Social Media Sentiment and the Market Response to Earnings Surprises

Nicolas Queneherve

Dissertation written under the supervision of Tural Karimli

Dissertation submitted in partial fulfilment of requirements for the MSc in management, at Universidade Católica Portuguesa and for the MSc in management at Grenoble Ecole de Management, 24/05/2026.

Preface

I would like to express my sincere gratitude to my supervisor, Professor Tural Karimli, for his guidance, constructive feedback, and continuous support throughout the development of this dissertation. His comments and methodological advice greatly contributed to improving both the empirical analysis and the overall structure of this work.

I would also like to acknowledge the contribution of Cookson, Lu, Mullins, and Niessner for making the Social Signal dataset publicly available. Their work on social-media-based sentiment and attention measures provided a central component of the empirical analysis developed in this dissertation.

In addition, I am grateful to my colleagues and the broader Católica Lisbon academic environment for their encouragement, discussions, and feedback on the initial research idea and subsequent developments of the project.

Finally, I would like to thank my family and my friends for their constant support and encouragement throughout my studies.

Use of Artificial Intelligence Tools

Artificial intelligence tools were used during the preparation of this dissertation as supplementary research and writing assistance instruments. In particular, ChatGPT (OpenAI) and Gemini (Google) were used to support language refinement, formatting and the review of econometric interpretations and methodological explanations.

However, the research question, theoretical framework, empirical design, methodological choices, interpretation of results, and overall structure of the dissertation were independently developed by the author. All creative, analytical, and critical components of the work originate from the author's own reasoning and academic judgment.

Artificial intelligence tools were therefore used as support instruments for review and presentation purposes rather than as substitutes for original research or independent analysis. The final content, conclusions, and interpretations presented in this dissertation remain entirely the responsibility of the author.

Abstract

Author: Nicolas Queneherve

Title: Social Media Sentiment and the Market Response to Earnings Surprises

Keywords: Investor Sentiment, Earnings Announcements, Behavioral Finance, Event Study, Retail Frictions

This dissertation examines whether investor sentiment extracted from social media platforms influences stock returns and alters market sensitivity to earnings surprises. While traditional finance theory assumes rapid information efficiency, behavioral models suggest that prevailing public mood can bias information processing.

Using an event-study framework from 2019 to 2021, this study merges stock returns with a multi-platform sentiment index built from Twitter, StockTwits, and Seeking Alpha. The empirical strategy evaluates short-term market reactions and post-announcement dynamics using firm-level controls and Fama-French risk factors.

The empirical findings confirm that scaled earnings surprises heavily dictate immediate cumulative returns. However, the interaction between pre-announcement investor sentiment and fundamental news remains statistically and economically negligible. Ultimately, the results indicate that structured corporate disclosure environments remain anchored in accounting fundamentals rather than speculative retail narratives.

Resumo

Autor: Nicolas Queneherve

Título: Social Media Sentiment and the Market Response to Earnings Surprises

Palavras-chave: Sentimento dos Investidores, Anúncios de Resultados, Finanças Comportamentais, Estudo de Eventos, Fricções de Retalho

Esta dissertação analisa se o sentimento dos investidores extraído das redes sociais influencia os retornos das ações e se modifica a sensibilidade do mercado às surpresas nos resultados. Enquanto a teoria financeira tradicional assume a eficiência imediata da informação, os modelos comportamentais sugerem que o estado de espírito predominante do público pode enviesar o processamento da informação.

Utilizando uma abordagem de estudo de eventos para o período de 2019 a 2021, este estudo combina os retornos acionistas com um índice de sentimento multiplataforma construído a partir do Twitter, StockTwits e Seeking Alpha. A estratégia empírica avalia as reações do mercado no curto prazo e as dinâmicas pós-anúncio, recorrendo a controlos ao nível da empresa e aos fatores de risco de Fama-French.

Os resultados empíricos confirmam que as surpresas nos resultados normalizadas determinam fortemente os retornos cumulativos imediatos. No entanto, a interação entre o sentimento dos investidores pré-anúncio e a informação fundamental revela-se estatística e economicamente insignificante. Em suma, os resultados indicam que os contextos estruturados de divulgação de resultados empresariais permanecem ancorados nos fundamentos contabilísticos e não em narrativas especulativas de retalho.

Contents

Introduction

Over the past decade, social media has become an increasingly important source of information for financial markets. Individual investors now share opinions, interpret news, and express expectations about firms in real time. As a result, sentiment extracted from these platforms provides a direct measure of investors' beliefs, complementing traditional information sources. This raises an important question: to what extent do these sentiment signals influence asset prices, and do they reflect informational content or behavioural biases?

In standard financial theory, asset prices are expected to incorporate publicly available information efficiently. In the semi-strong form of the Efficient Market Hypothesis, prices should adjust immediately and without bias to new information, such as earnings announcements (Fama, 1970). However, empirical evidence has long challenged this view. In particular, the post-earnings announcement drift (PEAD) documented by Ball and Brown (1968) and further analyzed by Bernard and Thomas (1989) shows that stock prices often continue to move in the direction of earnings surprises well after the announcement. This suggests that markets may underreact to information, at least in the short run.

One potential explanation for such patterns lies in investor sentiment. A growing literature shows that sentiment can affect asset prices, especially for firms that are more difficult to value or arbitrage (Baker and Wurgler, 2006). More recently, the availability of large-scale textual data has allowed researchers to measure sentiment directly from social media platforms. Evidence suggests that these signals contain information relevant for future returns and market reactions (Chen et al., 2014). Despite these advances, there is still limited evidence on how sentiment interacts with the market response to earnings surprises specifically.

This dissertation investigates whether investor sentiment influences stock returns around earnings announcements and whether it modifies the sensitivity of returns to earnings surprises. Using a firm-day panel dataset combining stock returns, earnings announcement dates, earnings surprises, and social media sentiment, the analysis examines the dynamics of returns during, and after earnings events. The empirical strategy focuses on short event windows and tests whether sentiment measured prior to the announcement predicts subsequent returns, both directly and through its interaction with earnings surprises.

The results provide a nuanced picture of the relationship between sentiment and earnings announcements. Earnings surprises can predict short-term returns around announcement periods, confirming the informational importance of earnings news. However, the analysis does not provide robust evidence that sentiment significantly amplifies or attenuates the market response to earnings surprises. At the same time, return dynamics around announcements suggest that sentiment may contribute to short-term price movements surrounding earnings events, particularly over immediate post-announcement windows.

So, this thesis contributes to the literature by examining the interaction between social-media-based sentiment and surprise in earnings announcements within an event-study framework. Rather than showing that sentiment systematically alters the earnings response coefficient, the findings suggest that the influence of sentiment around earnings events may be more limited than some behavioural explanations imply.

Literature review

2.1 Investor Sentiment and Behavioural Finance

Traditional asset pricing theory assumes that financial markets efficiently incorporate available information into stock prices. Under the Efficient Market Hypothesis developed by Fama (1970), prices should primarily reflect expectations about future cash flows and risk, leaving little room for systematic pricing distortions caused by investor psychology. In such a framework, irrational trades are expected to be offset by rational arbitrageurs, preventing sentiment from exerting a persistent influence on market prices.

Behavioural finance emerged partly as a reaction to this view. Rather than assuming fully rational investors, this literature argues that market participants may systematically deviate from rational expectations because of psychological biases, limited information processing, or excessive confidence in their own beliefs. Daniel et al. (1998), for instance, show how overconfidence and biased self-attribution can generate overreaction to information and predictable return patterns. These mechanisms imply that investors do not necessarily interpret news in an objective manner, particularly in environments characterized by uncertainty or ambiguous valuation.

At the same time, irrational behaviour alone is not sufficient to generate persistent mispricing. A large part of the behavioural finance literature therefore focuses on limits to arbitrage. Even if rational investors recognize that prices deviate from fundamentals, correcting these distortions may be costly or risky. In practice, arbitrage is often constrained precisely in the stocks where valuation is the most subjective.

Baker and Wurgler (2006) provide an empirical framework on investor sentiment and stock returns. Their argument differs from earlier approaches because it focuses less on aggregate market mispricing and more on cross-sectional sensitivity to sentiment. The authors show that periods of high sentiment affect firms that are difficult to value and difficult to arbitrage, such as young, small, volatile, or unprofitable firms.

A contribution of their framework is the idea that sentiment does not influence all firms uniformly. Some stocks are naturally more exposed to speculative demand because their valuation relies heavily on expectations about uncertain future growth. Baker and Wurgler repeatedly emphasize that these firms are the ones for which arbitrage is least effective. This interaction between subjective valuation and limited arbitrage remains one of the core mechanisms through which behavioural distortions may persist in financial markets.

Still, many sentiment-based studies focus on broad return patterns over relatively long horizons, making it difficult to isolate how sentiment interacts with specific information events. In that sense, the behavioural interpretation often relies on evidence of mispricing correction rather than direct observation of investor reactions at the moment information arrives.

This issue becomes relevant around earnings announcements. Unlike broad market fluctuations, earnings releases constitute discrete and observable information shocks. Investors receive new information that must be interpreted rapidly, often under conditions of uncertainty regarding future firm performance. One possible implication is that sentiment may affect not only prices themselves, but also the way investors process unexpected news.

The growing use of textual analysis in finance expanded the empirical study of sentiment. Rather than relying exclusively on indirect market-based proxies, researchers increasingly began extracting sentiment from financial communication. Tetlock (2007) represents a turning point in this literature. Using textual analysis of the Wall Street Journal column *Abreast of the Market*, Tetlock shows that media pessimism predicts downward pressure on stock prices followed by a later reversal.

What makes Tetlock's contribution important is not simply the predictive result itself, but the broader implication that textual tone contains measurable information about investor behaviour. The study suggests that financial media may influence markets not only by transmitting fundamental information, but also by shaping or amplifying investor sentiment. Tetlock further finds that unusually high levels of pessimism are associated with higher trading volume, which is consistent with theories of noise trading and temporary sentiment-driven pressure on prices.

Nevertheless, interpreting textual sentiment remains challenging. A recurrent difficulty in this literature is distinguishing sentiment from information. Negative media tone, for example, may reflect genuine deterioration in fundamentals rather than irrational pessimism. Tetlock explicitly discusses this tension and argues that the observed reversals are more consistent with temporary sentiment effects than with permanent information shocks. Still, the boundary between sentiment and information remains difficult to identify empirically.

The development of social media introduced another important shift in sentiment research. Compared to traditional financial newspapers, social media platforms allow investors to react continuously and publicly to market events also giving the chance to non-professional investors to express themselves. This increased both the quantity and frequency of sentiment-related information available to researchers.

Cookson et al. (2025) extend this literature by constructing sentiment and attention measures using investor activity across platforms such as Twitter, StockTwits, and Seeking Alpha. Their approach is particularly relevant because it distinguishes investor attention from investor sentiment, two concepts that are often conflated in earlier research. The authors show that sentiment extracted from social media displays extrapolative dynamics, with optimistic tone tending to follow strong recent performance before partially reversing later. The Social Signal dataset used in this dissertation is based on the data developed by Cookson et al. (2025). The way they are calculating sentiment about a firm is detailed in there study but more information can be find on data medley (https://data.mendeley.com/datasets/xffyybvw4j/1)

At the same time, the use of social media data raises additional questions regarding investor composition and information quality. Social media activity is largely retail-driven, while price formation around major corporate announcements is often dominated by institutional investors and algorithmic trading systems. As a result, it is not obvious whether social-media-based sentiment should materially influence the pricing of earnings news, especially in highly followed firms.

2.2 Social Media Sentiment and Financial Markets

The development of online financial communication expanded the empirical study of investor sentiment. Earlier research on sentiment relied primarily on indirect market-based proxies such as trading volume, IPO activity, closed-end fund discounts, or survey-based indicators. While these measures provided useful information about broad market conditions, they offered only limited insight into how investors formed beliefs in real time.

The rise of internet forums and social media platforms changed this dynamic. Investors began discussing financial markets publicly and continuously through online platforms, generating large quantities of textual data that could be analyzed directly. This created new opportunities for researchers attempting to measure investor sentiment more precisely and at higher frequency.

One of the earliest contributions in this area is Antweiler and Frank (2004), who study discussions on internet message boards related to stocks. Their results suggest that online discussions contain information relevant to financial markets, even though a substantial share of the content appears noisy or speculative. The authors find stronger effects on trading activity and volatility than on returns themselves, which already hints at a recurring issue in this literature: online communication may reflect investor attention or disagreement more clearly than it predicts future prices.

Later studies progressively moved away from simple message volume toward textual analysis of investor tone. Chen et al. (2014), for example, show that opinions transmitted through social media platforms contain predictive information about stock returns and earnings performance. Their findings support the idea that collective investor discussions may aggregate dispersed information not immediately incorporated into prices. At the same time, the authors emphasize that the informational value of social media is uneven across firms and periods, suggesting that online sentiment should not be interpreted as a reliable signal.

This ambiguity remains central to the literature. Social media simultaneously creates information and noise. On the one hand, investor discussions may reveal beliefs, expectations, or reactions before they become reflected in prices. On the other hand, online platforms also amplify speculative behaviour and emotional reactions Distinguishing between these different mechanisms remains empirically difficult.

A further complication concerns the composition of investors active on social media platforms. Unlike institutional research reports or analyst forecasts, online discussions are largely driven by retail investors. This distinction matters because retail investors may process information differently from institutional market participants. Retail sentiment can influence prices temporarily, particularly in speculative or attention-driven environments. However, it remains less clear whether these effects persist in settings dominated by institutional trading and rapid information incorporation.

The increasing sophistication of textual analysis techniques considerably accelerated this line of research. Rather than manually classifying articles or messages, researchers began extracting sentiment automatically using natural language processing and machine learning tools. The objective is generally to quantify the tone embedded in textual communication by identifying optimistic, pessimistic, or uncertainty-related language.

Despite important methodological advances, measuring sentiment through text remains imperfect. The same textual content may simultaneously reflect information, opinion, or emotional reaction. A pessimistic message, for example, may indicate irrational fear, but it may also contain accurate information about deteriorating fundamentals.

This issue is particularly visible in financial social media, where attention and sentiment are often difficult to separate. Increased discussion around a stock may simply reflect heightened investor focus following important news rather than genuine optimism or pessimism.

Cookson et al. (2025) provide one of the most comprehensive approaches in this literature by constructing separate market-wide sentiment and attention indexes using investor activity across Twitter, StockTwits, and Seeking Alpha. Their methodology relies on millions of social media posts aggregated across platforms, allowing the authors to capture investor communication at daily frequency.

An important contribution of their study is the explicit distinction between sentiment and attention. The authors argue that these concepts reflect different economic mechanisms and should not be treated as equivalent variables. Their empirical results support this view. High sentiment tends to follow positive recent returns before reversing over subsequent weeks, while attention displays different dynamics linked more closely to trading activity and market stress.

The construction of the sentiment index itself also differs from earlier approaches. Rather than relying on a single platform, Cookson et al. aggregate signals across multiple investor communities using principal component analysis. The intuition behind this methodology is that the common component shared across platforms may better capture broad investor tone while filtering out platform-specific noise. Their results suggest that sentiment extracted from social media exhibits strong extrapolative behaviour, where optimism tends to increase following positive market performance.

Still, interpreting these measures requires caution. A large fraction of social-media-driven activity appears connected to retail investor behaviour, speculative trading, and short-term attention shocks. Cookson et al. themselves note that their attention index is strongly related to retail trading intensity. This raises an important question regarding the extent to which social-media-based sentiment can influence markets characterized by strong institutional participation.

It is not obvious that social-media sentiment should materially amplify the market reaction to earnings surprises. On the contrary, one could argue that the precision and visibility of earnings information reduce the scope for sentiment-driven interpretation. At the same time, periods preceding earnings announcements are often characterized by uncertainty regarding firm performance, analyst expectations, and future guidance. Under these conditions, prevailing investor tone may still influence how market participants interpret the announcement once it becomes public.

2.3 Earnings Announcements and Market Reactions

Earnings announcements occupy a central role in financial markets because they provide one important source of firm-specific information available to investors. Quarterly earnings releases contain updated information regarding profitability, operating performance, and prospects, allowing market participants to revise their expectations about firm value. As a result, earnings announcements are associated with significant trading activity and sharp price adjustments.

The informational role of earnings announcements was documented early in the accounting and finance literature. Ball and Brown (1968) show that stock prices react to unexpected earnings information, establishing one of the foundational results in empirical accounting research. Their findings suggest that earnings announcements convey economically relevant information that becomes incorporated into stock prices through market reactions.

Subsequent research focused more specifically on the concept of earnings surprises, generally defined as the difference between realized earnings and prior market expectations. Within this framework, the market response depends not on the level of earnings itself, but on whether the announcement exceeds or falls short of expectations embedded in analyst forecasts and prior pricing.

Hayn (1995) further emphasizes that the informational content of earnings is not constant across firms or situations. The market reaction to earnings announcements varies according to factors such as firm profitability, uncertainty, and the credibility of reported information. This line of research gradually led to the development of the Earnings Response Coefficient (ERC) literature, which examines the sensitivity of stock returns to earnings surprises.

The ERC can be interpreted as the intensity with which investors incorporate unexpected earnings information into prices. In practical terms, a higher ERC implies that a given earnings surprise generates a larger stock price reaction. A substantial part of the literature therefore attempts to identify the factors capable of amplifying or attenuating this relationship.

Also, the efficiency of market reactions to earnings announcements remains debated. Under the semi-strong form of market efficiency developed by Fama (1970), publicly available earnings information should be incorporated into prices rapidly and without predictable delay. In such a framework, abnormal returns following earnings announcements should be limited to the immediate reaction window.

Empirical evidence, however, suggests that this adjustment is not always instantaneous. Bernard and Thomas (1989) document the existence of Post-Earnings Announcement Drift (PEAD), showing that stock returns continue to drift in the direction of the earnings surprise for several weeks following the announcement. Firms reporting positive surprises tend to experience continued positive returns, while firms with negative surprises exhibit continued underperformance.

This phenomenon appears difficult to reconcile with fully efficient markets. Several explanations have been proposed, including delayed information processing, investor underreaction, limited attention, and gradual diffusion of information across market participants.

Livnat and Mendenhall (2006) revisit this literature by examining how different definitions of earnings surprise influence post-announcement returns. Their findings suggest that analyst-based earnings surprises possess stronger explanatory power for future returns than simpler accounting-based measures, reinforcing the importance of market expectations in determining price reactions.

The existence of PEAD also has important implications for event-study methodology. Because the market reaction may extend beyond the announcement day itself, many studies distinguish between immediate reaction windows and longer post-announcement periods. This distinction motivates the use of separate return windows when analyzing earnings events.

In the present study, the short window surrounding the announcement is intended to capture the initial incorporation of earnings information into prices, while the longer post-announcement window allows for the possibility of delayed adjustment or gradual correction. This approach follows the broader event-study literature, which often separates immediate market reactions from subsequent return dynamics.

Beyond delayed adjustment, another important issue concerns the interpretation of earnings information itself. Earnings announcements do not simply reveal numerical results; they also require investors to evaluate guidance, future prospects, and the broader implications of firm performance. Under such conditions, heterogeneous interpretation of the same announcement may emerge across market participants.

Several studies argue that investor attention plays a significant role in this process. Peress (2008), for example, shows that media coverage influences the speed and intensity of information diffusion in financial markets. When investor attention is limited, information may not be incorporated immediately into prices, leaving room for temporary pricing inefficiencies. Related studies further emphasize the importance of investor attention in earnings settings. DellaVigna and Pollet (2009) show that earnings announcements released on Fridays generate weaker immediate reactions and stronger subsequent drift, consistent with lower investor attention at the time of disclosure. Similarly, Hirshleifer, Lim, and Teoh (2009) argue that competing information events can distract investors and delay the incorporation of earnings news into prices. Da, Engelberg, and Gao (2011) extend this literature by introducing Google Search Volume as a direct proxy for investor attention, showing that periods of elevated search activity are associated with stronger market reactions and trading intensity. Taken together, these studies suggest that attention constraints may influence not only the speed of information incorporation but also the magnitude of market reactions around earnings announcements.

This possibility creates a natural connection between the earnings literature and the sentiment literature discussed previously. If investor sentiment affects how market participants interpret information, then sentiment may plausibly influence the magnitude of the earnings response coefficient. Positive sentiment could amplify reactions to favorable earnings surprises, while pessimistic environments may attenuate them or intensify reactions to negative news.

At the same time, earnings announcements differ from many other settings commonly examined in behavioural finance. Unlike speculative market narratives or broad sentiment cycles, earnings releases provide highly structured and closely monitored information.

Institutional investors, analysts, and algorithmic trading systems process these announcements rapidly, potentially limiting the scope for sentiment-driven distortions.

This ambiguity is precisely what makes earnings announcements an interesting empirical setting. On the one hand, earnings surprises represent moments of uncertainty resolution where investor interpretation may matter. On the other hand, the precision and visibility of earnings information may dominate softer behavioural signals such as sentiment.

2.4 Sentiment and Earnings Announcements

The intersection between investor sentiment and earnings announcements has progressively attracted attention in the finance literature because earnings releases constitute moments where market participants must interpret new information under uncertainty. While the earnings literature traditionally focuses on the informational content of earnings surprises, later studies considered the possibility that the market response may also depend on prevailing investor beliefs and sentiment conditions.

A central idea in this literature is that sentiment may influence the Earnings Response Coefficient (ERC), namely the sensitivity of stock returns to unexpected earnings information. Under this framework, the same earnings surprise may generate different price reactions depending on the broader sentiment environment in which the announcement occurs. Optimistic investors may react more strongly to positive surprises, while pessimistic environments may amplify negative reactions or attenuate positive ones.

Mian and Sankaraguruswamy (2012) provide analyses of this mechanism by examining whether investor sentiment affects the market response to earnings announcements. Their findings suggest that sentiment plays a role in shaping the reaction to earnings news, particularly for firms characterized by high uncertainty and stronger limits to arbitrage. More specifically, the authors show that optimistic sentiment tends to amplify reactions to positive earnings surprises, while pessimistic environments are associated with weaker market responses.

An important contribution of this literature is the recognition that earnings announcements are not interpreted mechanically. Even though earnings releases contain objective numerical information, investors must still evaluate the persistence of earnings, the credibility of management guidance, and the implications for future firm performance. As a result, behavioural factors may influence not only whether prices react, but also how strongly they react.

At the same time, the evidence remains far from uniform across studies. Some papers document strong sentiment effects, while others find that these effects are concentrated only within particular categories of firms or during specific market conditions. This heterogeneity suggests that the interaction between sentiment and earnings announcements may depend heavily on the informational environment surrounding the firm.

In the study developed by Baker and Wurgler (2006), sentiment effects are strongest in the firms that are difficult to value and difficult to arbitrage. This emphasizes the role of uncertainty and investor attention in this context. Firms with greater valuation uncertainty, lower analyst coverage, or higher informational frictions are generally considered more susceptible to sentiment-driven mispricing. In such settings, investors may rely more heavily on narratives, beliefs, or market mood when interpreting earnings information.

Seok et al. (2019) extend this literature by focusing more specifically on firm-level sentiment and market reactions to earnings news. Their results suggest that sentiment extracted from investor communication can influence short-term post-announcement returns, supporting the idea that behavioural factors contribute to the pricing of earnings information. However, the effects documented in the study remain concentrated in particular subsets of firms and are not always stable across specifications or return horizons.

This instability is an important characteristic of the literature. While the theoretical intuition linking sentiment to earnings reactions appears plausible, empirical evidence often depends strongly on the choice of sentiment proxy, sample period, event window, and firm characteristics. Some studies identify amplification effects only for positive surprises, while others document asymmetries between optimistic and pessimistic market environments. In many cases, the economic magnitude of the interaction effect also appears relatively modest.

A related strand of research examines whether investor attention rather than sentiment itself drives these patterns like Cookson et al. (2024). Increased discussion surrounding earnings announcements may simply reflect heightened investor focus on firms experiencing unusual events. Under this interpretation, abnormal trading activity and return dynamics are linked less to sentiment and more to the speed of information diffusion.

Distinguishing sentiment from attention becomes especially difficult in the context of social media. Online investor discussions often combine emotional reactions, speculative opinions, and rapid responses to breaking information. As a result, social-media-based sentiment measures may capture several overlapping mechanisms simultaneously.

Recent developments in textual analysis nevertheless allowed researchers to examine these dynamics at much higher frequency and granularity than earlier sentiment studies. Social media platforms such as Twitter, StockTwits, and Seeking Alpha provide continuous streams of investor communication that can be analyzed directly around earnings announcements.

This evolution expanded the scope of sentiment research, but it also introduced additional challenges. Compared to traditional sentiment proxies, social-media-based measures are often more volatile and potentially more sensitive to short-term noise. Moreover, because online discussions are largely driven by retail investors, it remains uncertain whether social-media sentiment possesses enough informational relevance to materially influence price formation around major corporate disclosures.

This issue is particularly relevant in earnings settings. Earnings announcements are closely monitored by analysts, institutional investors, and algorithmic trading systems capable of processing information almost immediately after release. The extent to which retail-driven social sentiment can alter the pricing of earnings surprises remains unclear.

One possible interpretation is that sentiment matters before the announcement by shaping investor expectations rather than during the immediate reaction itself. Under this view, sentiment may already be embedded in analyst forecasts, investor positioning, or pre-announcement pricing. If this is the case, the incremental impact of sentiment at the moment of the announcement may be limited once hard information becomes public.

Another possibility is that social-media sentiment influences only particular segments of the market, such as small firms, speculative firms, or stocks characterized by limited analyst coverage. In more informationally efficient firms, the role of sentiment may be substantially attenuated by rapid institutional processing of earnings news.

The present study contributes to this literature by examining whether pre-announcement social-media sentiment influences the market reaction to earnings surprises. Using an event-study framework combined with a multi-platform sentiment measure, the analysis focuses specifically on whether the sensitivity of returns to earnings news varies across different sentiment environments.

Data

The empirical analysis combines several datasets covering stock returns, earnings announcements, analyst forecasts, firm characteristics, and investor sentiment. The construction starts from a firm-day panel and is then reorganized around earnings announcement events.

Daily stock returns are obtained from the CRSP database. These data provide information on returns, trading volume, and shares outstanding for listed firms. Returns are expressed in excess form by subtracting the daily risk-free rate. In addition, the Fama-French three factors (market, SMB, and HML) are included in the analysis to control for systematic sources of return variation.

Firm-level identifiers and industry classifications are taken from Compustat. The SIC code is used to assign each firm to one of the Fama-French 12 industry groups. This step is mainly intended to capture broad sector differences in return behaviour and information processing, rather than to introduce firm-level accounting detail into the analysis.

Earnings announcement dates come from Compustat (RDQ), while analyst forecasts and realized earnings are drawn from the I/B/E/S database. To quantify the informational content of earnings releases, an earnings surprise measure is constructed as the difference between reported earnings and the consensus analyst forecast:

$$\text{Surprise}_i = (\text{Actual}_i - \text{Mean Forecast}_i) / \text{Price}_i$$

This variable reflects the extent to which realized performance departs from expectations already embedded in prices. To ensure comparability across firms with different price levels, the earnings surprise measure is normalized by stock price before entering the empirical analysis. The resulting variable is then standardized within the sample prior to estimation.

Investor sentiment is measured using the "Social Signal" dataset developed by Cookson et al. (2025). This dataset aggregates sentiment extracted from multiple investor-focused platforms, including Twitter, StockTwits, and Seeking Alpha. At the firm-day level, sentiment signals from these platforms are combined using principal component analysis, and the first principal component captures the common variation across platforms. As shown in the original study, this component reflects a shared sentiment signal rather than platform-specific noise .

The variable used in this study, $zee\_sent\_pc\_w$, corresponds to a winsorized version of this first principal component. Winsorization is applied to limit the impact of extreme observations, which are frequent in sentiment data. The resulting measure can be interpreted as a standardized indicator of overall investor tone at the firm-day level, based on a combination of social media sources.

To focus on pre-announcement conditions, the sentiment measure is transformed into a backward-looking average. For each firm-day, the average sentiment is computed over the previous 30 trading observations:

$$\text{Sentiment}_{i,t}^{30d} = \frac{1}{N_{i,t}} \sum_{k=1}^{30} \text{Sentiment}_{i,t-k} \, .$$

Only lagged values are included, so the measure does not incorporate any information from or after the earnings announcement. In addition, a minimum requirement of 20 valid observations within the 30-day window is imposed. This restriction ensures that the average reflects a stable sentiment environment rather than a small number of observations.

The dataset is initially constructed at the firm-day level by combining returns, sentiment, and firm characteristics. This panel is then reorganized around earnings announcements. Each firm-day observation is matched to the nearest future earnings announcement date. This matching ensures that all explanatory variables are aligned relative to the timing of the event. Event time is defined as:

$$\text{Event Time} = \text{Date} - \text{RDQ}.$$

Using this definition, an event window of $[-30, +30]$trading days is constructed around each announcement. Within this window, cumulative excess returns are computed over two distinct horizons. The first, spanning from one day before to three days after the announcement, captures the immediate market response. The second, covering days four to thirty after the announcement, is used to examine whether any delayed adjustment or price continuation takes place.

After constructing these windows, the dataset is collapsed to the event level. Each observation then corresponds to a single firm-event pair. Firms can appear multiple times in the dataset, as they may have several earnings announcements over the sample period. To account for this structure, standard errors are clustered at the firm level in the regression analysis.

The final sample reflects the intersection of several data sources. The availability of analyst forecasts from I/B/E/S restricts the sample to firms with sufficient coverage. The sentiment construction further reduces the number of observations, as it requires a minimum history of past sentiment data. Additional filters are applied to remove observations with missing returns or control variables. After these steps, the final regression sample consists of 3,265 firm-event observations over the period 2019 to 2021. The sentiment data are obtained from the Social Signal dataset developed by Cookson et al. (2025), which provides firm-day social media sentiment and attention indexes from 2012 through 2021. The dataset aggregates information from StockTwits, Twitter, and Seeking Alpha using standardized first principal components of sentiment and attention. Although the database covers a longer horizon, the empirical analysis focuses on the 2019–2021 period. This choice is motivated both by data consistency and by the growing importance of social media within financial markets during recent years. In particular, the COVID-19 period coincided with a substantial increase in retail investor participation, online trading activity, and finance-related social media engagement, potentially strengthening the informational relevance of social-media-based sentiment measures during this period.

These restrictions are largely driven by data availability rather than discretionary choices. In particular, firms with limited analyst coverage or low social media activity are less likely to be included. This implies that the final sample is tilted toward firms that are more visible to both analysts and retail investors. Due to the data density requirements of social media indexes, the final sample exhibits a descriptive concentration within the Fama-French 12 Healthcare sector; however, potential structural biases are absorbed through the explicit inclusion of industry fixed effects in our regression specification .

Two additional features of the dataset are worth noting. First, the sentiment measure is constructed using only past information, which limits concerns about forward-looking bias. Second, event windows are allowed to overlap across announcements. This means that a given firm-day may contribute to multiple event windows when announcements are close in time. While this introduces some dependence across observations, it is addressed by clustering standard errors at the firm level.

Taken together, the dataset combines market outcomes, analyst expectations, and investor sentiment in a consistent framework. This structure makes it possible to examine how pre-announcement sentiment is related to the way markets incorporate earnings news, while keeping a clear separation between information available before and after the announcement.

Methodology

The empirical strategy follows an event study framework designed to examine how stock prices react to earnings announcements and whether this reaction varies with the pre-announcement sentiment environment. The approach combines a time alignment of returns around discrete events with cross-sectional regressions at the event level.

Event Definition and Time Alignment

We define the event as a firm's earnings announcement date, denoted RDQ_i. Each observation in the final dataset corresponds to a firm-event pair. To align observations across firms, calendar time is transformed into event time:

$$\text{Event Time}_{i,t} = t - RDQ_i,$$

where $t$denotes the trading date. Negative values correspond to days prior to the announcement, while positive values capture post-announcement periods. This transformation allows all firms to be analyzed within a common temporal structure, independent of their specific announcement dates.

An event window of $[-30, +30]$trading days is constructed around each announcement. This window is sufficiently wide to capture both pre-announcement conditions and post-announcement adjustments, while remaining focused on the immediate informational environment of the event.

Measurement of Returns

The market response to earnings announcements is summarized using cumulative excess returns. Daily excess returns are defined as:

$$ret\_excess_{i,t} = R_{i,t} - rf_t,$$

where $R_{i,t}$is the firm's daily return and $rf_t$is the risk-free rate.

Cumulative returns are then constructed over two distinct intervals. The first captures the short-term reaction:

$$CAR_i[-1, +3] = \sum_{t=-1}^{+3} \left(R_{i,t} - rf_t\right),$$

This window includes the day prior to the announcement, the announcement day, and the following three trading days. While price discovery is often rapid, this specification allows for short-lived adjustments that may not be fully incorporated on the announcement day itself.

The second interval focuses on post-announcement dynamics:

$$CAR_i[+4, +30] = \sum_{t=+4}^{+30} \left(R_{i,t} - rf_t\right).$$

This window is used to examine whether any continuation or reversal occurs after the initial response. In particular, it provides a way to assess whether the market reaction is followed by delayed adjustment or correction.

Returns are measured as excess returns rather than model-based abnormal returns. Systematic risk is accounted for through the inclusion of Fama-French factors in the regression stage. While an alternative approach would estimate abnormal returns using a market model prior to aggregation, including factor controls directly in the regression provides an equivalent adjustment in this setting and avoids imposing a separate estimation step.

Empirical Specification

To assess whether sentiment influences the market response to earnings announcements, the following cross-sectional regression is estimated:

$$CAR_i = \beta_1\, Surprise_i + \beta_2\, Sentiment_i + \beta_3(Surprise_i \times Sentiment_i) + \gamma X_i + \alpha_{industry} + \alpha_{year} + \varepsilon_i,$$

The dependent variable CAR_icorresponds either to the short-term window $[-1, +3]$or the post-announcement window $[+4, +30]$.

The variable $Surprise_i$ captures the earnings shock, measured as the difference between actual earnings and the consensus analyst forecast from I/B/E/S, normalized by the firm's stock price. The resulting measure is standardized within the sample before entering the regression analysis.

The variable $Sentiment_i$corresponds to the 30-day pre-announcement average sentiment measure. By construction, it reflects the information environment leading up to the announcement and excludes any contemporaneous or post-event information.

The interaction term $Surprise_i \times Sentiment_i$is the central variable of interest. It captures whether the sensitivity of returns to earnings news varies with the level of sentiment. In this framework, the coefficient $\beta_3$reflects how the Earnings Response Coefficient (ERC) changes as a function of sentiment.

This specification does not test whether sentiment directly drives returns. Instead, it evaluates whether sentiment is associated with a differential response to earnings information. A significant interaction term would indicate that identical earnings surprises are priced differently depending on the prevailing sentiment environment.

Control Variables and Fixed Effects

The vector $X_i$includes firm-level controls intended to capture variation in size and liquidity. Firm size is measured as the logarithm of market equity (ln ME), while liquidity is proxied by the logarithm of trading volume (ln VOL).

The regression also includes the Fama-French three factors (market, SMB, and HML). These factors control for systematic variation in returns that may be correlated with earnings announcements, ensuring that the estimated coefficients reflect firm-specific responses rather than broader market movements.

Industry fixed effects are included using the Fama-French 12 classification. These fixed effects absorb persistent differences across sectors, such as variation in investor attention, information disclosure practices, or typical earnings behaviour.

Year fixed effects are also incorporated to control for time-specific shocks. Given the relatively short sample period, these effects capture macroeconomic conditions, changes in market volatility, and other aggregate influences that may affect both sentiment and returns.

Firm fixed effects are not included in the baseline specification. The main reason is that the analysis relies on cross-sectional variation in both sentiment and earnings surprise. Including firm fixed effects would absorb a substantial portion of this variation, particularly for firms with relatively stable sentiment patterns. Industry fixed effects therefore provide a compromise, controlling for broad sector differences while preserving sufficient variation for identification.

Estimation and Inference

The regression model is estimated using ordinary least squares (OLS). Because firms may appear multiple times in the dataset due to repeated earnings announcements, observations are not independent within firms. To account for this, standard errors are clustered at the firm level.

Clustering allows for arbitrary correlation of the error term within firms across events. In practice, this is important in an event study setting where firm-specific characteristics may persist over time and affect multiple observations.

Interpretation and Identification

The empirical design is intended to capture statistical relationships rather than causal effects. In particular, the interaction coefficient $\beta_3$should be interpreted as measuring whether the response of returns to earnings surprises is associated with the sentiment environment, rather than implying that sentiment causally amplifies or dampens this response.

Although the inclusion of control variables and fixed effects mitigates concerns related to omitted variables, it does not eliminate them entirely. For example, sentiment may be correlated with unobserved factors such as firm visibility, investor attention, or the broader information environment. As a result, the findings should be interpreted as conditional associations.

Robustness Strategy

To assess the stability of the results, alternative specifications are considered. These include regressions excluding Fama-French factors and regressions excluding firm-level controls. The objective is to verify that the main findings are not driven by a specific set of controls or modeling choices. Additional robustness checks and heterogeneity analyses are presented in the results section.

The methodology combines an event study framework with cross-sectional regressions to examine how earnings information is incorporated into stock prices. By focusing on the interaction between sentiment and earnings surprise, the analysis tests whether the market response to fundamental news varies with the surrounding sentiment environment.

Results

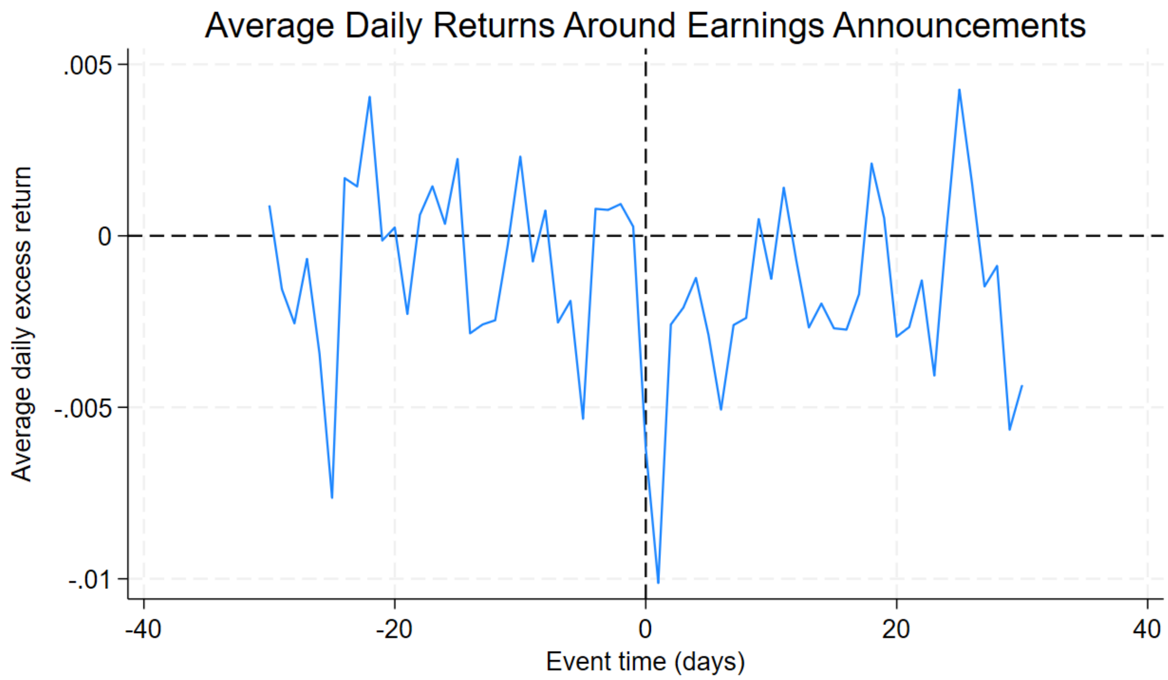
Overview and Empirical Motivation

The empirical analysis focuses on a specific question: whether the market response to surprise in earnings announcements depends on the sentiment environment preceding the event. While a large body of literature documents that sentiment can predict returns in general settings, it is less clear whether sentiment plays a role when information arrives in a discrete and observable form, such as in surprise in earnings announcements.

The underlying intuition motivating this analysis is straightforward. Earnings announcements represent moments of uncertainty resolution. When firms release earnings, investors receive new information that was not fully anticipated. In such contexts, one might expect behavioural factors, including sentiment, to play a stronger role. If investors are surprised, their interpretation of the news could depend on their prior beliefs or mood, potentially amplifying or dampening the price reaction.

The empirical results do not support this intuition. Across all specifications, earnings surprises generate a clear and statistically significant response in returns, but this response does not appear to vary with the sentiment environment. The remainder of this section develops this result in detail.

Figure 1. Average Daily Excess Returns Around Earnings Announcements



Notes: This figure displays average daily excess returns around earnings announcement dates over the event window [-30,+30]. The vertical dashed line indicates the announcement date (t = 0). The figure illustrates the short-term return dynamics surrounding earnings announcements.

Figure 1 shows that average daily excess returns fluctuate around the earnings announcement date, with a noticeable movement around the event window. The pattern is consistent with the use of an event-study framework, although the figure also indicates that return dynamics are noisy and not concentrated entirely on the announcement day.

Descriptive Statistics and Data Properties

Table 1. Summary Statistics

	mean	sd	min	p25	p50	p75	max	count
Event-day return	-0.004	0.084	-0.672	-0.038	-0.004	0.026	1.016	4334
CAR [-1,+3]	-0.021	0.125	-0.557	-0.089	-0.023	0.041	0.636	4334
CAR [+4,+30]	-0.027	0.223	-1.024	-0.139	-0.012	0.074	1.285	4334
Earnings surprise	0.000	1.000	-5.745	0.063	0.075	0.081	4.621	3462
Sentiment (30-day avg.)	0.044	0.436	-1.789	-0.166	0.110	0.344	1.231	4108
Log market equity	14.355	3.177	7.060	11.624	14.147	17.135	21.641	4334
Log volume	15.134	1.644	0.000	14.087	15.249	16.276	19.581	4334
Observations	3265							

Notes: This table reports summary statistics for the main variables used in the empirical analysis. Event-day return corresponds to the stock return on the earnings announcement date. CAR [-1,+3] and CAR [+4,+30] represent cumulative abnormal returns over the short and post-announcement windows, respectively. Earnings surprise is normalized by stock price and standardized within the sample. Sentiment is measured as the 30-day average of the social-media-based sentiment index.

Table 1 provides summary statistics for the main variables used in the analysis. Several aspects of the data are worth highlighting, as they directly relate to the empirical design.

First, return measures display substantial dispersion, which is typical in event studies focusing on earnings announcements. For the short window $[-1, +3]$, cumulative returns have a mean of approximately -2.15% and a standard deviation of 12.48%. The distribution is relatively wide, with the 1st percentile at -32.9% and the 99th percentile at 35.7%. These values indicate that some announcements are associated with large price movements, consistent with the idea that earnings can contain significant new information.

The longer window $[+4, +30]$exhibits even greater dispersion, with a standard deviation of 22.31%. The range is also wider, extending from approximately -102% to +128%. These extreme values suggest that the post-announcement period captures a broader set of dynamics, including potential continuation, reversal, or firm-specific shocks unrelated to the announcement itself.

Event-day returns display similar features. The standard deviation is approximately 8.36%, and the distribution is highly skewed (skewness = 1.77) with very high kurtosis (26.8). This indicates that extreme observations are present and may influence estimation. This is a common feature of financial return data and reinforces the importance of using robust standard errors.

Second, the earnings surprise variable is normalized by stock price and standardized before entering the regression analysis. As shown in Table 1, the standardized measure has a mean close to zero and a standard deviation of one, with values ranging from -5.745 to 4.621. This indicates that the variable preserves meaningful cross-sectional variation while remaining comparable across firms.

Third, the sentiment measure displays moderate variation but is considerably smoother than raw daily sentiment. This is a direct consequence of the 30-day averaging procedure. In earlier versions of the analysis, sentiment was measured using only the lagged value (t−1). That specification resulted in highly volatile estimates and unstable coefficients, reflecting the noisy nature of daily sentiment data. The shift to a 30-day average was therefore motivated by the need to capture a more stable measure of the information environment. The descriptive statistics confirm that this smoothing reduces extreme values while preserving meaningful variation.

Finally, firm characteristics such as size and trading volume exhibit a wide range. The median log market equity is approximately 14.15, with values ranging from 7.06 to 21.64. This indicates that the sample includes both relatively small and very large firms, although the distribution is somewhat tilted toward larger firms. This feature will be relevant when interpreting the results, as large firms tend to have more analyst coverage and more efficient information processing.

Overall, the descriptive statistics suggest that the dataset is consistent with standard event study applications, while also highlighting potential sources of noise and heterogeneity that must be taken into account in the analysis.

Correlation Analysis and Preliminary Evidence

Table 2. Correlation Matrix

	RET_adj	ret_short	ret_post	Earnings surprise	sent_30d	lnME	lnVOL
RET_adj	1	.4992052	.027834	.0303749	.0019096	.0671724	.0882808
ret_short	.4992052	1	.0716755	.0569021	.0059459	.1297086	.0940449
ret_post	.027834	.0716755	1	-.0019548	.040427	-.0188214	-.0191382
Earnings surprise	.0303749	.0569021	-.0019548	1	-.029289	.1047829	.0451409
sent_30d	.0019096	.0059459	.040427	-.029289	1	-.0944105	-.0196655
lnME	.0671724	.1297086	-.0188214	.1047829	-.0944105	1	.5739627
lnVOL	.0882808	.0940449	-.0191382	.0451409	-.0196655	.5739627	1

Notes: This table presents pairwise correlations between the main variables used in the regression analysis. The relatively low correlations between sentiment, earnings surprise, and the interaction term suggest that multicollinearity is not a major concern in the baseline specification.

Table 2 reports pairwise correlations between the main variables. While these correlations do not establish causal relationships, they provide useful information about the structure of the data and potential econometric concerns.

The correlation between event-day returns and short-window cumulative returns is relatively high (approximately 0.50). This is expected, as a substantial portion of the total reaction occurs on the announcement day itself. In contrast, the correlation between short-window returns and post-announcement returns is much lower (approximately 0.07), indicating that the longer window captures distinct dynamics rather than simply extending the immediate reaction.

The correlation between earnings surprise and returns is positive but modest, which is consistent with the regression results reported later. This suggests that while earnings surprises are an important driver of returns, they do not fully explain cross-sectional variation.

A key result concerns the relationship between sentiment and earnings surprise. The correlation between these variables is effectively zero (-0.012), indicating that sentiment does not simply proxy for earnings news. This is important for identification, as it suggests that the interaction term is not mechanically driven by overlap between the two variables.

The interaction term itself shows a moderate correlation with earnings surprise (0.197), which is expected given its construction. However, this does not appear to generate multicollinearity concerns. Variance Inflation Factors (VIFs) are all low, with values close to 1 for the main variables (1.06 for surprise, 1.13 for sentiment, and 1.05 for the interaction). This indicates that the regression coefficients are estimated with sufficient precision and that the lack of significance for the interaction term is not driven by inflated standard errors.

Taken together, the correlation analysis provides reassurance that the main variables are not mechanically linked and that the regression results are unlikely to be driven by simple co-movements or multicollinearity.

Main Regression Results

Table 3. Baseline Regression Results

	(1) Event return	(2) CAR [-1,+3]	(3) CAR [+4,+30]
Earnings surprise (std.)	0.00111	0.00589***	0.00503
	(0.00169)	(0.00191)	(0.00512)
Sentiment (30-day avg.)	0.000859	-0.00280	-0.00218
	(0.00310)	(0.00499)	(0.00797)
Surprise x Sentiment	0.0000345	-0.00228	-0.0254*
	(0.00490)	(0.00563)	(0.0143)
Log market equity	0.000779	0.00326***	-0.00398**
	(0.000831)	(0.000953)	(0.00154)
Log volume	0.00526***	0.00246	-0.00308
	(0.00202)	(0.00197)	(0.00315)
Market factor	0.0114***	0.00948***	0.00981***
	(0.00110)	(0.00188)	(0.00296)
SMB	0.0105***	0.0141***	-0.0116**
	(0.00185)	(0.00329)	(0.00457)
HML	-0.00354***	-0.00751***	0.00367
	(0.00112)	(0.00203)	(0.00298)
Observations	3265	3265	3265

Notes: Standard errors in parentheses
All regressions include industry (FF12) and year fixed effects.
Standard errors are clustered at the firm level.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
This table reports the baseline regression results examining the relationship between earnings surprises, investor sentiment, and stock returns around earnings announcements. Columns (1)–(3) correspond respectively to event-day returns, short-window cumulative abnormal returns, and post-announcement cumulative abnormal returns. All regressions include industry (FF12) and year fixed effects. Standard errors are clustered at the firm level.

Table 3 presents the main regression results for the short window $[-1, +3]$. The specification includes earnings surprise, sentiment, their interaction, firm-level controls, Fama-French factors, and fixed effects.

The coefficient on earnings surprise is positive and statistically significant in the short event window, with an estimate of 0.00589. This indicates that a one-standard-deviation increase in earnings surprise is associated with an increase of approximately 0.59 percentage points in cumulative returns over the [-1,+3] window. This result is consistent with standard findings in the literature and confirms that earnings announcements contain value-relevant information.

The sentiment variable, by contrast, is not statistically significant. The estimated coefficient in the short-window specification is -0.00280. This suggests that, once earnings surprise and other controls are included, pre-announcement sentiment does not have a direct effect on short window returns in this setting.

The central result concerns the interaction term between sentiment and earnings surprise. In the short-window specification, the estimated coefficient is -0.00228, with a standard error of 0.00563. The coefficient is statistically insignificant, indicating that the market response to earnings surprises does not appear to vary meaningfully with the pre-announcement sentiment environment.

The magnitude of the coefficient is also small relative to its variability. This suggests that even if an interaction effect exists, it is likely to be economically limited in the short window specification.

It is important to interpret this result carefully. The absence of statistical significance does not imply that the true effect is exactly zero. However, the combination of a small coefficient and a low t-statistic suggests that the data do not provide meaningful evidence of an interaction effect.

Statistical Power and Interpretation of the Null Result

Given the absence of a statistically significant interaction effect, it is necessary to assess whether this result reflects a true absence of relationship or simply a lack of statistical power.

The regression is based on 3,265 firm-event observations, with standard errors clustered at the firm level (703 clusters). The coefficient on the interaction term is -0.00228, with a standard error of 0.00563, yielding a t-statistic of -0.41.

The magnitude of the t-statistic is informative. A value close to zero indicates that the estimated coefficient is small relative to its variability, rather than simply imprecisely estimated. If the true effect were economically meaningful but difficult to detect, one would expect a t-statistic closer to 1.5 or 2. In contrast, the observed value suggests that the estimated effect is both statistically and economically limited.

This does not imply that sentiment has no effect in general. Rather, it indicates that within this empirical setting—earnings announcements combined with a pre-event sentiment measure—there is no detectable evidence that sentiment modifies the sensitivity of returns to earnings surprises.

At the same time, the relatively modest sample size implies that smaller effects may go undetected. The intersection of CRSP, I/B/E/S, and sentiment data imposes natural constraints on the number of usable observations. As a result, the findings should be interpreted as evidence of no strong interaction effect, rather than definitive proof of absence.

Robustness Analysis

Table 4. Robustness Checks

	(1) Full	(2) Without FF factors	(3) No controls
Earnings surprise (std.)	0.00589***	0.00640***	0.00682***
	(0.00191)	(0.00193)	(0.00191)
Sentiment (30-day avg.)	-0.00280	-0.00497	-0.00311
	(0.00499)	(0.00504)	(0.00498)
Surprise × Sentiment	-0.00228	-0.00314	-0.00134
	(0.00563)	(0.00599)	(0.00567)
Log market equity	0.00326***	0.00324***	
	(0.000953)	(0.000960)	
Log volume	0.00246	0.00177	
	(0.00197)	(0.00199)	
Market factor	0.00948***		0.00915***
	(0.00188)		(0.00194)
SMB	0.0141***		0.0143***
	(0.00329)		(0.00336)
HML	-0.00751***		-0.00746***
	(0.00203)		(0.00203)
Observations	3265	3265	3265

Notes: Standard errors in parentheses
Standard errors are clustered at the firm level.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
This table presents robustness checks for the baseline specification. Alternative regressions are estimated by excluding Fama-French risk factors or firm-level control variables. The interaction between sentiment and earnings surprise remains statistically insignificant across specifications. Standard errors are clustered at the firm level.

To evaluate whether the main findings depend on specific modeling choices, a series of robustness checks are conducted. These involve re-estimating the baseline regression under alternative specifications, including models that exclude Fama-French factors and models that omit firm-level controls.

Across all specifications, the coefficient on earnings surprise remains positive and statistically significant, with only minor variation in magnitude. This reinforces the conclusion that earnings news is the primary driver of returns in the short window.

The interaction term between sentiment and earnings surprise remains statistically insignificant in all cases. Its magnitude also remains stable, with coefficients consistently close to zero. This suggests that the absence of an interaction effect is not driven by the inclusion of particular control variables or by the way risk is accounted for.

The stability of the interaction coefficient is particularly important. If the result were sensitive to specification choices, one could argue that the null finding reflects model misspecification. Instead, the evidence indicates that the result is robust across a range of reasonable alternatives.

Heterogeneity by Firm Size

Table 5. Heterogeneity by Firm Size

	(1) Small firms	(2) Large firms
Earnings surprise (std.)	0.00519***	0.0195
	(0.00198)	(0.0301)
Sentiment (30-day avg.)	0.00169	-0.00303
	(0.00984)	(0.00793)
Surprise × Sentiment	-0.00239	0.00531
	(0.00583)	(0.0686)
Log market equity	0.0110***	0.000332
	(0.00355)	(0.00147)
Log volume	0.00119	0.000992
	(0.00351)	(0.00223)
Market factor	0.00939***	0.0104***
	(0.00269)	(0.00185)
SMB	0.0221***	0.00609*
	(0.00591)	(0.00338)
HML	-0.00756**	-0.00662***
	(0.00376)	(0.00203)
Observations	1310	1955

Notes: Standard errors in parentheses
Standard errors are clustered at the firm level.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
This table reports heterogeneity analyses across firm size groups defined using the median value of log market equity. The results suggest that earnings surprises remain significant across subsamples, while the interaction between sentiment and earnings surprises remains statistically insignificant. Standard errors are clustered at the firm level.

The next step is to examine whether the interaction effect emerges in specific segments of the sample. Firms are divided into two groups based on the median level of market equity, and the regression is estimated separately for each group.

Earnings surprise remains a significant predictor of returns in both subsamples. Interestingly, the magnitude of the coefficient is somewhat larger for larger firms than for smaller firms. This differs from the typical expectation that smaller firms exhibit stronger reactions due to higher information uncertainty. One possible explanation is that analyst coverage and forecast accuracy differ across firms, affecting how surprises are measured and interpreted.

More importantly, the interaction term remains statistically insignificant in both groups. The estimated coefficients are small and imprecisely estimated, with no clear pattern across firm sizes. This suggests that the absence of an interaction effect is not driven by aggregation across heterogeneous firms.

The heterogeneity analysis therefore reinforces the main result. Even in environments where informational frictions are expected to be stronger, sentiment does not appear to modify the market response to earnings announcements.

Role of Sentiment Measurement

An important aspect of the analysis concerns the construction of the sentiment variable. The use of a 30-day average reflects a deliberate attempt to reduce the noise inherent in daily sentiment measures.

In earlier specifications, sentiment was measured using only the lagged value ($t-1$). This approach resulted in highly volatile estimates and unstable coefficients. The lack of precision in those results motivated the shift toward a smoothed measure. The descriptive statistics confirm that the 30-day average produces a more stable distribution.

However, this choice also introduces a trade-off. While smoothing reduces noise, it may also attenuate short-term variation in sentiment that could be relevant for immediate reactions. In other words, the sentiment measure captures a broader environment rather than high-frequency shifts in investor mood.

This raises the possibility that the null interaction result is partly driven by measurement limitations. If sentiment affects the interpretation of earnings news at a very short horizon, a smoothed measure may fail to capture this effect. Conversely, if sentiment reflects broader market narratives, it may not align closely with firm-specific earnings expectations.

Interpretation and Economic Implications

The central finding of the analysis is that earnings surprises generate a statistically significant price response, but this response does not vary with the sentiment environment.

This result is somewhat at odds with the initial intuition. One might expect that when investors face unexpected information, their reactions would be influenced by prior beliefs or mood. In particular, sentiment could act as a lens through which earnings news is interpreted, leading to stronger or weaker reactions depending on whether sentiment is positive or negative.

The empirical evidence does not support this mechanism. Instead, the results suggest that earnings announcements are processed in a way that is largely independent of the sentiment environment.

One possible interpretation is that earnings announcements provide information that is sufficiently precise and salient to dominate softer signals. When firms release earnings, the information is widely disseminated, closely monitored, and quickly incorporated into prices. In such a setting, the role of sentiment may be limited.

Another possibility is that the sentiment measure, while capturing general investor tone, does not reflect firm-specific expectations about earnings. Because it aggregates signals from multiple platforms and users, it may capture broad market sentiment rather than beliefs about a particular firm's performance. This possibility is especially relevant because the sentiment variable is constructed around open source and non-professional investors platform, which don't affect institutional investors.

It is also worth noting that the absence of an interaction effect does not imply that sentiment is irrelevant in financial markets. Rather, it suggests that its role may be context-dependent. Sentiment may matter more in settings where information is ambiguous, less frequent, or more difficult to interpret. In contrast, earnings announcements represent structured and widely anticipated events, where the scope for sentiment-driven interpretation may be more limited.

Post-Announcement Dynamics

The analysis of the $[+4, +30]$window provides additional insight. The coefficient on earnings surprise is not statistically significant in this window, and there is no evidence of a systematic continuation or reversal pattern.

This suggests that the market incorporates earnings information relatively quickly, leaving limited scope for delayed adjustment. In particular, the absence of post-announcement drift indicates that the initial reaction captures most of the information contained in the announcement.

The interaction term is weakly significant in the post-announcement window, with a negative coefficient. However, this result is only significant at the 10% level and does not appear consistently across the short-window or robustness specifications. It should therefore be interpreted cautiously rather than as strong evidence of sentiment-driven post-announcement drift.

Findings statement

Across all specifications, the results point to a consistent pattern. Earnings surprises are a significant driver of short-term returns, but there is no evidence that sentiment modifies this relationship. The interaction effect is small and statistically insignificant in the main short-window specification and remains insignificant across robustness checks and firm-size subsamples.

These findings suggest that, in the context of earnings announcements, the market response to fundamental information does not depend on the broader sentiment environment. While sentiment may influence returns in other settings, its role appears limited when information arrives in a structured and widely observed form and in an institutional context.

Limitations

While the empirical design is structured to examine the relationship between sentiment and the market response to earnings announcements, several limitations must be acknowledged. These relate to measurement, sample construction, return specification, and interpretation of the results.

Measurement of Sentiment

A first limitation concerns the construction of the sentiment variable. The measure used in this study is derived from the Social Signal dataset and aggregates sentiment across multiple social media platforms, including Twitter, StockTwits, and Seeking Alpha. While this approach provides a broad and diversified proxy for investor tone, it also introduces several constraints.

First, sentiment extracted from social media is inherently noisy. Individual messages may reflect speculation, disagreement, or non-informational content. Although the underlying dataset applies aggregation and filtering techniques, the resulting variable remains an imperfect proxy for investor beliefs.

Second, the sentiment measure is not firm-specific in a strict sense. Because it combines signals across platforms and users, it captures a general tone surrounding a firm rather than precise expectations about its earnings. As a result, the measure may not align closely with the type of information that is relevant for interpreting earnings announcements.

Third, the use of a 30-day backward-looking average reflects a deliberate modeling choice. Rather than capturing high-frequency fluctuations in sentiment, the objective is to measure a persistent sentiment environment. Earlier specifications based on one-day lagged sentiment produced unstable and noisy estimates, which motivated the adoption of a longer window. This smoothing reduces volatility but may attenuate short-term sentiment effects. In that sense, the analysis focuses on whether a sustained sentiment environment influences price reactions, rather than whether daily mood fluctuations do so.

Sample Selection and Visibility Bias

The final sample consists of 3,265 firm-event observations, reflecting the intersection of CRSP returns, I/B/E/S analyst forecasts, and sentiment data. This imposes natural restrictions on the sample.

Firms without analyst coverage or with limited social media activity are excluded. As a result, the sample is tilted toward firms that are more visible, more actively traded, and more frequently discussed by investors. These firms are also more likely to be followed by institutional investors and subject to efficient pricing mechanisms.

This feature has important implications. If sentiment effects are strongest in environments where information is scarce or difficult to process, the focus on highly visible firms may reduce the likelihood of detecting such effects. In this sense, the estimated interaction effect can be interpreted as a lower bound on the influence of sentiment in less-covered segments of the market.

Measurement of Returns

Returns are measured as excess returns rather than model-based abnormal returns. Systematic risk is accounted for through the inclusion of Fama-French factors directly in the regression specification. This corresponds to a one-step factor-adjusted approach, in which risk adjustment is handled within the estimation rather than through a separate pre-event model.

While this approach is widely used and avoids additional estimation noise, it differs from the traditional two-step event study methodology. In particular, it does not explicitly estimate firm-specific betas over a pre-event window. As a result, risk adjustment is performed at the regression stage rather than in the construction of the return measure.

This represents a trade-off between simplicity and precision. The inclusion of factor controls mitigates concerns related to systematic risk, but it does not fully replicate a classical market-model-based abnormal return calculation.

Statistical Power and Interpretation of the Interaction

A central result of the analysis is the absence of a statistically significant interaction between sentiment and earnings surprise. The interaction term in the short-window have a coefficient of -0.00228, with a standard error of 0.00563 and a t-statistic of -0.41.

A t-statistic of this magnitude indicates that the estimated effect is both statistically insignificant and economically small relative to its variability. The coefficient is less than one standard error away from zero, implying that it is statistically indistinguishable from zero in this sample.

This finding should be interpreted carefully. On one hand, the lack of significance suggests that there is no strong evidence that sentiment modifies the market response to earnings news. On the other hand, the sample size—while typical for event-level studies—remains limited compared to large panel analyses using high-frequency sentiment data.

The results therefore indicate the absence of a detectable interaction effect, rather than definitive proof that no such effect exists. However, given the small magnitude of the estimated coefficient, any potential effect is likely to be economically limited in this setting.

Identification and Omitted Variables

The empirical framework relies on ordinary least squares estimation and identifies statistical associations rather than causal relationships. Although the inclusion of control variables and fixed effects reduces concerns related to omitted variables, it does not eliminate them.

In particular, sentiment may be correlated with unobserved factors such as investor attention, information quality, or firm visibility. These factors could influence both sentiment and returns, potentially confounding the estimated relationships.

As a result, the findings should be interpreted as conditional correlations within the data rather than as evidence of a causal mechanism linking sentiment to market reactions.

Role of Investor Type

An additional limitation concerns the nature of the sentiment data. Social media platforms are dominated by retail investors, while earnings announcements are primarily processed by institutional investors and algorithmic trading systems.

This mismatch may limit the ability of the sentiment measure to influence prices at the time of the announcement. If institutional investors incorporate earnings information rapidly and efficiently, retail sentiment may have little impact on the immediate price response, even if it reflects genuine beliefs or expectations.

The underlying sentiment index is constructed entirely from decentralized, open-source, and non-professional digital platforms, specifically Twitter, StockTwits, and Seeking Alpha, which introduces an inherent degree of retail noise into the signal . Because these platforms are open to the public, the conversational data they generate reflect a highly heterogeneous mix of investor sophistication and trading motives :

- Twitter (X): As a general-interest microblogging network, Twitter captures a broad but highly unfiltered stream of public discourse. Financial text on this platform is heavily exposed to high-frequency speculative noise, sudden sentiment cascades, and artificial inflation from automated bot networks or viral retail trends.
- StockTwits: While domain-specific and organized strictly around asset ticker tags, StockTwits operates as an open-access forum populated overwhelmingly by retail day-traders and short-term momentum speculators . The platform's crowd dynamics capture immediate emotional mood swings and speculative retail positioning rather than structured expectations of future corporate performance.
- Seeking Alpha: As a crowdsourced equity research platform, Seeking Alpha provides long-form investment theses that are more structurally sophisticated than microblogging posts. However, it remains an open platform written by independent, freelance contributors whose professional credentials, analytical quality, and valuation frameworks are entirely unregulated.

This consideration provides a potential explanation for the absence of an interaction effect, particularly in the short window surrounding the announcement.

Timing of Sentiment Effects

Finally, the analysis focuses on whether sentiment affects the reaction to earnings announcements. However, sentiment may play a role at an earlier stage.

In particular, sentiment may influence how expectations are formed prior to the announcement rather than how information is processed at the time of the event. If sentiment is already incorporated into analyst forecasts or investor expectations, its effect would not appear in the interaction with realized surprises.

This interpretation is consistent with the empirical findings. The absence of an interaction effect does not imply that sentiment is irrelevant, but rather that its influence may operate upstream, before the earnings announcement occurs.

Overall, the analysis is subject to limitations related to measurement, sample construction, and empirical design. These limitations do not invalidate the results but constrain their interpretation. In particular, the findings indicate that there is no evidence of a meaningful interaction between sentiment and earnings surprise within the observed sample, while leaving open the possibility that sentiment may operate through other channels or in different market settings.

Conclusion

The growing importance of social media in financial markets has substantially expanded the study of investor sentiment and behavioural finance. Online platforms now allow investors to react to information continuously, express expectations publicly, and influence market discussions in real time. As a result, sentiment extracted from social media has increasingly been viewed as a potential determinant of stock returns and market dynamics. At the same time, earnings announcements remain among the most important and closely monitored information events in financial markets, making them a natural setting in which to examine whether investor sentiment affects the incorporation of new information into prices.

This dissertation investigated whether pre-announcement investor sentiment influences stock returns around earnings announcements and whether it modifies the market response to earnings surprises. Using a firm-day panel dataset combining stock returns, earnings announcement dates, analyst-based earnings surprises, and social-media-derived sentiment measures, the analysis employed an event-study framework to examine return dynamics before, during, and after earnings announcements. The empirical strategy focused on both short announcement windows and post-announcement periods in order to distinguish immediate reactions from delayed adjustments.

The results provide several important findings. First, earnings surprises significantly predict short-term stock returns around announcement periods, confirming the central role of earnings information in price formation. Firms reporting stronger-than-expected earnings generally experience positive abnormal returns, while negative surprises are associated with weaker performance. This result is consistent with the broader earnings announcement literature and supports the informational relevance of analyst expectations.

Second, the analysis does not provide robust evidence that investor sentiment significantly modifies the relationship between earnings surprises and stock returns. The interaction term between sentiment and earnings surprise remains statistically insignificant across specifications, subsamples, and robustness checks. Although the original motivation of the study was that sentiment might amplify investor reactions when unexpected information arrives, the empirical evidence suggests that this effect is either limited or difficult to detect within the present setting.

Importantly, the absence of a strong interaction effect does not imply that sentiment is irrelevant in financial markets more generally. Instead, the findings suggest that highly structured information events such as earnings announcements may remain largely dominated by fundamentals and rapid information incorporation. Earnings releases provide precise and widely disseminated information that is immediately processed by analysts, institutional investors, and algorithmic trading systems. Under these conditions, the influence of broader social-media-driven sentiment may be substantially attenuated.

At the same time, the results also indicate that sentiment may still contribute to short-term return dynamics surrounding earnings events. This interpretation is broadly consistent with behavioural finance theories emphasizing limits to arbitrage and heterogeneous information processing.

The findings also contribute to the broader debate regarding market efficiency and behavioural finance. Much of the sentiment literature argues that investor psychology can generate temporary deviations from fundamental value, particularly in environments characterized by uncertainty or speculative behaviour. The present study does not reject this possibility entirely, but it suggests that the role of sentiment may be more context-dependent than some behavioural explanations imply. In earnings announcement settings, hard information appears to dominate softer behavioural signals once the announcement becomes public.

Methodologically, this dissertation contributes to the literature by combining an event-study framework with a multi-platform social media sentiment measure derived from recent advances in textual analysis. Unlike traditional sentiment proxies based on surveys or aggregate market variables, the sentiment measure used in this study reflects direct investor communication extracted from online platforms. This allows for a more granular examination of investor tone surrounding firm-specific events. In addition, the distinction between immediate reaction windows and longer post-announcement periods provides further insight into the timing of market adjustment around earnings releases.

Several limitations nevertheless remain. Social-media-based sentiment measures are inherently noisy and may partly reflect investor attention or speculative activity rather than pure sentiment. In addition, the sample is restricted to firms with sufficient analyst coverage and social media activity, which may bias the analysis toward larger and more visible firms. The empirical design also identifies statistical associations rather than causal relationships. These limitations imply that the findings should be interpreted cautiously and primarily as evidence regarding the presence or absence of detectable interaction effects within the observed sample.

Future research could extend this analysis in several directions. One possibility would be to examine intraday reactions around earnings announcements using higher-frequency data, which may capture very short-lived sentiment effects that daily returns fail to identify. Another extension would involve comparing retail-driven sentiment measures with institutional investor expectations in order to better distinguish different sources of market reaction. Additional work could also explore whether sentiment effects become stronger during periods of market stress, extreme uncertainty, or speculative episodes where behavioural dynamics are likely to dominate.

Overall, this dissertation suggests that while social-media-based sentiment contains meaningful information about investor behaviour, its ability to modify the market response to earnings announcements appears limited within the present empirical setting. The findings therefore support a more nuanced interpretation of behavioural finance in which sentiment may influence certain dimensions of market activity without fundamentally overriding the role of earnings information in price formation.

References

Antweiler, W., & Frank, M. Z. (2004). Is All That Talk Just Noise? The Information Content of Internet Stock Message Boards. *The Journal of Finance*, *59*(3), 1259-1294. https://doi.org/10.1111/j.1540-6261.2004.00662.x

Baker, M., & Wurgler, J. (2006). Investor Sentiment and the Cross-Section of Stock Returns. *The Journal of Finance*, *61*(4), 1645-1680. https://doi.org/10.1111/j.1540-6261.2006.00885.x

Bird, R., & Yeung, D. (2012). How do investors react under uncertainty? *Pacific-Basin Finance Journal*, *20*(2), 310-327. https://doi.org/10.1016/j.pacfin.2011.10.001

Bouteska, A., Sharif, T., & Abedin, M. Z. (2024). Does investor sentiment create value for asset pricing? An empirical investigation of the KOSPI-listed firms. *International Journal of Finance & Economics*, *29*(3), 3487-3509. https://doi.org/10.1002/ijfe.2836

Chen, H., De, P., Hu, Y. (Jeffrey), & Hwang, B.-H. (2014). Wisdom of Crowds : The Value of Stock Opinions Transmitted Through Social Media. *The Review of Financial Studies*, *27*(5), 1367-1403. https://doi.org/10.1093/rfs/hhu001

Cookson, J. A., Lu, R., Mullins, W., & Niessner, M. (2025). *Market Signals from Social Media* (SSRN Scholarly Paper Nº 5187350). Social Science Research Network. https://doi.org/10.2139/ssrn.5187350

Daniel, K., Hirshleifer, D., & Subrahmanyam, A. (1998). Investor Psychology and Security Market Under- and Overreactions. *The Journal of Finance*, *53*(6), 1839-1885. https://doi.org/10.1111/0022-1082.00077

Hayn, C. (1995). The information content of losses. *Journal of Accounting and Economics*, *20*(2), 125-153. https://doi.org/10.1016/0165-4101(95)00397-2

Kim-Hahm, H., Abou-Zaid, A. S., & Mohd, A. (2025). News vs. Social Media : Sentiment Impact on Stock Performance of Big Tech Companies. *Journal of Risk and Financial Management*, *18*(12), 660. https://doi.org/10.3390/jrfm18120660

Livnat, J., & Mendenhall, R. R. (2006). Comparing the Post–Earnings Announcement Drift for Surprises Calculated from Analyst and Time Series Forecasts. *Journal of Accounting Research*, *44*(1), 177-205. https://doi.org/10.1111/j.1475-679X.2006.00196.x

Mian, G. M., & Sankaraguruswamy, S. (2012). Investor Sentiment and Stock Market Response to Earnings News. *The Accounting Review*, *87*(4), 1357-1384. https://doi.org/10.2308/accr-50158

Peress, J. (2008). *Media Coverage and Investors' Attention to Earnings Announcements* (SSRN Scholarly Paper Nº 1106475). Social Science Research Network. https://doi.org/10.2139/ssrn.1106475

Seok, S. I., Cho, H., & Ryu, D. (2019). Firm-specific investor sentiment and the stock market response to earnings news. *The North American Journal of Economics and Finance*, *48*, 221-240. https://doi.org/10.1016/j.najef.2019.01.014

Tetlock, P. C. (2007). Giving Content to Investor Sentiment : The Role of Media in the Stock Market. *The Journal of Finance*, *62*(3), 1139-1168. https://doi.org/10.1111/j.1540-6261.2007.01232.x

DellaVigna, S., & Pollet, J. M. (2009). *Investor Inattention and Friday Earnings Announcements*. The Journal of Finance, 64(2), 709–749. https://doi.org/10.1111/j.1540-6261.2009.01447.x

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. https://doi.org/10.1111/j.1540-6261.2009.01501.x

Da, Z., Engelberg, J., & Gao, P. (2011). *In Search of Attention*. The Journal of Finance, 66(5), 1461–1499. https://doi.org/10.1111/j.1540-6261.2011.01679.x

Fama, E. F. (1970). Efficient Capital Markets: A Review of Theory and Empirical Work. The Journal of Finance, 25(2), 383–417. https://doi.org/10.2307/2325486

Ball, R., & Brown, P. (1968). An Empirical Evaluation of Accounting Income Numbers. Journal of Accounting Research, 6(2), 159–178. https://doi.org/10.2307/2490232

Bernard, V. L., & Thomas, J. K. (1989). Post-Earnings-Announcement Drift: Delayed Price Response or Risk Premium? Journal of Accounting Research, 27, 1–36. https://doi.org/10.2307/2491062

Fama, E. F., & French, K. R. (1993). Common Risk Factors in the Returns on Stocks and Bonds. Journal of Financial Economics, 33(1), 3–56. https://doi.org/10.1016/0304-405X(93)90023-5

Appendices

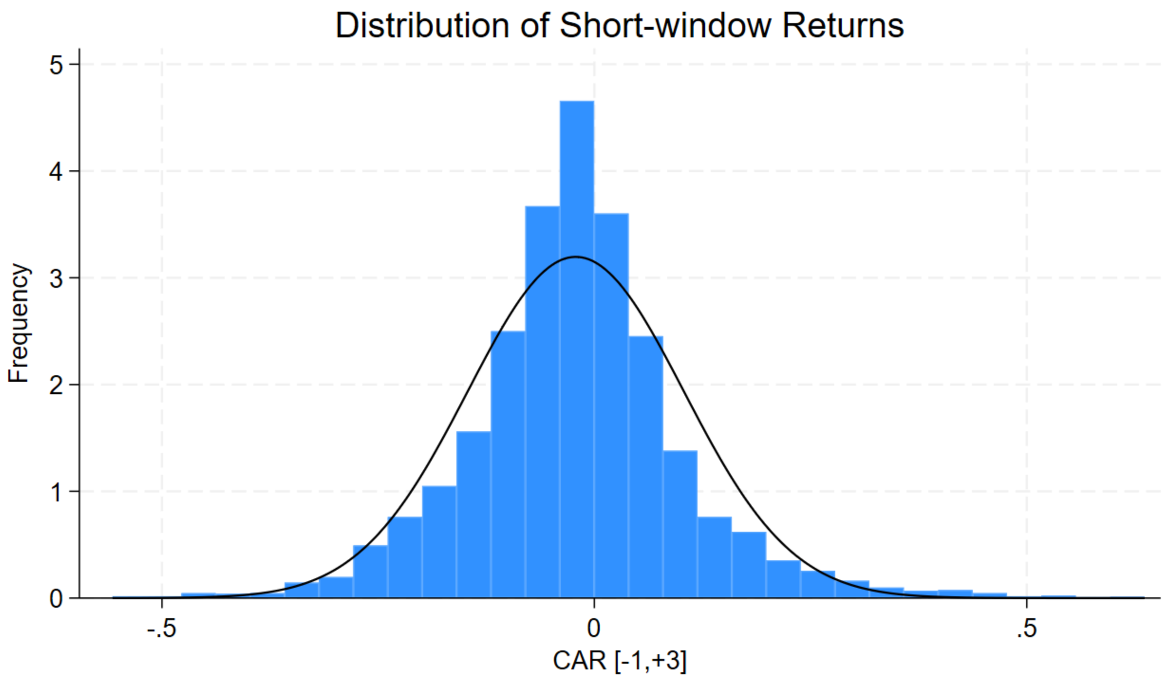
The appendices provide additional robustness specifications, descriptive figures, and variable definitions supporting the empirical analysis presented in the main text.

Appendix A. Variable Definitions

Variable	Definition
RET_adj	Event-day excess stock return measured on the earnings announcement date.
ret_short	Cumulative abnormal return over the short event window [-1,+3].
ret_post	Cumulative abnormal return over the post-announcement window [+4,+30].
surprise_std_v2	Earnings surprise measured as the difference between actual earnings and the consensus analyst forecast, normalized by stock price and standardized within the sample.
sent_30d	Thirty-day backward-looking average of the social-media sentiment index.
interaction_std_v2	Interaction term between standardized earnings surprise and pre-announcement sentiment.
lnME	Natural logarithm of firm market equity.
lnVOL	Natural logarithm of trading volume.
mktrf	Fama-French market excess return factor.
smb	Fama-French size factor (Small Minus Big).
hml	Fama-French value factor (High Minus Low).
ff12	Fama-French 12-industry classification fixed effects.
event_time	Trading-day distance relative to the earnings announcement date.
rdq_event	Earnings announcement date obtained from Compustat.
PRC	Stock price from CRSP used for normalization of earnings surprises.

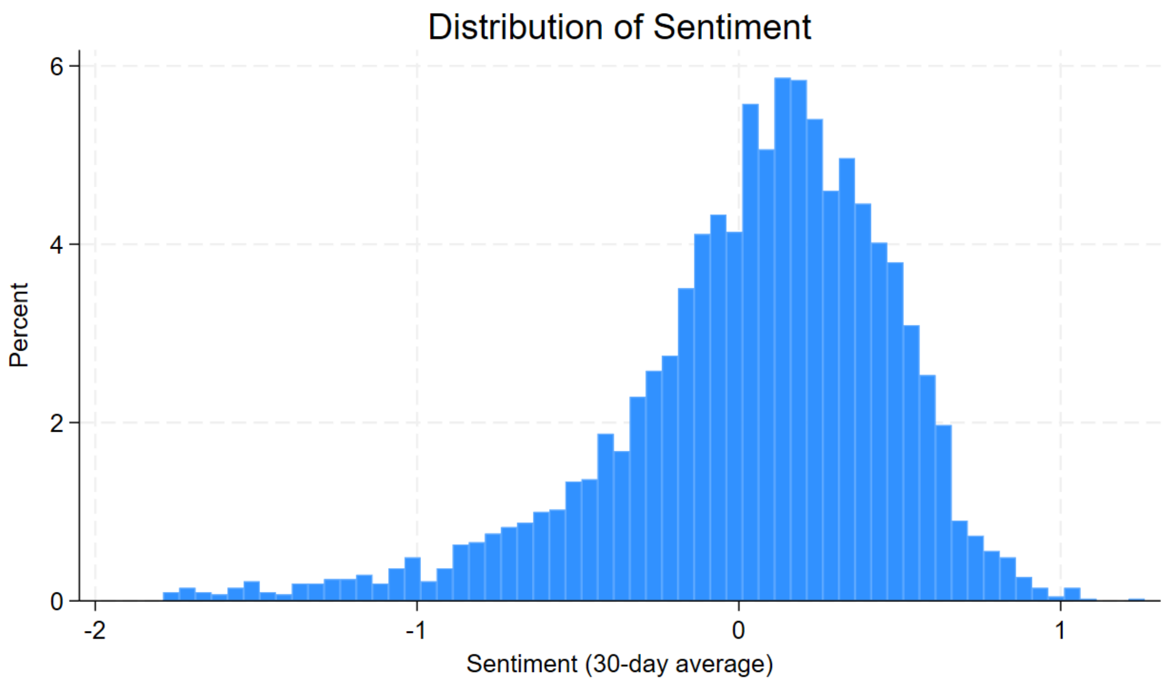
Appendix B. Additional figures

Figure 2. Distribution of Short-Window Returns



Notes : This figure displays the distribution of cumulative abnormal returns over the short event window [-1,+3]. The distribution is approximately centered around zero with moderate skewness and limited extreme observations, which is consistent with the return behavior typically observed in earnings announcement event studies.

Figure 3. Distribution of Pre-Announcement Sentiment



Notes : This figure reports the distribution of the 30-day pre-announcement sentiment measure constructed from the Social Signal dataset. The sentiment variable aggregates investor communication across multiple social media platforms and reflects the prevailing tone surrounding firms prior to earnings announcements.

Appendix C. Alternative Earnings Surprise Specification

The following appendix tables report an earlier version of the empirical analysis based on an alternative earnings surprise construction prior to normalization by stock price. In these specifications, earnings surprises were standardized within the sample but not scaled by firm price levels.

The final specification adopted in the main analysis therefore introduces a price-normalized earnings surprise measure to reduce the influence of extreme observations and ensure comparability between firms.

Importantly, the core empirical conclusions remain unchanged across specifications. Earnings surprises continue to predict short-term returns, while the interaction between sentiment and earnings surprises remains statistically insignificant. The appendix tables are presented for transparency and illustrate that the absence of a robust interaction effect is not driven by the specific construction of the earnings surprise variable.

Table 6. Baseline Regression Results

	(1) Event return	(2) CAR [-1,+3]	(3) CAR [+4,+30]
Earnings surprise (std.)	0.00402***	0.00760***	0.00139
	(0.00154)	(0.00207)	(0.00433)
Sentiment (30-day avg.)	0.000712	-0.00318	-0.00319
	(0.00306)	(0.00496)	(0.00804)
Surprise × Sentiment	-0.00556	-0.00254	-0.00256
	(0.00402)	(0.00476)	(0.0114)
Log market equity	0.000730	0.00325***	-0.00397**
	(0.000831)	(0.000952)	(0.00155)
Log volume	0.00534***	0.00259	-0.00325
	(0.00202)	(0.00198)	(0.00317)
Market factor	0.0114***	0.00954***	0.00999***
	(0.00111)	(0.00187)	(0.00298)
SMB	0.0105***	0.0140***	-0.0115**
	(0.00184)	(0.00330)	(0.00457)
HML	-0.00357***	-0.00748***	0.00391
	(0.00113)	(0.00202)	(0.00298)
Constant	-0.0899***	-0.136***	-0.0230
	(0.0236)	(0.0257)	(0.0460)
Observations	3265	3265	3265

Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
Standard errors are clustered at the firm level.

Table 7: Robustness Checks

	(1) Full	(2) Without FF factors	(3) No controls
Earnings surprise (std.)	0.00760***	0.00792***	0.00854***
	(0.00207)	(0.00205)	(0.00204)
Sentiment (30-day avg.)	-0.00318	-0.00539	-0.00570
	(0.00496)	(0.00502)	(0.00500)
Surprise × Sentiment	-0.00254	-0.00225	-0.00131
	(0.00476)	(0.00491)	(0.00486)
Log market equity	0.00325***	0.00324***	
	(0.000952)	(0.000960)	
Log volume	0.00259	0.00190	
	(0.00198)	(0.00200)	
Market factor	0.00954***		
	(0.00187)		
SMB	0.0140***		
	(0.00330)		
HML	-0.00748***		
	(0.00202)		
Constant	-0.136***	-0.127***	-0.0462***
	(0.0257)	(0.0260)	(0.00729)
Observations	3265	3265	3265

Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Appendix D. Alternative Sentiment Horizon Specification

The following appendix tables report alternative specifications based on a one-day lagged sentiment measure rather than the final 30-day backward-looking sentiment average used in the main analysis.

These earlier specifications were estimated during the initial stages of the empirical analysis in order to examine whether high-frequency sentiment fluctuations influenced the market response to earnings announcements. While the lagged sentiment measure generated stronger direct sentiment coefficients, the estimates were substantially more volatile and sensitive to short-term noise commonly associated with daily social-media-based sentiment indicators.

Most importantly, the central conclusion of the dissertation remains unchanged across specifications. Earnings surprises continue to display significant explanatory power for short-term returns, while the interaction between sentiment and earnings surprises remains statistically insignificant. The appendix tables are therefore presented for transparency and to illustrate that the absence of a robust interaction effect does not depend on the choice of sentiment horizon.

Figure 4 Distribution of pre-event sentiment (-1 day lag without average)

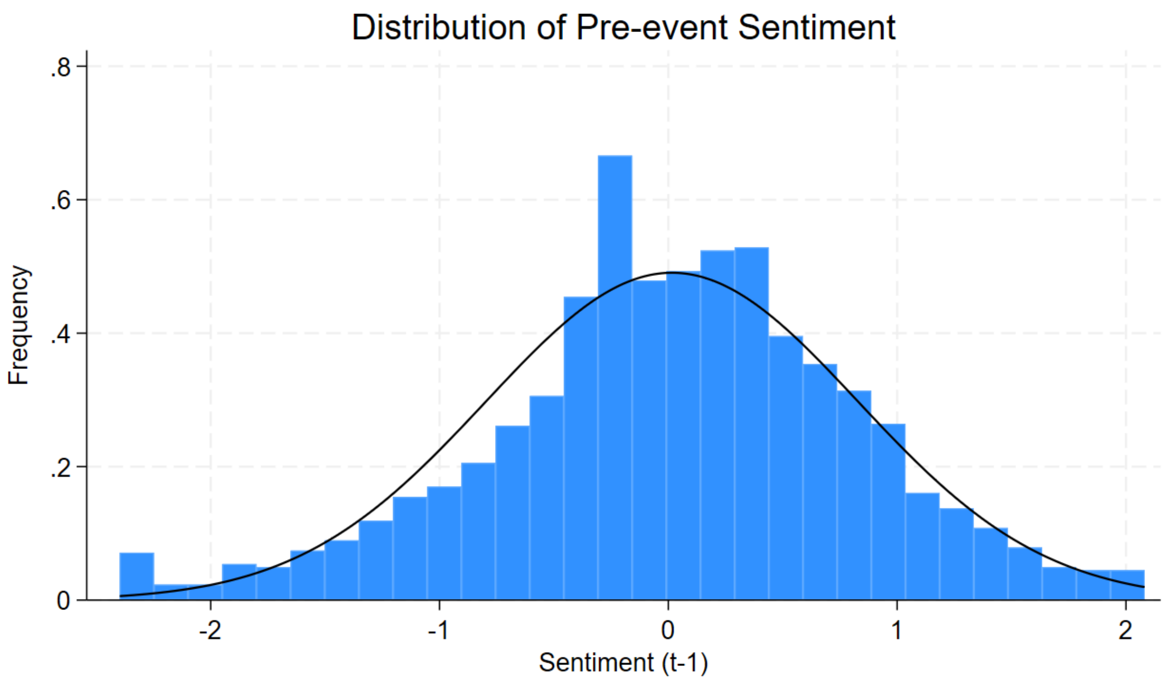


Table 8: Sentiment, Earnings Surprise, and Returns

	(1) Event return	(2) CAR [-1,+3]	(3) CAR [+4,+30]
Earnings surprise	0.003^{***}	0.006^{***}	0.001
	(0.001)	(0.001)	(0.003)
Sentiment (t−1)	0.001	0.009^{***}	0.004
	(0.001)	(0.002)	(0.003)
Surprise × Sentiment	-0.002	-0.001	0.001
	(0.001)	(0.002)	(0.003)
Log market equity	0.001	0.003^{***}	-0.003^{*}
	(0.001)	(0.001)	(0.001)
Log volume	0.003^{***}	0.002	-0.003
	(0.001)	(0.002)	(0.003)
Market factor	0.010^{***}	0.013^{***}	0.013^{***}
	(0.001)	(0.001)	(0.001)
SMB	0.010^{***}	0.014^{***}	0.024^{***}
	(0.002)	(0.002)	(0.001)
HML	-0.004^{***}	-0.001	0.002^{***}
	(0.001)	(0.001)	(0.001)
Constant	-0.069^{***}	-0.098^{***}	0.051
	(0.012)	(0.019)	(0.036)
Observations	3,462	3,462	3,462
R-squared	0.094	0.135	0.327

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 9: Robustness: Short-window sentiment effect

	(1) Full model	(2) No FF factors	(3) No controls
Earnings surprise	0.006***	0.005***	0.005***
	(0.001)	(0.001)	(0.001)
Sentiment (t-1)	0.009***	0.011***	0.011***
	(0.002)	(0.003)	(0.003)
Surprise x Sentiment	-0.001	-0.000	-0.000
	(0.002)	(0.002)	(0.002)
Log market equity	0.003***	0.004***	
	(0.001)	(0.001)	
Log volume	0.002	0.002	
	(0.002)	(0.002)	
MKT	0.013***		
	(0.001)		
SMB	0.014***		
	(0.002)		
HML	-0.001		
	(0.001)		
Constant	-0.098***	-0.099***	-0.016***
	(0.019)	(0.021)	(0.002)
Observations	3,462	3,462	3,462
R-squared	0.135	0.020	0.010

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$