	-0.1289	-0.6645***	-0.6205*
	(0.3725)	(0.2544)	(0.1521)	(0.1024)	(0.1726)	(0.2382)	(0.3185)
Total Posts	-0.0183	-0.0130	-0.0040	0.0022	0.0034	0.0039	-0.0050
	(0.0125)	(0.0086)	(0.0038)	(0.0029)	(0.0052)	(0.0084)	(0.0105)
Percentage Earnings Surprise	-0.0642	-0.0437*	-0.0066	-0.0030	0.0089	0.0015	-0.0505
	(0.0448)	(0.0244)	(0.0182)	(0.0120)	(0.0234)	(0.0356)	(0.0475)
Earnings Surprise	0.1010	0.0126	0.2809	0.3503*	0.3584	0.4316	0.4414
	(0.4958)	(0.2671)	(0.2820)	(0.2094)	(0.2731)	(0.4184)	(0.5480)
Negative Surprise Dummy	0.2854	0.1375	0.2718	-0.1310	0.2404	0.0052	0.2598
	(0.4663)	(0.3019)	(0.2057)	(0.1581)	(0.2001)	(0.2938)	(0.3631)
CAR 5 Days Pre Event				0.1747***	-0.2180***	-0.2078***	-0.3194***
				(0.0333)	(0.0525)	(0.0638)	(0.0847)
CAR 10 Days Pre Event				0.0383	0.1044***	0.1896***	0.0359
				(0.0352)	(0.0389)	(0.0649)	(0.0846)
CAR 20 Days Pre Event				0.0532**	0.0731***	0.0421	0.0078
				(0.0210)	(0.0242)	(0.0382)	(0.0494)
Average NLP Score	2.0162*	-0.0750	0.3822	0.3282	0.6867	1.8546**	1.9702*
	(1.0564)	(0.7582)	(0.4487)	(0.3553)	(0.5754)	(0.8569)	(1.0401)
IQR of Dictionary Score	1.3165	-0.2312	-1.2507	-2.5629***	-2.0007	0.5442	-0.3588
	(2.2383)	(1.6709)	(1.2710)	(0.9728)	(1.2711)	(1.7699)	(2.3481)
IQR of NLP Score	6.8496**	2.8059	0.1566	-0.6265	-3.9320**	-2.4890	-0.0429
	(2.8744)	(2.0412)	(1.5519)	(0.9247)	(1.7224)	(1.8852)	(2.3406)
Average Dictionary Score	10.5314***	5.2976**	1.0279	-2.6902**	-3.1335*	-4.2783	-6.0510**
	(3.1598)	(2.1065)	(1.7366)	(1.2262)	(1.6882)	(2.6031)	(3.0684)
Percentage of Negative Posts	-4.5712	-6.9630***	0.2866	3.5206***	4.5210***	8.5483***	10.6101***
	(3.5390)	(2.4220)	(1.7964)	(1.1626)	(1.6699)	(2.5166)	(3.2822)
R2	0.0488	0.0194	0.0057	0.1629	0.0530	0.0466	0.0597
Adj_R2	0.0396	0.0099	-0.0039	0.1524	0.0410	0.0346	0.0478
F_stat	5.3104	2.0461	0.5936	15.4671	4.4434	3.8824	5.0418
F_pval	1.148e-07	0.0262	0.8202	1.11e-16	2.196e-07	3.577e-06	1.051e-08
N	1046	1046	1046	1046	1046	1046	1046

Tab. 10: Panel regression results using fixed effects and standard OLS for all sentiment variables across all event windows. Robust entity clustered standard errors were used for all regressions. *** indicates significance at the 1%, ** indicates significance at the 5% level, and * indicates significance at the 10% level.

Variable	20-Day Pre	10-Day Pre	5-Day Pre	Event Period	5-Day Post	10-Day Post	20-Day Post
				FE Results			
Average NLP Score	7.4233**	5.2282*	4.6679***	0.2156	-2.6413	-7.3044***	-11.8386***
	(3.5890)	(2.7114)	(1.5795)	(1.1875)	(1.6890)	(2.5960)	(3.3801)
IQR of Dictionary Score	0.9559	0.6076	-0.4408	-3.5043***	-1.9881	-1.4656	-5.2555*
	(2.4584)	(1.7679)	(1.2703)	(0.9463)	(1.6219)	(2.1854)	(2.9819)
IQR of NLP Score	8.9588***	3.9325	0.9936	-1.3011	-6.6652***	-9.1739***	-5.4577*
	(3.3770)	(2.4000)	(1.7226)	(1.3136)	(1.8537)	(2.4076)	(3.0396)
Average Dictionary Score	15.7280***	7.6268***	4.2004**	-0.9884	-1.7963	-4.2660	-9.9248***
	(3.7654)	(2.4711)	(1.6899)	(1.4463)	(2.0130)	(2.9693)	(3.7408)
Percentage of Negative Posts	2.5784	2.4590	6.5250*	1.9165	-4.2236	-12.9780**	-18.5239***
	(8.1378)	(6.2795)	(3.8667)	(2.8159)	(3.5419)	(5.3950)	(7.0530)
R2	0.0187	0.0154	0.0122	0.0124	0.0146	0.0146	0.0140
Adj_R2	0.0144	0.0111	0.0079	0.0081	0.0103	0.0103	0.0097
F_stat	4.0274	3.3021	2.6181	2.6646	3.1386	3.1447	3.0074
F_pval	0.001263	0.005775	0.02311	0.02107	0.008084	0.007984	0.01057
N	1149	1149	1149	1149	1149	1149	1149
				OLS Results			
Average NLP Score	2.3439***	0.6583	0.5056	0.3840	0.9256**	0.9891*	1.0897
	(0.7134)	(0.5279)	(0.3336)	(0.2903)	(0.4086)	(0.5992)	(0.6793)
IQR of Dictionary Score	1.3207	-0.0312	-1.0145	-3.1447***	-1.3101	-0.2402	-2.7905
	(2.0495)	(1.5717)	(1.1435)	(0.9186)	(1.3030)	(1.7519)	(2.2822)
IQR of NLP Score	6.7021**	4.4317**	0.4165	-1.2010	-4.0215***	-4.8288***	-3.8621*
	(2.6009)	(1.9400)	(1.3693)	(0.8756)	(1.4387)	(1.6989)	(2.0824)
Average Dictionary Score	10.6514***	4.1069*	1.3889	-1.5614	-0.7946	-1.3583	-4.9403
	(3.0684)	(2.1139)	(1.6162)	(1.3261)	(1.6784)	(2.5295)	(3.1498)
Percentage of Negative Posts	-5.6869*	-5.1275**	-0.1169	2.8930**	3.5428**	3.7342	6.9274**
	(3.3819)	(2.3711)	(1.5824)	(1.1298)	(1.6296)	(2.4808)	(3.1678)
R2	0.0445	0.0137	0.0027	0.0126	0.0110	0.0112	0.0270
Adj_R2	0.0404	0.0094	-0.0016	0.0083	0.0067	0.0069	0.0227
F_stat	10.6647	3.1733	0.6264	2.9169	2.5444	2.6005	6.3505
F_pval	4.941e-10	0.007506	0.6797	0.01267	0.02669	0.02389	7.949e-06
N	1149	1149	1149	1149	1149	1149	1149

Tab. 11: Panel regression results using fixed effects and standard OLS for all literature variables across all event windows. Robust entity clustered standard errors were used for all regressions. *** indicates significance at the 1%, ** indicates significance at the 5% level, and * indicates significance at the 10% level.

Variable	20-Day Pre	10-Day Pre	5-Day Pre	Event Period	5-Day Post	10-Day Post	20-Day Post
				FE Results			
Log of Total Posts	0.3661	0.5161*	0.0690	-0.3041***	-0.4649**	-0.9937***	-1.0276***
	(0.4043)	(0.2872)	(0.1429)	(0.1030)	(0.1834)	(0.2661)	(0.3299)
Total Posts	-0.0188	-0.0146	-0.0048	0.0060**	0.0062	0.0076	-0.0021
	(0.0160)	(0.0106)	(0.0038)	(0.0028)	(0.0062)	(0.0099)	(0.0116)
Percentage Earnings Surprise	-0.0388	-0.0501*	-0.0198	-0.0199*	-0.0082	-0.0210	-0.0626
	(0.0526)	(0.0286)	(0.0183)	(0.0114)	(0.0250)	(0.0368)	(0.0472)
Earnings Surprise	-0.0838	-0.0359	0.2706	0.3843*	0.4966	0.3159	0.1068
	(0.5358)	(0.2791)	(0.2889)	(0.2151)	(0.3018)	(0.3954)	(0.6275)
Negative Surprise Dummy	0.2218	-0.0814	0.1359	-0.1243	0.2853	-0.0662	0.1803
	(0.5345)	(0.3425)	(0.2323)	(0.1840)	(0.2299)	(0.3306)	(0.4459)
CAR 5 Days Pre Event				0.1864***	-0.2091***	-0.1841***	-0.3043***
				(0.0347)	(0.0524)	(0.0643)	(0.0877)
CAR 10 Days Pre Event				0.0420	0.1128***	0.1870***	0.0484
				(0.0369)	(0.0397)	(0.0688)	(0.0880)
CAR 20 Days Pre Event				0.0471**	0.0660***	0.0388	-0.0000
				(0.0218)	(0.0241)	(0.0397)	(0.0491)
R2	0.0042	0.0061	0.0020	0.1610	0.0484	0.0444	0.0393
Adj_R2	-0.0006	0.0013	-0.0028	0.1545	0.0411	0.0371	0.0319
F_stat	0.7975	1.1677	0.3832	22.8554	6.0590	5.5382	4.8748
F_pval	0.5515	0.3231	0.8605	1.11e-16	1.271e-07	7.251e-07	6.51e-06
N	1047	1047	1047	1047	1047	1047	1047
				OLS Results			
Log of Total Posts	0.5534***	0.2016**	0.0241	-0.0622	-0.0376	0.1096	0.3981***
	(0.1049)	(0.0940)	(0.0427)	(0.0402)	(0.0645)	(0.1146)	(0.1102)
Total Posts	-0.0255***	-0.0117	-0.0043*	0.0007	0.0007	-0.0083	-0.0196***
	(0.0079)	(0.0076)	(0.0025)	(0.0028)	(0.0046)	(0.0093)	(0.0071)
Percentage Earnings Surprise	-0.0507	-0.0395	-0.0114	-0.0159*	-0.0101	-0.0121	-0.0644
	(0.0505)	(0.0246)	(0.0170)	(0.0096)	(0.0239)	(0.0403)	(0.0419)
Earnings Surprise	0.2812	0.1448	0.3163	0.3385*	0.3536	0.3818	0.3626
	(0.4791)	(0.2493)	(0.2868)	(0.2014)	(0.2762)	(0.4358)	(0.5783)
Negative Surprise Dummy	0.3471	0.0543	0.2680	-0.1044	0.2565	0.1689	0.4833
	(0.4670)	(0.2895)	(0.2054)	(0.1528)	(0.2001)	(0.2942)	(0.3640)
CAR 5 Days Pre Event				0.1814***	-0.2082***	-0.1907***	-0.3001***
				(0.0332)	(0.0513)	(0.0624)	(0.0832)
CAR 10 Days Pre Event				0.0322	0.0982***	0.1635**	0.0019
				(0.0350)	(0.0375)	(0.0663)	(0.0844)
CAR 20 Days Pre Event				0.0536**	0.0703***	0.0477	0.0179
				(0.0213)	(0.0231)	(0.0383)	(0.0491)
R2	0.0368	0.0088	0.0035	0.1541	0.0401	0.0289	0.0453
Adj_R2	0.0322	0.0041	-0.0013	0.1476	0.0327	0.0215	0.0379
F_stat	7.9701	1.8546	0.7279	23.6648	5.4193	3.8704	6.1625
F_pval	2.202e-07	0.09968	0.6026	1.11e-16	1.046e-06	0.0001656	8.638e-08
N	1047	1047	1047	1047	1047	1047	1047

Tab. 12: Panel regression results using fixed effects and standard OLS for literature and NLP based variables across all event windows. Robust entity clustered standard errors were used for all regressions. *** indicates significance at the 1%, ** indicates significance at the 5% level, and * indicates significance at the 10% level.

Variable	20-Day Pre	10-Day Pre	5-Day Pre	Event Period	5-Day Post	10-Day Post	20-Day Post
				FE Results			
Log of Total Posts	0.1622	0.4370	0.0403	-0.2749***	-0.3218*	-0.8422***	-0.9918***
	(0.4245)	(0.2904)	(0.1552)	(0.1040)	(0.1928)	(0.2628)	(0.3373)
Total Posts	-0.0140	-0.0103	-0.0029	0.0054**	0.0045	0.0044	-0.0051
	(0.0164)	(0.0105)	(0.0042)	(0.0027)	(0.0062)	(0.0097)	(0.0116)
Percentage Earnings Surprise	-0.0335	-0.0326	-0.0110	-0.0203	-0.0012	-0.0227	-0.0779*
	(0.0521)	(0.0286)	(0.0182)	(0.0125)	(0.0249)	(0.0381)	(0.0464)
Earnings Surprise	-0.1602	-0.1823	0.1965	0.3937*	0.4717	0.3585	0.2354
	(0.5472)	(0.2838)	(0.2913)	(0.2179)	(0.3065)	(0.4198)	(0.6480)
Negative Surprise Dummy	0.2117	-0.1090	0.1103	-0.1188	0.2802	-0.0637	0.2196
	(0.5338)	(0.3437)	(0.2336)	(0.1850)	(0.2334)	(0.3335)	(0.4526)
CAR 5 Days Pre Event				0.1864***	-0.2121***	-0.1877***	-0.3026***
				(0.0348)	(0.0518)	(0.0638)	(0.0881)
CAR 10 Days Pre Event				0.0421	0.1079***	0.1882***	0.0584
				(0.0367)	(0.0402)	(0.0675)	(0.0868)
CAR 20 Days Pre Event				0.0477**	0.0721***	0.0428	-0.0040
				(0.0218)	(0.0245)	(0.0394)	(0.0486)
Average NLP Score	1.9326*	2.2032***	0.9443	-0.2119	-0.3071	-1.2662	-1.6660
	(0.9916)	(0.8310)	(0.5841)	(0.3929)	(0.5953)	(0.8572)	(1.0618)
IQR of NLP Score	8.0699**	1.9670	-0.0085	-0.9810	-6.1989***	-6.3021***	0.1432
	(3.6186)	(2.4850)	(1.8650)	(1.2924)	(2.0313)	(2.4198)	(2.9946)
R2	0.0091	0.0138	0.0053	0.1616	0.0582	0.0494	0.0413
Adj_R2	0.0024	0.0072	-0.0014	0.1535	0.0491	0.0402	0.0320
F_stat	1.2542	1.9068	0.7246	18.3059	5.8699	4.9391	4.0910
F_pval	0.2702	0.06535	0.6512	1.11e-16	1.225e-08	5.353e-07	1.549e-05
N	1046	1046	1046	1046	1046	1046	1046
				OLS Results			
Log of Total Posts	0.0156	0.1501	-0.0208	-0.1449*	-0.0364	-0.2656	-0.1767
	(0.3527)	(0.2449)	(0.1412)	(0.0855)	(0.1588)	(0.2294)	(0.2637)
Total Posts	-0.0176	-0.0109	-0.0033	0.0027	0.0034	0.0011	-0.0078
	(0.0130)	(0.0101)	(0.0038)	(0.0031)	(0.0057)	(0.0102)	(0.0093)
Percentage Earnings Surprise	-0.0569	-0.0400	-0.0108	-0.0138	-0.0008	-0.0028	-0.0592
	(0.0489)	(0.0245)	(0.0173)	(0.0103)	(0.0232)	(0.0362)	(0.0462)
Earnings Surprise	0.2579	0.1421	0.3085	0.3192	0.3056	0.2955	0.2767
	(0.4867)	(0.2468)	(0.2855)	(0.1989)	(0.2696)	(0.4242)	(0.5571)
Negative Surprise Dummy	0.2332	0.0435	0.2601	-0.1154	0.2747	0.1172	0.3880
	(0.4538)	(0.2903)	(0.2027)	(0.1544)	(0.1988)	(0.2946)	(0.3632)
CAR 5 Days Pre Event				0.1815***	-0.2093***	-0.1916***	-0.2992***
				(0.0333)	(0.0511)	(0.0627)	(0.0833)
CAR 10 Days Pre Event				0.0326	0.0952**	0.1660**	0.0082
				(0.0352)	(0.0382)	(0.0668)	(0.0845)
CAR 20 Days Pre Event				0.0532**	0.0737***	0.0457	0.0114
				(0.0213)	(0.0236)	(0.0383)	(0.0491)
Average NLP Score	0.5863	0.0621	0.1127	0.2692	0.5377*	1.2049***	1.3265**
	(0.6930)	(0.4745)	(0.3044)	(0.1907)	(0.3030)	(0.4237)	(0.5461)
IQR of NLP Score	6.5520***	0.5978	0.2279	0.0970	-2.7737*	0.5423	3.4744*
	(2.3910)	(1.7720)	(1.2755)	(0.8031)	(1.4764)	(1.6142)	(2.0895)
R2	0.0415	0.0089	0.0036	0.1555	0.0476	0.0348	0.0499
Adj_R2	0.0351	0.0022	-0.0031	0.1473	0.0384	0.0255	0.0407
F_stat	6.4302	1.3345	0.5380	19.0729	5.1799	3.7367	5.4424
F_pval	1.981e-07	0.2304	0.8061	1.11e-16	1.95e-07	6.007e-05	6.7e-08
N	1046	1046	1046	1046	1046	1046	1046

Tab. 13: Panel regression results using fixed effects and standard OLS for literature and dictionary based variables across all event windows. Robust entity clustered standard errors were used for all regressions. *** indicates significance at the 1%, ** indicates significance at the 5% level, and * indicates significance at the 10% level.

Variable	20-Day Pre	10-Day Pre	5-Day Pre	Event Period	5-Day Post	10-Day Post	20-Day Post
				FE Results			
Log of Total Posts	0.7496*	0.7220**	0.1882	-0.2356*	-0.4922***	-1.2471***	-1.2768***
	(0.4338)	(0.3088)	(0.1759)	(0.1214)	(0.1883)	(0.2582)	(0.3345)
Total Posts	-0.0246	-0.0167	-0.0058	0.0056*	0.0076	0.0114	0.0013
	(0.0159)	(0.0104)	(0.0039)	(0.0029)	(0.0062)	(0.0095)	(0.0113)
Percentage Earnings Surprise	-0.0358	-0.0375	-0.0106	-0.0118	0.0024	-0.0238	-0.0693
	(0.0523)	(0.0298)	(0.0191)	(0.0135)	(0.0261)	(0.0395)	(0.0500)
Earnings Surprise	-0.0524	-0.0512	0.2533	0.3700*	0.4656	0.2973	0.1034
	(0.5412)	(0.2851)	(0.2840)	(0.2206)	(0.3095)	(0.3994)	(0.6378)
Negative Surprise Dummy	0.2561	-0.0695	0.1323	-0.1159	0.2838	-0.0909	0.1701
	(0.5299)	(0.3433)	(0.2321)	(0.1829)	(0.2355)	(0.3345)	(0.4526)
CAR 5 Days Pre Event				0.1854***	-0.2097***	-0.1818***	-0.2999***
				(0.0344)	(0.0524)	(0.0649)	(0.0888)
CAR 10 Days Pre Event				0.0400	0.1091***	0.1852***	0.0472
				(0.0370)	(0.0403)	(0.0685)	(0.0875)
CAR 20 Days Pre Event				0.0477**	0.0692***	0.0433	0.0037
				(0.0217)	(0.0245)	(0.0388)	(0.0485)
IQR of Dictionary Score	-4.8648*	-4.1864**	-2.8329*	-1.8464*	-1.0694	3.3182*	4.0155
	(2.5745)	(1.9407)	(1.4653)	(1.0260)	(1.4895)	(1.9405)	(2.8882)
Average Dictionary Score	5.6045**	0.8499	0.2865	-0.6379	-2.6861**	-3.4854*	-2.8383
	(2.4317)	(1.8550)	(1.3538)	(1.0486)	(1.3460)	(1.9543)	(2.5737)
R2	0.0105	0.0106	0.0061	0.1641	0.0518	0.0493	0.0419
Adj_R2	0.0038	0.0039	-0.0006	0.1561	0.0426	0.0401	0.0326
F_stat	1.4436	1.4523	0.8382	18.6536	5.1877	4.9231	4.1509
F_pval	0.1842	0.1808	0.5556	1.11e-16	1.966e-07	5.709e-07	1.225e-05
N	1046	1046	1046	1046	1046	1046	1046
				OLS Results			
Log of Total Posts	0.8318***	0.4298***	0.1269	-0.0424	-0.1157	-0.2452	0.0265
	(0.1854)	(0.1500)	(0.0939)	(0.0739)	(0.1032)	(0.1630)	(0.1726)
Total Posts	-0.0317***	-0.0151*	-0.0056*	0.0015	0.0040	-0.0011	-0.0122
	(0.0087)	(0.0078)	(0.0029)	(0.0031)	(0.0049)	(0.0092)	(0.0077)
Percentage Earnings Surprise	-0.0531	-0.0318	-0.0068	-0.0087	-0.0002	-0.0126	-0.0662
	(0.0469)	(0.0254)	(0.0178)	(0.0113)	(0.0239)	(0.0379)	(0.0461)
Earnings Surprise	0.2746	0.1094	0.2970	0.3152	0.3272	0.4016	0.3870
	(0.4939)	(0.2593)	(0.2852)	(0.2021)	(0.2787)	(0.4328)	(0.5837)
Negative Surprise Dummy	0.3881	0.0978	0.2883	-0.0935	0.2547	0.1135	0.4251
	(0.4771)	(0.2964)	(0.2077)	(0.1531)	(0.1993)	(0.2879)	(0.3599)
CAR 5 Days Pre Event				0.1813***	-0.2089***	-0.1913***	-0.2996***
				(0.0330)	(0.0514)	(0.0626)	(0.0835)
CAR 10 Days Pre Event				0.0298	0.0959**	0.1665**	0.0051
				(0.0350)	(0.0382)	(0.0666)	(0.0848)
CAR 20 Days Pre Event				0.0552***	0.0730***	0.0496	0.0195
				(0.0212)	(0.0235)	(0.0381)	(0.0492)
IQR of Dictionary Score	-1.2483	-2.7482*	-1.4254	-1.4150	-1.2854	2.2566	2.5690
	(1.8862)	(1.4125)	(1.1079)	(0.8789)	(1.1468)	(1.4041)	(1.9282)
Average Dictionary Score	3.3176*	0.9708	0.2448	-1.0951	-2.5644***	-3.5177***	-3.4664***
	(1.7782)	(1.2256)	(0.9829)	(0.7000)	(0.8036)	(1.0807)	(1.3327)
R2	0.0395	0.0124	0.0051	0.1575	0.0451	0.0364	0.0499
Adj_R2	0.0330	0.0057	-0.0016	0.1493	0.0359	0.0271	0.0407
F_stat	6.1021	1.8621	0.7683	19.3602	4.8955	3.9092	5.4355
F_pval	5.292e-07	0.07248	0.6141	1.11e-16	6.16e-07	3.081e-05	6.893e-08
N	1046	1046	1046	1046	1046	1046	1046

Tab. 14: Panel regression results using fixed effects and standard OLS for literature variables and the percentage of negative posts across all event windows. Robust entity clustered standard errors were used for all regressions. *** indicates significance at the 1%, ** indicates significance at the 5% level, and * indicates significance at the 10% level.

Variable	20-Day Pre	10-Day Pre	5-Day Pre	Event Period	5-Day Post	10-Day Post	20-Day Post
				FE Results			
Log of Total Posts	0.3462	0.4748*	0.0539	-0.3023***	-0.4763**	-0.9919***	-1.0077***
	(0.4052)	(0.2830)	(0.1443)	(0.1025)	(0.1850)	(0.2657)	(0.3307)
Total Posts	-0.0170	-0.0108	-0.0034	0.0058**	0.0073	0.0075	-0.0040
	(0.0162)	(0.0104)	(0.0040)	(0.0027)	(0.0061)	(0.0097)	(0.0115)
Percentage Earnings Surprise	-0.0349	-0.0420	-0.0168	-0.0203*	-0.0058	-0.0214	-0.0667
	(0.0531)	(0.0281)	(0.0182)	(0.0118)	(0.0252)	(0.0375)	(0.0471)
Earnings Surprise	-0.1504	-0.1747	0.2201	0.3910*	0.4538	0.3226	0.1814
	(0.5453)	(0.2822)	(0.2919)	(0.2197)	(0.3118)	(0.4126)	(0.6491)
Negative Surprise Dummy	0.2110	-0.1039	0.1277	-0.1231	0.2778	-0.0650	0.1934
	(0.5331)	(0.3412)	(0.2319)	(0.1837)	(0.2306)	(0.3325)	(0.4499)
CAR 5 Days Pre Event				0.1863***	-0.2084***	-0.1842***	-0.3056***
				(0.0347)	(0.0523)	(0.0644)	(0.0879)
CAR 10 Days Pre Event				0.0425	0.1095***	0.1875***	0.0542
				(0.0367)	(0.0400)	(0.0671)	(0.0865)
CAR 20 Days Pre Event				0.0469**	0.0671***	0.0386	-0.0020
				(0.0217)	(0.0241)	(0.0392)	(0.0488)
Percentage of Negative Posts	-2.0812	-4.3313**	-1.5784	0.2086	-1.3294	0.2077	2.3164
	(2.2573)	(1.9106)	(1.3543)	(0.9044)	(1.3579)	(1.9909)	(2.4184)
R2	0.0047	0.0114	0.0035	0.1610	0.0493	0.0444	0.0401
Adj_R2	-0.0010	0.0057	-0.0023	0.1537	0.0411	0.0361	0.0318
F_stat	0.7536	1.8433	0.5559	20.3009	5.4907	4.9189	4.4222
F_pval	0.6066	0.08779	0.7656	1.11e-16	2.195e-07	1.797e-06	1.09e-05
N	1047	1047	1047	1047	1047	1047	1047
				OLS Results			
Log of Total Posts	0.5486***	0.2247**	0.0256	-0.0636	-0.0261	0.1061	0.3814***
	(0.1022)	(0.0918)	(0.0433)	(0.0420)	(0.0679)	(0.1184)	(0.1111)
Total Posts	-0.0260***	-0.0097	-0.0041	0.0005	0.0018	-0.0086	-0.0211***
	(0.0084)	(0.0074)	(0.0028)	(0.0028)	(0.0048)	(0.0092)	(0.0075)
Percentage Earnings Surprise	-0.0515	-0.0359	-0.0111	-0.0161*	-0.0083	-0.0126	-0.0670
	(0.0499)	(0.0242)	(0.0171)	(0.0097)	(0.0237)	(0.0411)	(0.0414)
Earnings Surprise	0.2920	0.0935	0.3130	0.3418*	0.3273	0.3900	0.4008
	(0.4840)	(0.2466)	(0.2886)	(0.2044)	(0.2823)	(0.4441)	(0.5813)
Negative Surprise Dummy	0.3419	0.0792	0.2696	-0.1059	0.2683	0.1652	0.4661
	(0.4708)	(0.2947)	(0.2066)	(0.1538)	(0.1979)	(0.2912)	(0.3638)
CAR 5 Days Pre Event				0.1812***	-0.2066***	-0.1912***	-0.3024***
				(0.0332)	(0.0513)	(0.0627)	(0.0833)
CAR 10 Days Pre Event				0.0326	0.0948**	0.1645**	0.0069
				(0.0349)	(0.0377)	(0.0653)	(0.0837)
CAR 20 Days Pre Event				0.0535**	0.0717***	0.0473	0.0159
				(0.0212)	(0.0231)	(0.0380)	(0.0490)
Percentage of Negative Posts	0.4683	-2.2325	-0.1424	0.1386	-1.1239	0.3505	1.6340
	(1.7952)	(1.4893)	(0.9705)	(0.6403)	(0.9068)	(1.3222)	(1.4840)
R2	0.0369	0.0109	0.0035	0.1542	0.0410	0.0290	0.0459
Adj_R2	0.0313	0.0052	-0.0023	0.1468	0.0327	0.0206	0.0376
F_stat	6.6428	1.9070	0.6090	21.0194	4.9337	3.4424	5.5446
F_pval	6.576e-07	0.07673	0.7233	1.11e-16	1.654e-06	0.0003424	1.731e-07
N	1047	1047	1047	1047	1047	1047	1047

Tab. 15: Panel regression results using fixed effects and standard OLS for all explanatory variables across all event windows with the exception of pre event cumulative abnormal returns. Robust entity clustered standard errors were used for all regressions. *** indicates significance at the 1%, ** indicates significance at the 5% level, and * indicates significance at the 10% level.

Variable	20-Day Pre	10-Day Pre	5-Day Pre	Event Period	5-Day Post	10-Day Post	20-Day Post
Log of Total Posts	0.6740	0.7587**	0.2408	-0.1251	-0.3251*	-1.0862***	-1.4508***
	(0.4368)	(0.2998)	(0.1802)	(0.1246)	(0.1814)	(0.2434)	(0.3448)
Total Posts	-0.0202	-0.0144	-0.0063	0.0020	0.0046	0.0083	0.0032
	(0.0157)	(0.0099)	(0.0043)	(0.0027)	(0.0056)	(0.0085)	(0.0114)
Percentage Earnings Surprise	-0.0187	-0.0224	0.0047	-0.0147	-0.0127	-0.0571	-0.1107**
	(0.0539)	(0.0327)	(0.0188)	(0.0130)	(0.0241)	(0.0404)	(0.0458)
Earnings Surprise	-0.2757	-0.2517	0.1955	0.4400*	0.4043	0.2788	0.2021
	(0.5287)	(0.2884)	(0.2944)	(0.2487)	(0.3167)	(0.4182)	(0.6114)
Negative Surprise Dummy	0.2414	-0.0899	0.1280	-0.0784	0.2569	-0.1248	0.1425
	(0.5229)	(0.3406)	(0.2317)	(0.1847)	(0.2445)	(0.3378)	(0.4471)
Average NLP Score	10.1522***	7.5551***	4.9709***	0.0904	-4.4164**	-9.3607***	-14.0191***
	(3.2867)	(2.5099)	(1.6857)	(1.2687)	(1.7647)	(2.6312)	(3.3085)
IQR of Dictionary Score	-0.0718	-0.0515	-1.2019	-3.1634***	-1.8929	0.8914	-0.1672
	(2.8750)	(2.1433)	(1.5691)	(1.0860)	(1.4076)	(2.0527)	(3.0017)
IQR of NLP Score	8.6131**	2.3300	0.3770	-0.3961	-5.7213***	-6.3525***	-0.7834
	(3.5602)	(2.4951)	(1.8620)	(1.3824)	(1.9328)	(2.3817)	(3.0425)
Average Dictionary Score	17.5734***	10.9682***	4.3977**	-1.1930	-4.1456**	-8.4421***	-14.9520***
	(4.1724)	(2.5615)	(1.8226)	(1.5520)	(1.9136)	(2.9571)	(3.8549)
Percentage of Negative Posts	9.0927	6.2571	7.7446*	1.5320	-7.7797**	-16.2831***	-21.2700***
	(7.4221)	(5.5506)	(4.1005)	(2.9752)	(3.9445)	(5.8531)	(7.4137)
R2	0.0258	0.0284	0.0148	0.0151	0.0223	0.0363	0.0439
Adj_R2	0.0164	0.0190	0.0053	0.0055	0.0128	0.0270	0.0347
F_stat	2.5144	2.7775	1.4289	1.4523	2.1627	3.5775	4.3626
F_pval	0.005497	0.002176	0.1622	0.1525	0.01802	0.0001125	5.328e-06
N	1046	1046	1046	1046	1046	1046	1046