The Value of Stock Analysis in News*

Shuting Li[†] Zhen Shi[‡] Yusen Xia[§] Baozhong Yang[¶]

This Draft: November, 2025

This study examines whether financial journalism provides value-relevant insights beyond corporate disclosures. Using large language models, we match news coverage in WSJ articles to firm disclosures in 8-K filings and decompose news content into components that reiterate corporate releases and that provide incremental information and analysis. We find that WSJ articles incorporate substantial analytical commentary on macroeconomic conditions, competitive dynamics, regulatory challenges, and technological innovations. Sentiment in this incremental analysis predicts substantial short-term abnormal returns, long-term stock performance, and future operating performance. Effects are more pronounced among firms with high information asymmetry. Our findings demonstrate that financial journalism delivers material insights beyond regulatory disclosures, thereby enhancing price discovery in capital markets.

Keywords: News Media, Financial Analysis, Corporate Disclosure, Generative AI, Large Language Models (LLMs).

*We thank Tony Cookson and Marina Niessner for helpful comments.

[†]J. Mack Robinson College of Business, Georgia State University, Email: sli49@gsu.edu

[‡]J. Mack Robinson College of Business, Georgia State University, Email: zshi@gsu.edu

[§]J. Mack Robinson College of Business, Georgia State University, Email: ysxia@gsu.edu

[¶]J. Mack Robinson College of Business, Georgia State University, Email: bzyang@gsu.edu

1. Introduction

The role of news media in financial markets has become increasingly important as investors navigate an overwhelming volume of information from diverse sources, such as regulatory filings, corporate announcements, and social media platforms.[1] While firms are required to disclose material events through standardized filings such as the SEC Form 8-K, many investors access such information indirectly from financial news outlets, such as the Wall Street Journal (WSJ). These news outlets serve as critical information intermediaries by interpreting, contextualizing, and disseminating corporate disclosures to a broader audience. In this study, we aim to understand the potential value of such services provided by the news media, in particular: Does financial journalism provide valuable insights beyond regulatory disclosures? And if so, how can we measure the value of such financial analysis in news media?

The key challenge in analyzing the value of news analysis lies in isolating it from the fundamental information contained in the underlying event covered by the news. In this paper, we made a first attempt at disentangling these components in news analysis. Our analysis relies on two complementary sources surrounding corporate events: 8-K filings and WSJ articles. The 8-K filings reflect firms' own disclosures of corporate events, whereas the WSJ articles provide the corresponding news analysis and interpretation of those events. We develop a new methodology by using large language models to help decompose the content of news articles into the fundamental information and incremental analysis.

We begin by collecting all 8-K filings and WSJ articles from 2012 to 2021. To construct a comprehensive dataset linking 8-K filings with their corresponding WSJ articles, we first match filings and articles by firm and date, and then use ChatGPT as a tool to identify the underlying corporate events and verify that each WSJ article discusses the same event disclosed in the firm's 8-K filing. Employing propensity score matching, we carefully construct a control group of 8-K filings that lack WSJ coverage, enabling a precise comparison between firms with media coverage and those without. Our findings reveal that WSJ article sentiment is positively associated with abnormal stock returns in the 10-day window surrounding

[1] See, e.g., Cookson, Engelberg, and Mullins (2023), Cookson et al. (2024), Chen et al. (2014), Bartov, Faurel, and Mohanram (2018), and Kogan, Moskowitz, and Niessner (2023).

8-K filing dates, even after controlling for the sentiment embedded within the 8-K filings themselves. These results suggest that WSJ articles exert an additional influence on short-term market reactions beyond what is conveyed in regulatory disclosures. Notably, the effects are significant for both positive and negative sentiment, with a one-standard-deviation increase in positive sentiment associated with up to 1.7% higher to CARs. We employ FinBERT sentiment analysis in our main specifications to mitigate concerns that ChatGPT's training data may contain future information related to 8-K events. The results are robust to using ChatGPT-based sentiment.

Beyond short-term price effects, our analysis reveals that WSJ article sentiment has long-term predictive power. Specifically, WSJ sentiment forecasts long-term stock returns over horizons of 1, 2, 3, 6, and 12 months, with a one–standard-deviation increase in positive sentiment associated with up to a 2.54% increase in CARs. This finding suggests that investors initially underreact to the information conveyed in WSJ articles, and the market gradually incorporates it into prices over time. Moreover, WSJ article sentiment predicts future firm operating performance, with a one-standard-deviation increase in positive sentiment associated with up to a 0.66% improvement in ROA over the following four quarters, reinforcing the notion that WSJ articles convey valuable fundamental insights.

To further dissect the informational contribution of WSJ articles, we leverage ChatGPT to systematically classify WSJ content into three distinct components: information that directly overlaps with 8-K filings, additional analytical insights extending beyond the regulatory disclosures, and content unrelated to the 8-K event. This decomposition allows us to isolate the incremental value provided by financial journalism. Our findings reveal that the sentiment of additional analyses in WSJ articles is significantly associated with both short-term and long-term abnormal stock returns, as well as firm operating performance. These findings highlight the unique informational role of WSJ articles, demonstrating that investors respond to this supplementary analysis rather than merely reacting to repackaged regulatory content. Furthermore, the predictability of long-term stock returns suggests a persistent investor underreaction, with WSJ-provided information gradually incorporated into stock prices over time.

Delving deeper into the content of WSJ articles, we find that they extend far beyond corporate filings by covering a diverse array of contextual factors surrounding 8-K events. These include economic conditions and global market influences, sector-specific challenges and opportunities, consumer behavior and product market dynamics, regulatory and legal challenges, financial performance and market reactions, investment strategies and market trends, mergers and acquisitions, corporate social responsibility and sustainability, corporate governance and leadership changes, and technological innovations and adaptations. By integrating such information, WSJ articles help investors synthesize and interpret corporate disclosures within a broader market context, enabling a more informed assessment of firm value and future prospects.

Given our findings, we hypothesize that WSJ articles have a greater influence on stock prices in firms characterized by higher information asymmetry, such as smaller firms, those with limited analyst coverage, and firms with lower stock liquidity. To test this, we categorize firms into subsamples based on these measures of information asymmetry. Consistent with our predictions, the results indicate that WSJ articles exert a stronger impact on firms with greater opacity, suggesting that financial media plays a pivotal role in mitigating information gaps where regulatory disclosures alone may be insufficient.

To further validate our findings, we conduct a robustness check using ChatGPT-based sentiment. ChatGPT classifies each 8-K filing and WSJ article into five tone categories, providing sentiment at the document level and capturing contextual meaning that extends beyond the sentence-level FinBERT measure. Using these scores, we find that the effect of WSJ coverage on stock performance remains robust, and the decomposition results closely mirror those obtained with FinBERT, reinforcing the reliability of our conclusions across different sentiment classification approaches.

Our study makes several unique contributions to the literature on the role of news media in the stock market. First, our findings highlight the critical role of the news media as an information intermediary in generating new insights. While previous research has debated whether media coverage merely disseminates existing information, generates new information about firm fundamentals, or serves as a proxy for investor sentiment or non-informational trading, our study provides direct empirical evidence supporting that WSJ creates valuable insights.[2] Specifically, using large language models (LLMs), we demonstrate that the WSJ conveys additional information not captured in 8-K filings which provides critical context and analysis such as macroeconomic conditions, industry trends, and regulatory developments. These additional insights enhance investors' understanding of corporate events, underscoring the interpretive and value-creating role of high-quality financial journalism in capital markets.

Furthermore, we underscore the role of financial media in influencing and predicting stock returns and risk (e.g., Calomiris and Mamaysky, 2019; Jeon, McCurdy, and Zhao, 2022; Adämmer and Schüssler, 2020).[3] Prior studies provide evidence that the sentiment and tone conveyed in financial news can predict stock returns, volatility, and broader market risk. Our study differs in two ways. First, while earlier work primarily focuses on short-term market reactions, we show that WSJ sentiment also predicts long-term stock and operating performance, particularly for firms characterized by greater information asymmetry. Second, we take a firm-specific, event-level approach by directly matching WSJ articles with 8-K filings that disclose corporate events. This design allows us to isolate how media framing and analysis, rather than the underlying fundamental information, influences investors' interpretation of specific firm disclosures.

Finally, our study contributes to the growing literature on the use of big data and machine learning in finance, accounting, and economics. Recent research highlights the potential of generative AI models to enable capital market participants to extract valuable insights from vast amounts of textual information (Bybee, 2023; Jha et al., 2024; Chen, Peng, and Zhou, 2024; Li et al., 2024). Several contemporaneous studies examine the impact of generative AI on firm value and stock returns (Eisfeldt, Schubert, and Zhang, 2023; Bertomeu et al., 2023; Chen, Kelly, and Xiu, 2022; Lopez-Lira and Tang, 2023; Babina et al., 2024). Our paper differs in two important aspects. First, we employ generative AI not merely for text classification or prediction tasks, but for complex information extraction that requires contextual understanding and comparison across sources. Second, our approach leverages

[2] For example, see Tetlock (2007); Tetlock, Saar-Tsechansky, and Macskassy (2008); Engelberg and Parsons (2011); Fang, Peress, and Zheng (2014); Obaid and Pukthuanthong (2022); Peress (2014); Kaniel and Parham (2017).

[3] Other related studies include Bali et al. (2018); Bybee, Kelly, and Su (2023); Glasserman, Li, and Mamaysky (2019); Fang and Peress (2009); Glasserman and Mamaysky (2019); Nyman, Kapadia, and Tuckett (2021).

advanced prompting strategies to analyze, align, and decompose financial text, enabling us to accurately identify and isolate relevant corporate events discussed in both regulatory filings and media coverage. In doing so, we demonstrate how LLMs can be systematically applied to enhance empirical research in capital markets.

2. Data Collection and Sample Construction

2.1. Data collection

Our analysis draws on two primary textual sources: 8-K filings and Wall Street Journal (WSJ) articles. For the 8-K filings, we collect the full set of 820,710 filings filed between 2010 and 2021 from the SEC's EDGAR database. We focus specifically on the main body of each filing while removing all attachments such as exhibits and graphics. For the news data, we collect all WSJ articles published between 2012 and 2021.[4] Similarly, we retain only the textual content and remove any exhibits or graphics.

Under SEC regulations, firms are required to file a Form 8-K when a specified triggering event occurs, or they may also use Form 8-K to voluntarily disclose corporate events that are not otherwise mandated. We list a comprehensive list of the items in Form 8-K in Table C.1 of Appendix C. The triggering events, along with their corresponding items, are provided in Sections 1-6 and 9. Once a triggering event occurs, the firm must file Form 8-K within four business days. Among the mandatory items, the most frequently triggered ones are *Item 2.02 Results of Operations and Financial Condition*, which is typically filed when firms announce earnings, and *Item 1.01 Entry into a Material Definitive Agreement*, which applies when firms enter into significant contracts such as merger agreements, long-term supply arrangements, or joint ventures. Among the mandatory items, we excluded *Item 9.01 Financial Statements and Exhibits* from our sample, as it is rarely triggered independently and typically serves as a supplement to other items, such as *Item 2.02*. Its primary purpose is to provide audited financial statements or pro forma data as attached exhibits, rather than to serve as the main disclosure of a corporate event.

Unlike the mandatory items, *Item 7.01* ensures *Fair Disclosure* by requiring firms to

[4]We focus on studying this recent 10-year period due to the cost of LLMs processing a large amount of textual files.

publicly release material nonpublic information shared with select parties, such as in investor presentations or analyst calls.[5] *Item 8.01* allows firms to voluntarily disclose information they consider important to shareholders or market that do not fall under any other specific item of Form 8-K, such as credit rating changes, major descriptions on operations.

We obtain daily stock returns and market capitalization data from the Center for Research in Security Prices (CRSP), and accounting data, such as book value of assets, from Compustat. Data on analyst coverage are drawn from the Institutional Brokers' Estimate System (I/B/E/S) summary files.

2.2. Matching 8-K filing and WSJ article

Our analysis relies on a comparison between firms' own disclosures and WSJ coverage of the same corporate events. The central premise of the comparison is to identify pairs of 8-K filings and WSJ articles that report on the same underlying corporate events. This task presents several challenges. First, not all events disclosed on 8-K filings receive media attention, and, conversely, not all events reported by the WSJ are disclosed by firms themselves. Firms only file an 8-K if the corporate events fall within the scope of Form 8-K triggering events or if they choose to disclose the event voluntarily. Second, even when the same event is covered by both sources, the disclosure and media coverage may appear on different dates and, more importantly, use divergent framing and language. These differences in timing and presentation complicate efforts to establish equivalence across the two sources.

To effectively identify 8-K and WSJ article pairs, we adopt the following approach. First, for each WSJ article, we extract the firms mentioned and their corresponding stock tickers, discarding articles that do not reference any publicly traded firms. This filtering yields a sample of 295,446 articles. Using these tickers, we map firms to CRSP's permanent company number (PERMNO) and Compustat's global company key (GVKEY). Similarly, for each 8-K filing, we obtain the firm's PERMNO and GVKEY based on its CIK identifier. We then merge 8-K filings with WSJ articles on PERMNO and GVKEY, requiring that a news article

[5] Disclosure under *Item 7.01* is voluntary in general, but becomes effectively mandatory when a firm makes selective disclosure of material nonpublic information, since Regulation FD requires that such information be made public. In other words, if a firm discloses material nonpublic information to certain parties, such as analysts, institutional investors, firms need to simultaneously make that information public.

be published within a window of four business days before to four business days after the 8-K filing date. This restriction reflects the SEC's requirement that most items be reported within four business days of the triggering event.[6] This procedure produces filing–news pairs that satisfy two criteria: (1) the WSJ article references the same firm that filed the 8-K, and (2) the article's publication date falls within the specified eight business days window around the filing date.

The filing–news pairs give us potential matches that may concern the same underlying corporate events. To verify which pairs truly describe the same events, we employ a textual alignment procedure using ChatGPT. We adopt ChatGPT for this task because its advanced natural language understanding capabilities allow it to capture semantic nuances and differences in framing between corporate disclosures and media reports. Compared with conventional keyword- or rule-based Natural Language Processing (NLP) methods, ChatGPT is better able to recognize when two texts describe the same event despite differences in terminology, emphasis, or narrative style, thereby improving the precision of our event-matching process.

We employ the prompt shown in Table B.1 in Appendix B to carry out the textual alignment procedure. The prompt directs the model to (i) summarize the corporate events described in both texts, and (ii) assess whether the two descriptions pertain to the same event. This process yields a refined set of filing–article pairs in which the 8-K disclosure and the corresponding news coverage can be confidently linked to the same event. We manually review a sample of the matched pairs to further validate the reliability of this approach.

We provide an example of an event disclosed by a firm and simultaneously covered by the WSJ in Table B.5 in Appendix B. The firm in focus is Moderna, and the event was covered by a WSJ article titled *Moderna Gets U.S. Funding for Development, Manufacturing*

[6] There are four exceptions: *Item 2.02 Results of Operations and Financial Condition*, *Item 7.01 Regulation FD Disclosure*, *Item 8.01 Other Events*, and *Item 9.01 Financial Statements and Exhibits*. Items 2.02 and 7.01 are exempted because both are treated as "furnished" rather than "filed" under the Exchange Act. Information that is furnished is intended to inform the market but does not trigger legal liability for misstatements under Section 18 of the Exchange Act. Item 8.01 is exempted because it is purely voluntary, used when management believes additional disclosure would be beneficial to investors, and therefore carries no mandated timing or content requirements. Item 9.01 is exempted because its timing depends on the availability of related information, such as financial statements for completed acquisitions, which may require additional preparation time.

of Experimental Coronavirus Vaccine shown in Panel B. Panel A presents Moderna's Form 8-K filing on April 16, 2020, the same day as the event date. In this filing, Moderna reported two items: *Item 1.01 Entry into a Material Definitive Agreement* and *Item 7.01 Regulation FD Disclosure*. Under Item 1.01, Moderna disclosed that it entered into a contract with the Biomedical Advanced Research and Development Authority (BARDA) to support the clinical development and manufacturing of its COVID-19 vaccine candidate. Under Item 7.01, the company noted that it had issued a press release announcing its entry into the BARDA contract, which was furnished pursuant to Regulation FD.

2.3. Matched sample construction

The matched filing–news pairs allows us to compare firms' own disclosures with media coverage of the same event. Beyond this within-event comparison, we also examine differences between firms whose events were covered by the WSJ and those whose events were not. A key concern in this cross-sectional analysis is selection bias, as media coverage is endogenously related to firm size and the perceived significance of the event. To mitigate this concern, we employ a Propensity Score Matching (PSM) approach. Specifically, for each event reported by the WSJ, we identify a control event from a non-covered firm with comparable characteristics, thereby constructing a matched sample that balances observable covariates between covered and non-covered events.

To construct a comparable control group, we begin by identifying non-covered events, defined as corporate events disclosed in 8-K filings that are not mentioned in WSJ articles within a (-4, +4) business-day window. We then estimate a probit model of the coverage indicator on firm characteristics and 8-K item indicators, where the independent variables consist of firm-level characteristics measured in the fiscal year prior to the filing date. The probit regression results are reported in Table C.2 of Appendix C. Using the fitted values as propensity scores, we implement 1:1 nearest-neighbor matching without replacement, requiring exact matches on the 8-K item category.

Table 1 reports the summary statistic for the matched sample, with the summary statistics in Panel A and the comparison between treated and control group in Panel B. As shown Panel B, the treated and control observations are generally similar in firm characteristics.

Although some characteristics remain statistically different, the magnitudes of differences are small. This is expected, as propensity score matching balances filings based on their likelihood of receiving WSJ coverage rather than matching each characteristic individually.

2.4. Sentiment measurement

Our analysis requires extracting the sentiment conveyed in firms' disclosures and media coverage. To this end, we employ FinBERT, a large language model (LLM) tailored for financial text (Huang, Wang, and Yang, 2023).[7] FinBERT is built upon Google's Bidirectional Encoder Representations from Transformers (BERT) architecture, which improves contextual understanding by processing text bidirectionally rather than sequentially from left to right or right to left. Unlike generic BERT models, FinBERT is pre-trained on domain-specific corpora such as 10-K and 10-Q filings, earnings call transcripts, and analyst reports. This financial specialization makes it particularly well-suited for classifying the tone of corporate disclosures and related news coverage.

Using FinBERT, we classify each sentence in 8-K filings and WSJ articles into one of three sentiment categories: positive, negative, or neutral. We assign sentiment scores of 1, –1, and 0 to sentences classified as positive, negative, and neutral, respectively, and then average these scores across all sentences in a document to obtain filing-level and article-level sentiment measures. This approach yields consistent, context-aware sentiment estimates for both corporate disclosures and media coverage, enabling systematic comparison of tones across the two sources.

Panel A of Table 1 provides summary statistics on the FinBERT sentiment of our matched sample. The average 8-K sentiment is 0.00 with a standard deviation of 0.06, indicating that 8-K disclosures are written in a predominantly neutral tone. By contrast, the average WSJ sentiment score is 0.01 with a standard deviation of 0.18, suggesting that WSJ articles convey a slightly more positive tone and exhibit greater variation in sentiment across the matched sample.

[7] We later also use ChatGPT-based sentiment measures and find qualitatively similar results. We chose to adopt FinBERT-based sentiment as our main measure because it was based on much earlier training data and is less subject to potential look-ahead biases.

3. The Incremental Effect of WSJ Coverage

3.1. Regression specification

To examine the incremental effect of WSJ sentiments on stock return and firm performance, we run the following regression on the matched sample:

$$
\begin{aligned}
y_{i,t} = \ & \beta_1 Treat \times WSJ_{pos} + \beta_2 Treat \times WSJ_{neg} + \ \beta_3 Treat \times 8K_{pos} + \beta_4 Treat \times 8K_{neg} \\
& + \beta_5 8K_{pos} + \beta_6 8K_{neg} + Controls + \alpha_{ind} + \alpha_t + \epsilon_{i,t},
\end{aligned}
\tag{1}
$$

where outcome variables $y_{i,t}$ include stock returns, measured by cumulative abnormal returns $CAR[-5,k]$, and operating performance, measured by return on assets ROA. $Treat$ is an indicator variable equal to one if the 8-K filing discloses event(s) covered by the WSJ. WSJ_{pos} (WSJ_{neg}) denotes the interaction between the absolute value of the WSJ sentiment score and a positive (negative) sentiment dummy, where the dummy equals one if the WSJ sentiment is classified as positive (negative). Similarly, $8K_{pos}$ ($8K_{neg}$) is defined as the interaction between the absolute value of the 8-K sentiment score and a positive (negative) sentiment dummy, with the dummy equal to one if the 8-K sentiment is positive (negative).

Therefore, β_5 and β_6 capture the baseline effects of 8-K tone for events that are not covered by the WSJ. β_3 and β_4 measure the incremental impact of 8-K sentiment when the event is also covered by the WSJ, relative to uncovered events. Our primary coefficients of interest are β_1 and β_2, corresponding to the interaction terms, $Treat \times WSJ_{pos}$ and $Treat \times WSJ_{neg}$. These coefficients identify the incremental effect of WSJ sentiment intensity on outcomes, relative to events disclosed only through the firm's 8-K filing.

We control for the natural logarithm of market capitalization $log(ME)$, book-to-market ratio, Z-score, institutional ownership, the dummy for S&P 1500 indicating if the stock is included in S&P 1500, ROA, investment-to-assets ratio, tangibility, momentum, fluidity, the natural logarithm of illiquidity $log(Illiq)$, the cumulative abnormal return three trading days to one day prior to the filing date $CAR[-3,-1]$, and the indicators of the eight major types of 8-K items. We also include industry and year fixed effects.

3.2. Effect on stock performance

We begin by examining the incremental effect of WSJ sentiment on stock performance, focusing on both short-term market reactions and long-term returns. The results, estimated using the baseline regression specification in Equation (1), are discussed below.

Table 2 reports the estimated short-term effects of WSJ coverage on stock returns. Panel A presents results using the 8-K filing date as the event date, while Panel B presents results using the WSJ publication date as the event date. Column (1) through (9) show the regression results with dependent variable $CAR[-5,k]$, where $CAR[-5,k]$ is the CAR from the five trading days prior to the event date to the k trading days following the event date (k ranges from -5 to 5) relative to Carhart (1997) four-factor asset pricing model. Our primary independent variables of interest are $Treat \times WSJ_{pos}$ and $Treat \times WSJ_{neg}$.

Panel A of Table 2 shows that WSJ sentiment has a significant incremental effect on short-term stock returns surrounding 8-K filing date. Specifically, the interaction between coverage and positive sentiment $Treat \times WSJ_{pos}$ is consistently positive and statistically significant at the 1% level across all event windows, indicating that firms with 8-K disclosures subsequently covered by positively framed WSJ articles experience stronger abnormal returns. Most importantly, the coefficients in columns (2) through (9) display an upward trend, indicating that this incremental effect strengthens as the event window expands. In contrast, the interaction with negative sentiment $Treat \times WSJ_{neg}$ is strongly negative and significant, suggesting that negatively framed coverage leads to lower abnormal returns. The coefficients on $Treat \times WSJ_{neg}$ decline steadily across columns (2) through (9), consistent with a strengthening adverse effect over longer windows. The magnitude of these effects is economically meaningful, with CARs ranging from about +0.8% to nearly +10% for positive sentiment and comparable declines for negative sentiment, depending on the event window.

By comparison, the coefficients on the interaction between coverage and 8-K sentiment $Treat \times 8K_{pos}$ and $Treat \times 8K_{neg}$, and the coefficients on 8-K sentiment alone $8K_{pos}$ and $8K_{neg}$ are weaker and mostly insignificant. Negative 8-K tone ($8Kneg$) is associated with lower CARs and becomes statistically significant in multiple windows as the horizon extends (e.g., $[-5,-2]$ to $[-5,3]$). Moreover, $Treat \times 8Kpos$ is positive and significant around the event date, indicating that when WSJ covers a filing, positive tone in the underlying 8-K contributes additional explanatory power for short-term returns. However, both effects are smaller in magnitude and less persistent than the corresponding WSJ sentiment coefficients. Overall, these results are consistent with the notion that media framing in the WSJ provides information beyond what is disclosed in the filings themselves and highlights the important role of media sentiment in shaping market reactions to corporate disclosures.

We repeat the analysis using the WSJ publication date rather than the 8-K filing date to address the concern that news articles may be released either earlier or later than the corresponding 8-K disclosure. The results, reported in Panel B of Table 2, are consistent with those based on the filing date. Specifically, the interaction between coverage and positive sentiment $Treat \times WSJ_{pos}$ remains positive and highly significant across all event windows, while the interaction with negative sentiment $Treat \times WSJ_{neg}$ continues to be strongly negative and significant. The magnitudes of the coefficients are comparable to those reported in Panel A, and both sets of estimates display similar patterns, with increasing magnitude as the event window expands. Taken together, these findings suggest that our results are robust to whether abnormal returns are measured around the 8-K filing date or the WSJ publication date, alleviating concerns that the choice of event date affects the observed effects.

Next, we extend the post-filing window to one year after the filing date to examine whether the abnormal returns eventually reverse over the longer horizon. Table 3 reports the estimated long-term effects of WSJ coverage on stock returns. Panel A presents results using the 8-K filing date as the event date, while Panel B uses the WSJ publication date.

Panel A of Table 3 shows that positively framed coverage exerts a persistent and economically significant impact on returns, with effects that remain substantial over longer horizons. The interaction between treatment and positive sentiment $Treat \times WSJ_{pos}$ is consistently positive and highly significant across all horizons, with CARs ranging from approximately 9.7% over a 10-day window to more than 14% over a 252-day window. In contrast, the interaction with negative sentiment $Treat \times WSJ_{neg}$ is strongly negative and statistically significant across multiple event windows (e.g., –10.8% over [–5, 5]), though the magnitude and significance become weaker at the six-month and one-year horizons.

The coefficients on the interaction between coverage and 8-K sentiment ($Treat \times 8K_{pos}$

and $Treat \times 8K_{neg}$) and the coefficients on 8-K sentiment alone ($8K_{pos}$ and $8K_{neg}$) remain small and statistically insignificant. Collectively, these findings suggest that media framing not only affects short-term market reactions but also predicts long-term stock performance. The short-term abnormal returns do not reverse over time, indicating they reflect fundamental information rather than sentiment or non-fundamental trading. Moreover, the persistent predictive power suggests that investors initially underreact to WSJ-provided information, which markets gradually incorporate over time.

Panel B of Table 3 presents the results when the WSJ publication date is used as the event date. The findings closely mirror those based on the 8-K filing date. The interaction between treatment and positive sentiment $Treat \times WSJ_{pos}$ remains consistently positive and highly significant across all horizons, while the interaction with negative sentiment $Treat \times WSJ_{neg}$ continues to be strongly negative and significant in the short and medium term, although the magnitude declines and becomes statistically weaker at longer horizons.

Taken together, the results from both panels demonstrate that the impact of media sentiment on stock returns is robust to whether the event date is defined by the 8-K filing or the WSJ publication. These findings highlight that media framing exerts a persistent and economically meaningful influence on market perceptions well beyond the initial disclosure.

3.3. Effect on operating performance

Next, we examine the incremental effect of WSJ sentiment on the operating performance of the focal firms using the regression specification in Equation (1). Similar to our analysis of stock returns, we are interested in both the short-term impact and the longer-term persistence of the effect. Accordingly, the dependent variable in our regression is the cumulative return on assets (ROA), measured from the fiscal quarter immediately following the event ROA_{t+1} through the fourth subsequent fiscal quarter $ROA_{[t+1,t+4]}$. $ROA_{[t+1,t+n]}$ is computed as the cumulative EBITDA, Earnings Before Interest, Taxes, Depreciation, and Amortization, from quarter $t+1$ through quarter $t+n$ following the (where n ranges from 2 to 4) scaled by total assets at quarter $t+1$.

Table 4 presents the results. The results indicate that positively framed coverage has a significant and persistent positive effect. The interaction between treatment and positive sentiment $Treat \times WSJ_{pos}$ is positive and statistically significant across all horizons, with the magnitude of the effect increasing from 1.07 percentage points in ROA_{t+1} to 3.65 percentage points over the four-quarter horizon. In contrast, the interaction with negative sentiment $Treat \times WSJ_{neg}$ is negative and statistically significant, with effects ranging from −1.48 percentage points in the quarter immediately following the event to −4.11 percentage points cumulatively over four quarters. In terms of economic magnitude, a one-standard-deviation increase in positive WSJ sentiment corresponds to a 0.19 to 0.66 percentage point improvement in ROA over the following year, while a one-standard-deviation increase in negative sentiment predicts a 0.27 to 0.74 percentage point decline.

Interestingly, the interaction between treatment and positive 8-K sentiment $Treat \times 8K_{pos}$ is also positive and significant in several specifications, suggesting that favorable tone in the filing itself may contribute to stronger operating performance. By contrast, the interaction with negative 8-K sentiment $Treat \times 8K_{neg}$ is statistically insignificant across all horizons, indicating that pessimistic tone in the filings alone does not meaningfully translate into weaker operating outcomes. Similarly, the coefficients on 8-K sentiment alone $8K_{pos}$ and $8K_{neg}$ are smaller in magnitude and generally less robust, highlighting the incremental informational role of WSJ coverage in predicting real outcomes.

Taken together, these findings suggest that media sentiment not only influences short-term market reactions but also predicts meaningful differences in firms' subsequent operating performance. Moreover, while positive tone in 8-K disclosures is associated with stronger outcomes, the incremental informational role of WSJ coverage is more robust and reliable in predicting operating performance, particularly in the case of negative sentiment.

4. WSJ Content Decomposition and Topic Analysis

4.1. WSJ content decomposition analysis

4.1.1. WSJ article content decomposition. Given that both the WSJ and the 8-K filings report on the same corporate event, we examine whether WSJ articles convey information beyond what is disclosed in the 8-K, and if so, whether the sentiment embedded in this additional information drives the documented effects.

To address this, we implement a text classification procedure that partitions each WSJ article covering an 8-K event into three categories: (1) Common information, defined as sentences that replicate or closely restate the disclosure in the 8-K; (2) Additional information, defined as sentences that introduce new insights or details related to the event, the firm, or other factors influencing the firm's performance; and (3) Irrelevant information, defined as sentences unrelated to the focal event or firm, or otherwise not classifiable as common or additional information.

We employ ChatGPT to classify sentences into the three categories in our matched sample. As discussed earlier, ChatGPT is particularly suitable for this task because of its ability to capture semantic nuance and contextual meaning, which enables more accurate identification of overlapping and novel content compared with conventional keyword- or rule-based NLP methods. To do this, we first pre-tokenize each WSJ article at the sentence level and assign an index to every sentence. We then apply a structured prompting procedure using ChatGPT to classify sentences into the three categories described above; the full prompt is provided in Table B.2 in Appendix B. The input consists of matched filing–news pairs, each of which covers at least one common underlying corporate event. The procedure first identifies the focal event described in both texts and assesses whether the WSJ article primarily addresses this event. If the article does, its content is partitioned into two categories: Common information and Additional information. If the article does not primarily focus on the event, its content is instead classified into three categories: Common information, Additional information, and Irrelevant information. For each category, ChatGPT returns the indices of the corresponding sentences, thereby producing a sentence-level classification scheme. To assess the reliability of this procedure, we conduct a manual validation of a sample of the results, which confirms the consistency and accuracy of the procedure.

4.1.2. Content decomposition regression. After classification, we compute segment-level sentiment scores by averaging the BERT-derived sentiment values across all sentences within each segment. This approach enables us to isolate the distinct contributions of common, additional, and irrelevant information to the incremental effect of WSJ sentiment. We then estimate the following regression on the matched sample to examine how sentiment in each segment affects firms' stock and operating performance:

$$\begin{aligned}
y_{i,t} = \ & \beta_1 \, Treat \times Common_{pos} + \beta_2 \, Treat \times Common_{neg} + \ \beta_3 \, Treat \times Additional_{pos} \\
& + \beta_4 \, Treat \times Additional_{neg} + \ \beta_5 \, Treat \times Irrelevant_{pos} + \beta_6 \, Treat \times Irrelevant_{neg} \\
& + \beta_7 \, Treat \times 8K_{pos} + \beta_8 \, Treat \times 8K_{neg} + \beta_9 8K_{pos} + \beta_{10} 8K_{neg} + Controls + \epsilon_{i,t}, \qquad (2)
\end{aligned}$$

where $Common_{pos}$ ($Common_{neg}$) denotes the interaction between the absolute value of the FinBERT sentiment score for the WSJ article's common information section and a positive (negative) sentiment dummy, with the dummy equal to one if that section is classified as positive (negative). Similarly, $Additional_{pos}$ ($Additional_{neg}$) and $Irrelevant_{pos}$ ($Irrelevant_{neg}$) denote the interactions between the absolute value of the FinBERT sentiment score for the additional information and irrelevant information sections, respectively, and their corresponding positive (negative) sentiment dummies. The outcome variables $y_{i,t}$ include stock performance and operating performance. $Treat$ is a dummy equal to one if the 8-K filing discloses event(s) covered by the WSJ.

Our primary coefficients of interest are β_3 and β_4, which correspond to the interaction terms $Treat \times Additional_{pos}$ and $Treat \times Additional_{neg}$. These two coefficients capture the incremental effect of the sentiment embedded in the WSJ article's additional information section, relative to events disclosed solely through the firm's 8-K filing. To isolate these effects, β_1 and β_2 measure the impact of sentiment in the common information section, while β_5 and β_6 capture the effect of sentiment in the irrelevant information section. In addition, β_7 and β_8 identify the incremental influence of 8-K sentiment when the event is also covered by the WSJ, while β_9 and β_{10} reflect the baseline effect of 8-K sentiment for events that are not covered by the WSJ. Taken together, this empirical setup enables us to disentangle the specific contribution of WSJ additional information from common information and unrelated content, while also accounting for the role of the firm's own disclosure tone.

Consistent with the baseline regression specification in Equation (1), we control for firm characteristics, the three-day pre-event return $CAR[-3, -1]$, and 8-K item types. In addition, we include industry and year fixed effects and cluster standard errors at the firm level.

4.1.3. Evidence on the decomposed impact of WSJ sentiment. We present the results of Equation (2) in Table 5. Panel A reports the estimates for stock performance, and Panel B reports the estimates for operating performance over a one-year horizon.

Panel A of Table 5 presents the decomposed effects of WSJ sentiment on stock performance across information categories. The dependent variables are $CAR[-5,5]$ through $CAR[-5,252]$. The results reveal several insights. First, even content that restates regulatory disclosures (common information) influences market reactions when framed by WSJ, with positive sentiment generating CARs of 3.0% to 3.6% and negative sentiment producing effects of -3.0% to -3.4%. This suggests that media framing and emphasis affect how investors interpret disclosed information. Second, and more importantly, the additional information section, which represents WSJ's incremental analytical content beyond regulatory filings, exhibits substantial and significant effects on stock returns. Positive sentiment in this incremental content yields large effects (3.4% to 5.1%), while negative sentiment produces comparably large negative effects (-2.7% to -3.8%). In contrast, sentiment in the irrelevant section is generally insignificant across horizons, with only two coefficients statistically (or marginally) significant. Taken together, these findings indicate that WSJ sentiment influences stock returns through both the curation of regulatory disclosures and the provision of incremental analytical insights, while peripheral content exhibits largely insignificant effects.

Panel B of Table 5 presents the decomposed effects of WSJ sentiment on firms' operating performance, measured by cumulative ROA over the four fiscal quarters following the event. The dependent variables are ROA_{t+1} through $ROA_{[t+1,t+4]}$. The results reveal that both content categories predict future operating performance. Sentiment in the common information section, which restates regulatory disclosures, exhibits significant predictive power: positive sentiment ($Treat \times Common_{pos}$) predicts higher ROA, with effects reaching 1.10 percentage points by the fourth quarter, while negative sentiment ($Treat \times Common_{neg}$) forecasts consistent negative effects ranging from -0.43 to -1.03 percentage points. Importantly, the additional information section shows even stronger predictive power for negative outcomes. While positive sentiment ($Treat \times Additional_{pos}$) predicts modest improvements in ROA, negative sentiment ($Treat \times Additional_{neg}$) strongly forecasts adverse performance, with effects reaching -2.54 percentage points by $t+4$. In contrast, sentiment in the irrelevant information section shows no predictive power for operating outcomes. Taken together, these findings indicate that WSJ sentiment, through both the framing of disclosed information and the provision of incremental analytical content, contains forward-looking insights about firm fundamentals, whereas peripheral content exhibits no predictive value.

Overall, the evidence indicates that the documented incremental effects of WSJ sentiment are primarily driven by sentiment in the common and additional information sections, both of which show strong and consistent impacts on stock and operating performance. By contrast, sentiment contained in the irrelevant section exhibits little to no effect, underscoring that the incremental influence of WSJ coverage arises from value-relevant content.

4.2. Topic analysis

Given that the incremental effects of WSJ sentiment are primarily driven by the common and additional information segments, two natural follow-up questions arise: What information are typically discussed in both 8-K filings and WSJ articles, and what additional information does the WSJ cover beyond the common restatements? To answer these questions, we conduct a topic analysis. The idea is to identify the most prevalent themes in both 8-K filings and WSJ articles, and to classify each 8-K filing and WSJ article into these categories. This allows us to examine not only which topics are commonly discussed across the two sources but, more importantly, which additional topics the WSJ introduces when covering certain types of common topics.

Our analysis begins with topic identification of 8-K filings, as the triggering events associated with each item and section are clearly defined by SEC, and, as a result, firms filing a Form 8-K have a well-defined framework for determining which section and item correspond to a given event.[8] This regulatory clarity enables us to systematically summarize topics across sections. In addition, the structure of the form is informative in itself, as the section titles convey direct insight into the type of corporate event being reported.

We first construct a random sample of 5,000 8-K items to identify the most common ten topics. When construct the sample, we exclude *Section 7 Regulation FD Disclosure*

[8]The SEC provides explicit guidance on the scope of each item, and the items are organized under nine sections of Form 8-K. See https://www.sec.gov/files/form8-k.pdf

and *Section 9 Financial Statements and Exhibits* from our analysis, as companies often file other items in conjunction with these two sections, and the corporate event itself is typically described in other sections.[9] We conduct the topic identification at 8-K item level, as companies sometimes file multiple items for a single event, or a single item may encompass multiple types of events. In creating this sample, we ensure that the distribution of sections in the sample mirrors that of the full 8-K dataset.

Next, we employ ChatGPT to analyze the text of each 8-K item and identify five key topics per item. The prompt used for this step is presented in Panel A of Table B.4 in Appendix B. This procedure generates over 20,000 topics, with some recurring across multiple items. We then summarize the identified topics into eight categories to ensure that they collectively cover the majority of topics while remaining mutually exclusive. To achieve this, we instruct ChatGPT, using the prompt in Panel C of Table B.4 to consolidate the topics into eight broad categories based on their frequency.[10] The output includes the eight categories and their definitions, as shown in Panel D of Table B.4. Finally, using this framework, we ask ChatGPT to assign each 8-K item to the most appropriate category and count the number of links between each section–category pair.

What additional information does the WSJ typically provide for each type of corporate event? To address this question, we next apply the same procedure used for 8-K filings to the WSJ additional information sections. Our focus is on the incremental content provided by WSJ, rather than the common information already shared across the two sources. Accordingly, topic extraction and categorization are conducted at the level of the additional information segment. Specifically, we begin by identifying the key topics discussed in the relevant segments of WSJ articles. To train the categorization, we randomly select 2,000 WSJ articles matched to 8-K filings and, for each article, use ChatGPT to analyze the additional information segment and extract 10 key topics, following the prompt outlined in Panel B of Table B.4. We then summarize these topics into 10 broader categories using the prompt in Panel C, with

[9]For example, the exhibits are generally included in Section 9, and press releases are typically found in Section 7. As an example, companies that announce earnings will usually file Item 2.02 and Item 9.01, with earnings summarized under Item 2.02 and exhibits listed under Item 9.01.

[10]The prompt in Panel C of Table B.4 summarizes the input topics into ten categories. For our analysis, we aimed to consolidate the topics into eight broader categories. Accordingly, when summarizing the identified topics from 8-K items using the same prompt, we modified it to request eight categories instead of ten.

the resulting categories presented in Panel E of Table B.4. Next, we categorize all additional information segments in the matched sample into these 10 categories, following the prompts in Panel E. Finally, we count the number of pairings between each 8-K category and its corresponding WSJ additional information category.

We present the mapping between 8-K sections, the consolidated 8-K topics, and the additional information topics discussed in WSJ articles in Figure 1. The diagram is organized into three columns: the left column lists the sections of the 8-K filings with *Section 4 Matters Related to Accountants and Financial Statements* and *Section 6 Asset-Backed Securities* combined due to their low frequency. The middle column displays the eight categories of corporate events disclosed in the filings, while the right column presents the ten topics emphasized in the additional information segments of the corresponding WSJ articles. The width of each link reflects the relative frequency of connections between 8-K sections, event categories, and WSJ topics.

The two leftmost columns in Figure 1 illustrate the types of corporate events that firms typically disclose under each 8-K section. As shown, some sections, such as Sections 2 and 5, tend to be concentrated in specific corporate events, whereas others, such as Sections 8 and 1, cover a wide variety of events with broader distributions. Specifically, *Section 2 Financial Information* accounts for the largest share, mapping predominantly into the topic *Financial Reporting and Disclosure*, as well as related categories such as *Debt Financing and Restructuring*. Similarly, *Section 5 Corporate Governance and Management* flows mainly into *Executive Management and Compensation* and *Corporate Governance and Compliance*. By contrast, *Section 8 Other Events* and *Section 1 Registrant's Business and Operations* exhibit wider distributions, linking to a diverse set of categories including *Mergers and Acquisitions*, *Regulatory and Legal Proceedings*, and *Equity Transactions and Stock Issues*.

The two rightmost columns in Figure 1 illustrate, for each type of corporate event, the additional topics that WSJ articles tend to emphasize beyond the information shared in both sources. While 8-K events such as *Financial Reporting and Disclosure* and *Debt Financing and Restructuring* continue to appear in the news, the WSJ often broadens the discussion to include themes that extend beyond the original filings. For example, filing topics related to *M&A* connect not only to corresponding coverage of *Mergers and Acquisitions* activity but also to

Sector-Specific Challenges and Opportunities and Regulatory and Legal Challenges, reflecting a wider contextual framing. Similarly, topics tied to *Corporate Governance and Compliance* in 8-K filings are frequently linked to *Corporate Social Responsibility and Sustainability* in WSJ articles. Overall, this mapping highlights that WSJ coverage builds upon the information disclosed in 8-Ks while adding perspectives on consumer behavior, technological innovation, and broader market dynamics, thereby situating corporate events within a wider economic and strategic context.

5. Additional Analysis

5.1. Heterogeneity by information asymmetry

5.1.1. Heterogeneity in the impact of WSJ sentiment. We show that WSJ sentiment has an incremental effect on stock and operating performance, and that this effect is primarily driven by the sentiment expressed in segments containing common information and additional discussion. We further examine heterogeneity in these effects across firms with different levels of information asymmetry. We conjecture that the documented WSJ sentiment effect should be more pronounced for firms with higher levels of information asymmetry, as markets have more to learn from WSJ coverage when firms face greater opacity. We measure information asymmetry using firm size, analyst coverage, and stock liquidity, and estimate Equation (1) separately for subsamples of firms with higher and lower information asymmetry. We hypothesize that the effects should be stronger among smaller firms, firms followed by fewer analysts, and firms with less liquid stocks.

We begin by comparing corporate events between small and large firms, where firm size is measured by log(ME). The results are reported in Panel A of Table 6. For the low log(ME) subsample, the interaction between treatment and positive sentiment $Treat \times WSJ_{pos}$ is consistently positive and highly significant across all event windows, with CARs ranging from about 13.8% to over 22.5%. Conversely, the interaction with negative sentiment $Treat \times WSJ_{neg}$ is strongly negative and significant, reaching magnitudes of roughly –11% to –15% for shorter horizons. In contrast, the effects for the high log(ME) subsample are much weaker in both economic and statistical terms: although coefficients on positive sentiment remain positive and occasionally significant, their magnitudes are less than half those observed for smaller firms, and the corresponding negative sentiment effects are smaller as well. Importantly, the difference-in-coefficient tests confirm that both positive and negative sentiment effects are significantly larger for smaller firms. These findings are consistent with the hypothesis that WSJ sentiment has a greater influence when firms face higher information asymmetry, as is typically the case for smaller firms.

Panel B of Table 6 reports the comparison of corporate events between firms with higher analyst coverage and those with lower analyst coverage. For firms with low analyst coverage, the interaction between treatment and positive sentiment $Treat \times WSJ_{pos}$ is large, positive, and highly significant across all horizons, with CARs ranging from 13.5% to over 22%. The corresponding negative sentiment interaction $Treat \times WSJ_{neg}$ is significantly negative, with effects between –8% and –14% over shorter horizons. By contrast, the coefficients for firms with high analyst coverage are substantially smaller in magnitude, and while they remain statistically significant in some windows, the difference-in-coefficient tests confirm that the incremental effect of WSJ sentiment is significantly stronger for firms with fewer analysts following them.

A similar pattern emerges when firms are partitioned by stock liquidity, as shown in Panel C of Table 6. For firms with low liquidity, positive sentiment effects reach magnitudes of 14–23%, while negative sentiment effects range between –10% and –14% in shorter horizons. In contrast, for firms with high liquidity, the corresponding coefficients are both smaller in magnitude and less consistently significant. Again, the difference-in-coefficient tests highlight that the incremental impact of WSJ sentiment is significantly greater among firms with lower liquidity. Taken together, these results support the conjecture that WSJ sentiment exerts a larger influence on market reactions when firms face greater information asymmetry, as proxied by lower analyst coverage and weaker stock liquidity.

5.1.2. Heterogeneity in the decomposed impact of WSJ sentiment. Next, we examine the three segments of WSJ articles and investigate the decomposed impact of sentiment in each segment on stock performance using Equation (2). The results are reported in Table 7.

Panel A of Table 7 examines whether the decomposed effects of WSJ sentiment vary by firm size. For smaller firms, i.e., firms with $\log(ME)$ below the median, positive sentiment in both the common information and additional discussion segments has a strong and statistically significant impact on CARs, with effects ranging from roughly 3.6% to over 6% depending on the event window. Negative sentiment in these segments is also consistently significant, producing abnormal returns between –4% and –5% for common information and between –3% and –4% for additional discussion. By contrast, the estimates for large firms are weaker in magnitude: while positive common sentiment remains significant, the coefficients are about half the size of those for smaller firms, and the incremental effects of additional discussion are generally smaller and less persistent. The difference-in-coefficient tests reinforce this asymmetry, showing that both positive and negative sentiment effects in the common and additional segments are significantly larger for smaller firms.

Panels B of Table 7 explores heterogeneity in the decomposed effects of WSJ sentiment by analyst coverage. Firms with low analyst coverage exhibit much stronger responses to sentiment in both the common information and additional discussion segments. Positive sentiment in these segments generates large and statistically significant CARs, often exceeding 5%, while negative sentiment yields comparably large negative returns of roughly –4% to –5%. In contrast, the effects for high coverage firms are smaller in magnitude, and although some coefficients remain significant, the difference-in-coefficient tests confirm that both positive and negative sentiment effects in additional discussion segments are significantly stronger for low coverage firms.

Panel C presents similar patterns when firms are partitioned by stock liquidity. Among low liquidity firms, positive sentiment in common information and additional discussion produces economically large and highly significant abnormal returns—often above 6% in the longer horizons—while negative sentiment drives returns downward by –4% to –5%. For high liquidity firms, sentiment effects are again present but muted, with weaker magnitudes and less consistent significance. The difference-in-coefficient tests reinforce that both positive and negative sentiment effects are stronger for less liquid firms, particularly in the additional discussion segments.

Taken together, these findings indicate that the heterogeneity observed across information asymmetry groups can be largely attributed to the sentiment conveyed in the common and additional segments.

5.2. Robustness: ChatGPT-based sentiment

In the previous analyses, we rely on sentiments measured using FinBERT. To assess the robustness of our findings, we also employ ChatGPT to re-estimate our baseline specifications. Using ChatGPT provides complementary validation, as it is designed to capture broader contextual and semantic nuances, thereby offering an independent check on the reliability of our results. Moreover, compared to the FinBERT measure, which is constructed at the sentence level and then aggregated to the segment or article level, the ChatGPT-based measure is generated directly at the segment or article level, as ChatGPT evaluates longer text spans rather than individual sentences. This allows it to produce a smoother measure that is less sensitive to local fluctuations in tone.

We identify the 8-K filing sentiment using prompt shown in Panel A of Table B.3 in Appendix B. We input the text of each 8-K filing and instruct ChatGPT to evaluate it by assigning one of five sentiment scores, ranging from –1 (very negative) to +1 (very positive), with intermediate values capturing mild sentiment or neutrality. This procedure yields a filing-level measure of 8-K sentiment. We apply a similar procedure to identify WSJ article sentiment, using the prompt in Panel B of Table B.3 in Appendix B, to generate article-level sentiment scores across five categories.

Using the GPT-based sentiments, we repeat our baseline analysis to examine the impact of WSJ sentiment on stock performance using Equation (1), with the results reported in Table 8. The results show that positive WSJ sentiment $Treat \times WSJ_{pos}$ is associated with consistently positive and statistically significant CARs across all event windows, ranging from roughly 1.1% in the shortest window to over 5% over the full year. In contrast, negative WSJ sentiment $Treat \times WSJ_{neg}$ is strongly negative and highly significant across horizons, with CARs between –5% and –7.7%, indicating a pronounced adverse market reaction to negatively framed coverage. By comparison, the sentiment identified directly from 8-K filings yields weaker and less consistent results, with coefficients that are generally smaller and often insignificant. Overall, the findings based on GPT-classified sentiment are consistent with those obtained using FinBERT-based sentiment, reinforcing the evidence that WSJ sentiment exerts a significant influence on stock performance.

We also conduct subsample analyses using GPT-based WSJ sentiment to test the robustness of the documented heterogeneity effect, that corporate events of firms subject to higher information asymmetry experience stronger market reactions over the one-year horizon. Specifically, we repeat the baseline analysis following the specification in Equation (1) across the three sets of subsamples sorted by firm size, analyst coverage, and stock liquidity. The results are reported in Table 9.

In Panel A, small firms exhibit large and highly significant market reactions: positive sentiment $Treat \times WSJ_{pos}$ is associated with CARs of roughly 6–8% across horizons, while negative sentiment $Treat \times WSJ_{neg}$ produces declines of similar magnitude. In contrast, the effects for larger firms are much weaker, and the difference-in-coefficient tests confirm that both positive and negative sentiment impacts are significantly larger among smaller firms.

Panel B shows a similar pattern when firms are sorted by analyst coverage. For firms with few analysts following them, both positive and negative sentiment yield stronger CARs than for firms with greater analyst coverage, though the differences are less pronounced than in the size subsample. Panel C highlights the role of liquidity: low-liquidity firms exhibit the strongest market reactions, with positive sentiment effects of 6–9% and negative sentiment effects of –5% to –9%, compared to much smaller magnitudes for high-liquidity firms. Again, the difference-in-coefficient tests confirm that the incremental effects of WSJ sentiment are concentrated in firms with lower liquidity. Taken together, the subsample analyses demonstrate that the results based on GPT-classified WSJ sentiment are robust across different partitions of the sample by firm size, analyst coverage, and stock liquidity.

6. Conclusion

This paper investigates the role of WSJ coverage in shaping market responses to corporate disclosures and predicting long-term firm valuation. By matching 8-K filings with related WSJ articles based on the focal corporate events, we study how media sentiment affects the market's interpretation of firm-reported events. Using FinBERT to extract sentiment, we show that WSJ sentiment provides incremental information beyond what is disclosed in the filings themselves. Specifically, positively framed coverage enhances abnormal returns, while negatively framed coverage lowers them, with effects that are both statistically and economically significant.

We further document that the incremental role of WSJ sentiment is concentrated in the segments that convey common information and additional discussion. This finding highlights that the media not only restate corporate disclosures but also supplements them with broader context, which in turn is incorporated by markets and reflected on pricing. Our long-horizon analyses indicate that these effects persist well beyond the short-run event window with no evidence of reversal, suggesting that media sentiment reflects fundamental information and also predicts longer-term stock and operating performance.

A series of heterogeneity tests reinforce the robustness of our results. The effects of WSJ sentiment are consistently observed across subsamples defined by firm size, analyst coverage, and stock liquidity, with stronger estimates for firms facing greater information asymmetry. Moreover, when we re-estimate our analyses using GPT-based sentiment classification rather than FinBERT, we obtain highly consistent results, underscoring the reliability of our findings across different sentiment identification methods.

Overall, our findings underscore the complementary role of financial media in the price discovery process, reinforcing that WSJ articles provide value-relevant insights that extend beyond corporate disclosures. This has important implications for investors, analysts, and policymakers, as it highlights the media's role in mitigating information asymmetry, shaping investor sentiment, and predicting long-term firm valuation, rather than merely redistributing or entertaining existing narratives.

References

Adämmer, P., and R. A. Schüssler. 2020. Forecasting the equity premium: mind the news! *Review of Finance* 24:1313–55.

Amihud, Y. 2002. Illiquidity and stock returns: cross-section and time-series effects. *Journal of financial markets* 5:31–56.

Babina, T., A. Fedyk, A. He, and J. Hodson. 2024. Artificial intelligence, firm growth, and product innovation. *Journal of Financial Economics* 151:103745–.

Bali, T. G., A. Bodnaruk, A. Scherbina, and Y. Tang. 2018. Unusual news flow and the cross section of stock returns. *Management Science* 64:4137–55.

Bartov, E., L. Faurel, and P. S. Mohanram. 2018. Can twitter help predict firm-level earnings and stock returns? *The Accounting Review* 93:25–57.

Bertomeu, J., Y. Lin, Y. Liu, and Z. Ni. 2023. Capital market consequences of generative ai: Early evidence from the ban of chatgpt in italy. *Available at SSRN 4452670* .

Bybee, L. 2023. Surveying generative ai's economic expectations. *arXiv preprint arXiv:2305.02823* .

Bybee, L., B. Kelly, and Y. Su. 2023. Narrative asset pricing: Interpretable systematic risk factors from news text. *The Review of Financial Studies* 36:4759–87.

Calomiris, C. W., and H. Mamaysky. 2019. How news and its context drive risk and returns around the world. *Journal of Financial Economics* 133:299–336.

Carhart, M. M. 1997. On persistence in mutual fund performance. *The Journal of finance* 52:57–82.

Chen, H., P. De, Y. Hu, and B.-H. Hwang. 2014. Wisdom of crowds: The value of stock opinions transmitted through social media. *The review of financial studies* 27:1367–403.

Chen, S., L. Peng, and D. Zhou. 2024. Wisdom or whims? decoding investor trading strategies with large language models. *Zicklin School of Business, Baruch College Working Papers* .

Chen, Y., B. T. Kelly, and D. Xiu. 2022. Expected returns and large language models. *Available at SSRN 4416687* .

Cookson, J. A., J. E. Engelberg, and W. Mullins. 2023. Echo chambers. *The Review of Financial Studies* 36:450–500.

Cookson, J. A., R. Lu, W. Mullins, and M. Niessner. 2024. The social signal. *Journal of Financial Economics* 158:103870–.

Eisfeldt, A. L., G. Schubert, and M. B. Zhang. 2023. Generative ai and firm values. Working Paper, National Bureau of Economic Research.

Engelberg, J., and C. A. Parsons. 2011. The causal impact of media in financial markets. *The Journal of Finance* 66:67–97.

Fang, L., and J. Peress. 2009. Media coverage and the cross-section of stock returns. *The journal of finance* 64:2023–52.

Fang, L. H., J. Peress, and L. Zheng. 2014. Does media coverage of stocks affect mutual funds' trading and performance? *The Review of Financial Studies* 27:3441–66.

Glasserman, P., F. Li, and H. Mamaysky. 2019. Time variation in the news–returns relationship. *Journal of Financial and Quantitative Analysis* 1–37.

Glasserman, P., and H. Mamaysky. 2019. Does unusual news forecast market stress? *Journal of Financial and Quantitative Analysis* 54:1937–74.

Hoberg, G., G. Phillips, and N. Prabhala. 2014. Product market threats, payouts, and financial flexibility. *The Journal of Finance* 69:293–324.

Huang, A. H., H. Wang, and Y. Yang. 2023. Finbert: A large language model for extracting information from financial text. *Contemporary Accounting Research* 40:806–41.

Jeon, Y., T. H. McCurdy, and X. Zhao. 2022. News as sources of jumps in stock returns: Evidence from 21 million news articles for 9000 companies. *Journal of Financial Economics* 145:1–17.

Jha, M., J. Qian, M. Weber, and B. Yang. 2024. Chatgpt and corporate policies. Working Paper, National Bureau of Economic Research.

Kaniel, R., and R. Parham. 2017. Wsj category kings–the impact of media attention on consumer and mutual fund investment decisions. *Journal of Financial Economics* 123:337–56.

Kogan, S., T. J. Moskowitz, and M. Niessner. 2023. Social media and financial news manipulation. *Review of Finance* 27:1229–68.

Li, K., F. Mai, R. Shen, C. Yang, and T. Zhang. 2024. Dissecting corporate culture using generative ai–insights from analyst reports. *Available at SSRN 4558295* .

Lopez-Lira, A., and Y. Tang. 2023. Can chatgpt forecast stock price movements? return predictability and large language models. *arXiv preprint arXiv:2304.07619* .

Nyman, R., S. Kapadia, and D. Tuckett. 2021. News and narratives in financial systems: exploiting big data for systemic risk assessment. *Journal of Economic Dynamics and Control* 127:104119–.

Obaid, K., and K. Pukthuanthong. 2022. A picture is worth a thousand words: Measuring investor sentiment by combining machine learning and photos from news. *Journal of Financial Economics* 144:273–97.

Peress, J. 2014. The media and the diffusion of information in financial markets: Evidence from newspaper strikes. *the Journal of Finance* 69:2007–43.

Tetlock, P. C. 2007. Giving content to investor sentiment: The role of media in the stock market. *The Journal of finance* 62:1139–68.

Tetlock, P. C., M. Saar-Tsechansky, and S. Macskassy. 2008. More than words: Quantifying language to measure firms' fundamentals. *The journal of finance* 63:1437–67.

Figure 1. Relationships Between 8-K Sections, 8-K Topics, and WSJ Additional Information Topics

This figure summarizes the relationships between the 8-K filing sections (the left column), the topics covered in each section (the middle column), and the topics discussed in the additional information featured in WSJ articles (the right column). Our matched sample includes matched 8-K filings and WSJ articles spanning the period from 2012 to 2021. The left column displays the seven 8-K sections (excluding 8-K sections 7 and 9), while the center column lists 8 topics covered in the 8-K filings. The right column outlines 10 topics discussed in the additional information presented in WSJ articles. The height of each entity (8-K sections, 8-K topics, and WSJ topics) corresponds to the number of relevant segments, while the width of each link represents the number of relationships between the 8-K sections, 8-K topics, and WSJ additional information topics.

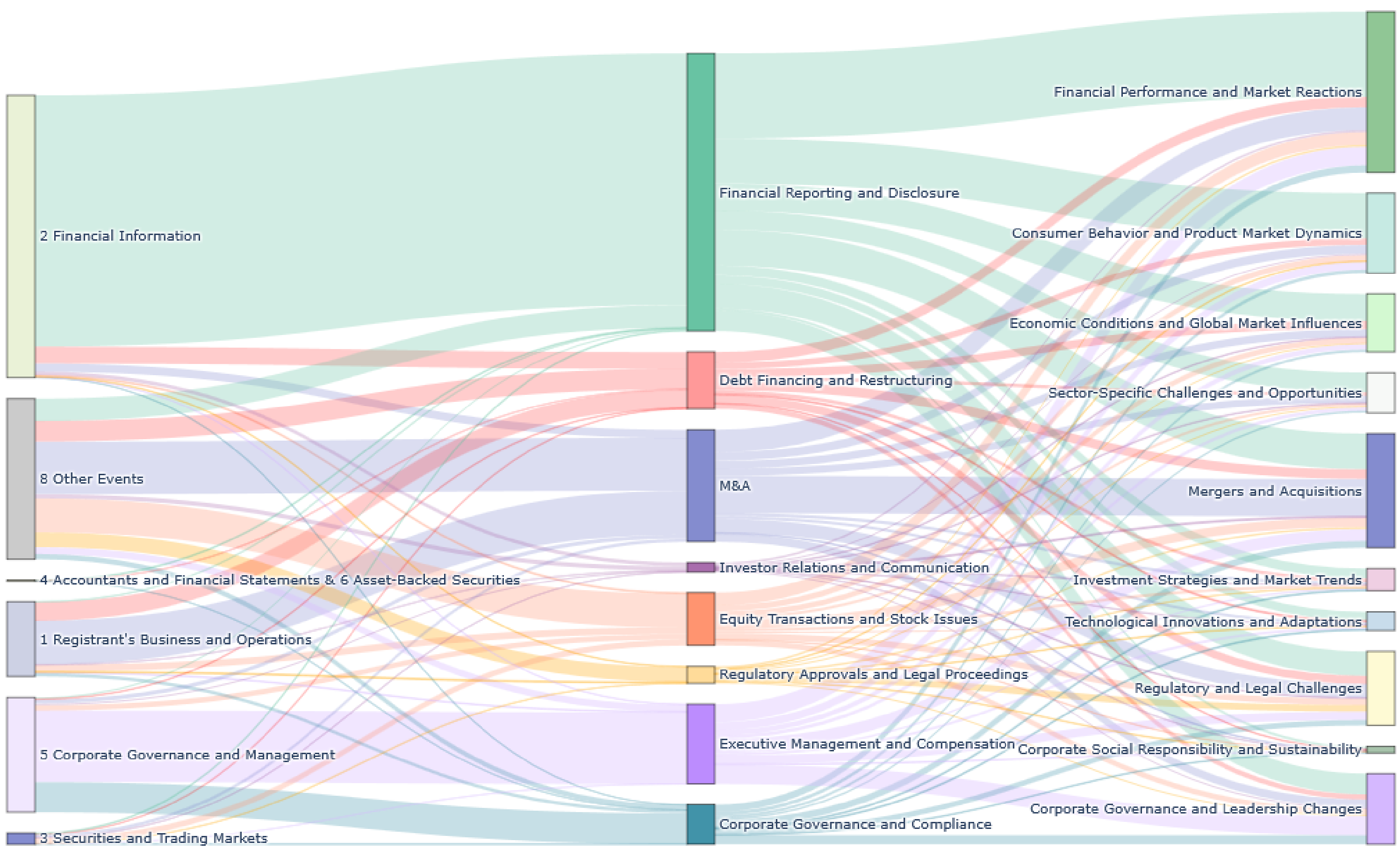

Table 1. Summary statistics

Panel A presents the summary statistics for the variables used in our analysis. The matched sample consists of 16,936 8-K filings filed between 2012 and 2021. The treated filings consist of those with disclosed corporate events covered by WSJ articles, while the control filings are a subset of non-treated filings (those with disclosed corporate events not covered by WSJ articles), selected to closely match the treated filings using propensity score matching based on a set of firm characteristics and 8-K items. We report the firm characteristics and the FinBERT sentiment scores of both the 8-K filings and the WSJ articles. All variables are defined in the Appendix A. Panel B compares the characteristics of the treated and control groups within the matched sample.

Panel A. Summary Statistics of the matched sample

Variable	N	Mean	S.D.	Min	0.250	Median	0.750	Max
Firm characteristics								
log(ME)	16,936	9.450	1.710	1.290	8.300	9.580	10.680	14.250
Book-to-Market	16,936	0.420	0.480	-12.120	0.160	0.340	0.600	7.280
Z Score	16,936	1.630	1.820	-57.200	0.900	1.620	2.450	15.890
Market Leverage	16,936	0.240	0.190	0.000	0.100	0.190	0.340	0.990
Institutional Ownership	16,936	0.750	0.240	0.000	0.670	0.800	0.900	3.400
S&P 1500	16,936	0.790	0.410	0.000	1.000	1.000	1.000	1.000
ROA	16,936	0.120	0.150	-4.500	0.080	0.130	0.180	0.900
Investment-to-assets	16,936	0.130	0.270	-0.170	0.040	0.070	0.150	7.400
Tangibility	16,936	0.300	0.250	0.000	0.090	0.200	0.480	0.970
Momentum	16,936	0.080	0.500	-0.920	-0.060	0.050	0.170	47.650
Fluidity	16,936	6.010	3.160	0.510	3.650	5.340	7.680	22.650
log(Illiq)	16,936	-22.700	1.960	-27.220	-24.060	-23.100	-21.690	-7.830
Bert Sentiment								
8-K Bert	16,936	0.000	0.060	-0.800	0.000	0.000	0.000	1.000
WSJ Bert	8,314	0.010	0.180	-1.000	-0.090	0.000	0.120	0.750

Panel B. Characteristics of treated group and control group

	Mean			
	Treatment	Control	Diff.	t-stat
log(ME)	9.44	9.46	-0.02	-0.62
Book-to-Market	0.42	0.42	0.00	0.09
Z Score	1.67	1.58	0.08	2.91
Market Leverage	0.24	0.23	0.01	2.49
Institutional Ownership	0.76	0.74	0.02	4.68
S&P 1500	0.79	0.78	0.01	1.62
ROA	0.13	0.12	0.01	3.15
Investment-to-assets	0.12	0.13	-0.01	-2.55
Tangibility	0.30	0.29	0.00	0.8
Momentum	0.08	0.08	0.00	0.08
Fluidity	5.99	6.03	-0.04	-0.84
log(Illiq)	-22.73	-22.68	-0.06	-1.87

Table 2. Impact of WSJ coverage of corporate events on short-term stock performance

This table presents short-term effects of WSJ coverage on stock returns at filing level using the matched sample over the period from 2012 to 2021. The treated filings consist of those with disclosed corporate events covered by WSJ articles, while the control filings are a subset of non-treated filings selected to closely match the treated filings based on a set of firm characteristics and 8-K items. The dependent variable $CAR[-5, k]$ is the cumulative abnormal return from the five trading days prior to the event date to the k trading days following the event date relative to Carhart (1997) four-factor asset pricing model. WSJ_{pos} (WSJ_{neg}) denotes the interaction between the absolute value of the WSJ FinBERT sentiment score and a positive (negative) sentiment dummy, where the dummy equals one if the WSJ sentiment is classified as positive (negative). $8K_{pos}$ ($8K_{neg}$) is defined as the interaction between the absolute value of the 8-K FinBERT sentiment score and a positive (negative) sentiment dummy, with the dummy equal to one if the 8-K sentiment is positive (negative). Panel A presents the regression results centered around the 8-K filing date, while Panel B shows the regression results centered around the publication date of the WSJ article. All regressions include controls for 8-K sentiment, filing items, and firm characteristics. All variables are defined in the Appendix A. Standard errors are adjusted for heteroskedasticity and clustered at the firm level, with t-statistics reported in parentheses below the coefficients. Coefficients marked with ***, **, and * are significant at the 1%, 5%, and 10% levels, respectively.

Panel A. CARs around 8K filing date

VARIABLES	(1) CAR [-5]	(2) CAR [-5, -3]	(3) CAR [-5, -2]	(4) CAR [-5, -1]	(5) CAR [-5, 0]	(6) CAR [-5, 1]	(7) CAR [-5, 2]	(8) CAR [-5, 3]	(9) CAR [-5, 5]
Treat×WSJ$_{Pos}$	0.229	0.833**	1.033***	0.795***	5.022***	9.525***	9.347***	9.665***	9.664***
	(1.32)	(2.49)	(2.86)	(2.94)	(6.24)	(6.03)	(5.95)	(6.10)	(5.93)
Treat×WSJ$_{Neg}$	-0.282	-1.725**	-1.699**	-1.109**	-6.711***	-9.693***	-9.871***	-10.290***	-10.805***
	(-0.84)	(-2.36)	(-2.41)	(-2.03)	(-7.41)	(-8.36)	(-8.10)	(-8.28)	(-8.09)
Treat×8K$_{Pos}$	0.467	0.627	-0.286	1.335	4.288**	6.722*	6.609*	6.815*	5.239
	(0.61)	(0.44)	(-0.16)	(1.07)	(2.09)	(1.93)	(1.83)	(1.83)	(1.34)
Treat×8K$_{Neg}$	0.959	-1.425	-1.396	-0.086	-0.311	-1.261	-0.945	0.617	-2.628
	(0.63)	(-0.57)	(-0.58)	(-0.04)	(-0.10)	(-0.38)	(-0.27)	(0.18)	(-0.71)
8K$_{Pos}$	-0.758	-0.564	-0.565	-0.796	-1.784	-1.508	-2.041	-2.087	-1.924
	(-1.33)	(-0.74)	(-0.69)	(-1.13)	(-1.50)	(-1.14)	(-1.41)	(-1.34)	(-1.19)
8K$_{Neg}$	-1.140	-1.651	-2.059*	-1.634	-2.761*	-3.261*	-3.416*	-4.687**	-2.330
	(-1.43)	(-1.30)	(-1.71)	(-1.52)	(-1.73)	(-1.91)	(-1.94)	(-2.45)	(-1.21)
Controls	Y	Y	Y	Y	Y	Y	Y	Y	Y
Industry FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
Year FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
Observations	16,901	16,901	16,901	16,901	16,900	16,895	16,892	16,891	16,888
R-squared	0.005	0.059	0.290	0.782	0.519	0.397	0.385	0.355	0.330

Panel B. CARs around WSJ date

VARIABLES	(1) CAR [-5]	(2) CAR [-5, -3]	(3) CAR [-5, -2]	(4) CAR [-5, -1]	(5) CAR [-5, 0]	(6) CAR [-5, 1]	(7) CAR [-5, 2]	(8) CAR [-5, 3]	(9) CAR [-5, 5]
Treat×WSJ_{Pos}	0.223	0.830**	1.024***	0.790***	4.991***	9.524***	9.311***	9.642***	9.669***
	(1.29)	(2.48)	(2.84)	(2.92)	(6.20)	(6.03)	(5.92)	(6.08)	(5.93)
Treat×WSJ_{Neg}	-0.286	-1.735**	-1.712**	-1.122**	-6.747***	-9.708***	-9.913***	-10.302***	-10.800***
	(-0.85)	(-2.37)	(-2.42)	(-2.06)	(-7.44)	(-8.37)	(-8.14)	(-8.29)	(-8.08)
Treat×$8K_{Pos}$	0.496	0.948	0.100	1.623	4.622**	7.085**	7.008*	7.386**	6.003
	(0.64)	(0.65)	(0.06)	(1.28)	(2.24)	(2.02)	(1.92)	(1.96)	(1.52)
Treat×$8K_{Pos}$	0.574	-2.283	-2.159	-0.684	-0.690	-1.968	-1.511	0.040	-3.261
	(0.38)	(-0.92)	(-0.90)	(-0.32)	(-0.23)	(-0.60)	(-0.44)	(0.01)	(-0.88)
$8K_{Pos}$	-0.778	-0.879	-0.947	-1.072	-2.116*	-1.852	-2.439	-2.643*	-2.659
	(-1.34)	(-1.06)	(-1.07)	(-1.42)	(-1.74)	(-1.36)	(-1.62)	(-1.65)	(-1.55)
$8K_{Neg}$	-0.748	-0.811	-1.305	-1.065	-2.428	-2.566	-2.855	-4.080**	-1.678
	(-1.00)	(-0.68)	(-1.15)	(-1.04)	(-1.57)	(-1.56)	(-1.64)	(-2.15)	(-0.88)
Controls	Y	Y	Y	Y	Y	Y	Y	Y	Y
Industry FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
Year FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
Observations	16,900	16,900	16,900	16,900	16,899	16,894	16,891	16,890	16,887
R-squared	0.005	0.058	0.289	0.782	0.519	0.397	0.384	0.354	0.329

Table 3. Impact of WSJ coverage of corporate events on stock performance over one-year horizon

This table presents long-term effects of WSJ coverage on stock returns at filing level using the matched sample over the period from 2012 to 2021. The treated filings consist of those with disclosed corporate events covered by WSJ articles, while the control filings are a subset of non-treated filings selected to closely match the treated filings based on a set of firm characteristics and 8-K items. The dependent variable $CAR[-5, k]$ is the cumulative abnormal return from the five trading days prior to the event date to the k trading days following the event date relative to Carhart (1997) four-factor asset pricing model. WSJ_{pos} (WSJ_{neg}) denotes the interaction between the absolute value of the WSJ FinBERT sentiment score and a positive (negative) sentiment dummy, where the dummy equals one if the WSJ sentiment is classified as positive (negative). $8K_{pos}$ ($8K_{neg}$) is defined as the interaction between the absolute value of the 8-K FinBERT sentiment score and a positive (negative) sentiment dummy, with the dummy equal to one if the 8-K sentiment is positive (negative). Panel A presents the regression results centered around the 8-K filing date, while Panel B shows the regression results centered around the publication date of the WSJ article. All regressions include controls for 8-K sentiment, filing items, and firm characteristics. All variables are defined in the Appendix A. Standard errors are adjusted for heteroskedasticity and clustered at the firm level, with t-statistics reported in parentheses below the coefficients. Coefficients marked with ***, **, and * are significant at the 1%, 5%, and 10% levels, respectively.

Panel A. CARs around 8K filing date

VARIABLES	(1) CAR [-5, 5]	(2) CAR [-5, 10]	(3) CAR [-5, 21]	(4) CAR [-5, 42]	(5) CAR [-5, 63]	(6) CAR [-5, 126]	(7) CAR [-5, 252]
Treat×WSJ_{Pos}	9.664***	9.783***	10.063***	10.919***	10.389***	10.808***	14.106***
	(5.93)	(5.63)	(5.56)	(5.38)	(4.83)	(3.63)	(3.52)
Treat×WSJ_{Neg}	-10.805***	-10.227***	-8.617***	-10.219***	-9.630***	-6.583	-2.149
	(-8.09)	(-7.21)	(-5.56)	(-4.52)	(-3.43)	(-1.55)	(-0.32)
Treat×$8K_{Pos}$	5.239	3.253	9.711**	9.587	7.376	18.916	20.481
	(1.34)	(0.79)	(2.06)	(1.51)	(0.87)	(1.60)	(1.17)
Treat×$8K_{Neg}$	-2.628	-3.187	0.482	-2.824	-6.676	-1.988	20.544
	(-0.71)	(-0.68)	(0.10)	(-0.49)	(-0.99)	(-0.22)	(1.42)
$8K_{Pos}$	-1.924	-1.658	-2.770	-0.624	0.093	-4.752	0.005
	(-1.19)	(-0.91)	(-1.24)	(-0.09)	(0.01)	(-0.63)	(0.00)
$8K_{Neg}$	-2.330	-2.120	-1.573	1.771	5.512	2.609	-14.620
	(-1.21)	(-0.90)	(-0.52)	(0.53)	(1.33)	(0.46)	(-1.63)
Controls	Y	Y	Y	Y	Y	Y	Y
Industry FE	Y	Y	Y	Y	Y	Y	Y
Year FE	Y	Y	Y	Y	Y	Y	Y
Observations	16,888	16,869	16,817	16,676	16,550	16,145	15,162
R-squared	0.330	0.287	0.240	0.180	0.153	0.117	0.072

Panel B. CARs around WSJ date

VARIABLES	(1) CAR [-5, 5]	(2) CAR [-5, 10]	(3) CAR [-5, 21]	(4) CAR [-5, 42]	(5) CAR [-5, 63]	(6) CAR [-5, 126]	(7) CAR [-5, 252]
Treat×WSJ$_{Pos}$	9.669***	9.762***	10.012***	10.854***	10.334***	10.768***	14.003***
	(5.93)	(5.62)	(5.53)	(5.35)	(4.81)	(3.61)	(3.50)
Treat×WSJ$_{Neg}$	-10.800***	-10.242***	-8.690***	-10.296***	-9.695***	-6.644	-2.228
	(-8.08)	(-7.22)	(-5.62)	(-4.56)	(-3.46)	(-1.56)	(-0.33)
Treat×8K$_{Pos}$	6.003	4.162	10.789**	10.837*	8.816	20.042*	24.138
	(1.52)	(1.00)	(2.28)	(1.67)	(1.02)	(1.67)	(1.32)
Treat×8K$_{Neg}$	-3.261	-3.947	0.149	-3.154	-7.055	-1.897	20.649
	(-0.88)	(-0.84)	(0.03)	(-0.55)	(-1.05)	(-0.21)	(1.43)
8K$_{Pos}$	-2.659	-2.562	-3.883*	-1.887	-1.360	-5.856	-3.703
	(-1.55)	(-1.31)	(-1.69)	(-0.27)	(-0.18)	(-0.75)	(-0.29)
8K$_{Neg}$	-1.678	-1.288	-1.278	2.047	5.878	2.503	-14.802*
	(-0.88)	(-0.56)	(-0.43)	(0.61)	(1.44)	(0.44)	(-1.66)
Controls	Y	Y	Y	Y	Y	Y	Y
Industry FE	Y	Y	Y	Y	Y	Y	Y
Year FE	Y	Y	Y	Y	Y	Y	Y
Observations	16,887	16,867	16,816	16,675	16,549	16,144	15,161
R-squared	0.329	0.285	0.238	0.180	0.153	0.117	0.072

Table 4. Impact of WSJ coverage of corporate events on operating performance

This table presents the effects of WSJ coverage on firm's operating performance at filing level using the matched sample over the period from 2012 to 2021. The treated filings consist of those with disclosed corporate events covered by WSJ articles, while the control filings are a subset of non-treated filings selected to closely match the treated filings based on a set of firm characteristics and 8-K items. The dependent variables, *ROA*, are the cumulative return on assets (ROA) calculated from the fiscal quarter following the event $t+1$ through the fourth fiscal quarter after the event $t+4$. WSJ_{pos} (WSJ_{neg}) denotes the interaction between the absolute value of the WSJ FinBERT sentiment score and a positive (negative) sentiment dummy, where the dummy equals one if the WSJ sentiment is classified as positive (negative). $8K_{pos}$ ($8K_{neg}$) is defined as the interaction between the absolute value of the 8-K FinBERT sentiment score and a positive (negative) sentiment dummy, with the dummy equal to one if the 8-K sentiment is positive (negative). All regressions include controls for 8-K sentiment, filing items, and firm characteristics. All variables are defined in the Appendix A. Standard errors are adjusted for heteroskedasticity and clustered at the firm level, with t-statistics reported in parentheses below the coefficients. Coefficients marked with ***, **, and * are significant at the 1%, 5%, and 10% levels, respectively.

VARIABLES	(1) ROA [t+1]	(2) ROA [t+1, t+2]	(3) ROA [t+1, t+3]	(4) ROA [t+1, t+4]
Treat×WSJ$_{Pos}$	1.068***	1.695***	2.787***	3.654**
	(3.72)	(2.66)	(2.64)	(2.48)
Treat×WSJ$_{Neg}$	-1.484***	-3.108***	-4.410***	-4.113***
	(-3.87)	(-4.24)	(-4.40)	(-3.02)
Treat×8K$_{Pos}$	1.503*	4.576**	6.885**	11.476**
	(1.79)	(2.31)	(2.23)	(1.97)
Treat×8K$_{Neg}$	-0.753	-0.938	-0.411	-1.224
	(-0.85)	(-0.51)	(-0.15)	(-0.34)
8K$_{Pos}$	-0.585	-1.607	-3.416	-6.458
	(-0.88)	(-0.95)	(-1.11)	(-1.35)
8K$_{Neg}$	-0.519	-1.841	-4.014	-5.883*
	(-0.76)	(-1.13)	(-1.55)	(-1.68)
Controls	Y	Y	Y	Y
Industry FE	Y	Y	Y	Y
Year FE	Y	Y	Y	Y
Observations	16,398	15,937	15,431	14,901
R-squared	0.512	0.516	0.521	0.510

Table 5. Decomposed impact of WSJ sentiment on stock and operating performance

This table reports the decomposed effects of WSJ sentiment across information categories on firms' stock market and operating performance. Each WSJ article in the matched sample is segmented using ChatGPT into three sections: common information, additional information, and irrelevant information. Sentiment for each section is measured using the FinBERT model, with section-level sentiment scores calculated as the average across all sentences within the section. Panel A presents the decomposed effects on stock returns, where the dependent variable is $CAR[-5, k]$, the cumulative abnormal return from five trading days prior to the filing date through k rading days after the filing date, estimated relative to the four-factor asset pricing model of Carhart (1997). Panel B reports the decomposed effects on operating performance, where the dependent variable is cumulative return on assets (ROA), defined as the sum of quarterly ROA from the first fiscal quarter following the filing date $t+1$ through the fourth subsequent quarter $t+4$. $Treat$ is a dummy equal to one if the 8-K filing discloses event(s) covered by the WSJ. $Common_{pos}$ ($Common_{neg}$) is the interaction between the absolute value of the FinBERT sentiment score for the WSJ article's common information section and a positive (negative) sentiment dummy, with the dummy equal to one if that section is classified as positive (negative). $Additional_{pos}$ ($Additional_{neg}$) and $Irrelevant_{pos}$ ($Irrelevant_{neg}$) are the interactions between the absolute value of the FinBERT sentiment score for the additional information and irrelevant information sections, respectively, and their corresponding positive (negative) sentiment dummies. All regressions include controls for 8-K sentiment, filing items, and firm characteristics. All variables are defined in the Appendix A. Standard errors are adjusted for heteroskedasticity and clustered at the firm level, with t-statistics reported in parentheses below the coefficients. Coefficients marked with ***, **, and * are significant at the 1%, 5%, and 10% levels, respectively.

Panel A. Stock Performance

VARIABLES	(1) CAR [-5, 5]	(2) CAR [-5, 10]	(3) CAR [-5, 21]	(4) CAR [-5, 42]	(5) CAR [-5, 63]	(6) CAR [-5, 126]	(7) CAR [-5, 252]
Treat×Common$_{pos}$	3.012***	3.031***	3.036***	2.723***	2.287***	2.986***	3.579**
	(5.23)	(5.07)	(4.73)	(3.79)	(3.02)	(2.91)	(2.52)
Treat×Common$_{neg}$	-3.419***	-3.075***	-3.004***	-3.242***	-3.620***	-2.979***	-3.057
	(-6.61)	(-5.77)	(-5.26)	(-4.30)	(-4.00)	(-2.62)	(-1.59)
Treat×Additional$_{pos}$	3.589***	3.740***	3.384***	4.342***	4.934***	3.721*	5.147*
	(3.63)	(3.44)	(3.13)	(2.87)	(2.97)	(1.77)	(1.91)
Treat×Additional$_{neg}$	-3.198***	-3.184***	-2.756***	-3.849***	-2.741*	-2.622	1.644
	(-3.75)	(-3.76)	(-2.88)	(-3.34)	(-1.84)	(-1.16)	(0.44)
Treat×Irrelevant$_{pos}$	2.955	2.593	5.351**	4.545	5.772	7.009	4.832
	(1.59)	(1.32)	(2.29)	(1.53)	(1.62)	(1.64)	(0.68)
Treat×Irrelevant$_{neg}$	-1.687	-2.823*	-1.884	-0.731	-2.596	-1.406	-1.925
	(-1.19)	(-1.82)	(-1.20)	(-0.34)	(-0.98)	(-0.41)	(-0.41)
Treat×8K$_{pos}$	4.235	2.317	9.450**	8.506	6.066	17.852	18.394
	(1.05)	(0.54)	(1.96)	(1.31)	(0.72)	(1.52)	(1.06)
Treat×8K$_{neg}$	-3.943	-3.522	0.732	-1.305	-6.847	-0.947	22.087
	(-1.06)	(-0.76)	(0.16)	(-0.23)	(-1.00)	(-0.10)	(1.50)
Controls	Y	Y	Y	Y	Y	Y	Y
Industry FE	Y	Y	Y	Y	Y	Y	Y
Year FE	Y	Y	Y	Y	Y	Y	Y
Observations	16,888	16,869	16,816	16,675	16,549	16,143	15,161
R-squared	0.332	0.288	0.241	0.182	0.155	0.118	0.072

Panel B. Operating Performance

VARIABLES	(1) ROA [t+1]	(2) ROA [t+1, t+2]	(3) ROA [t+1, t+3]	(4) ROA [t+1, t+4]
Treat×Common$_{pos}$	0.193*	0.435**	0.809**	1.102**
	(1.94)	(2.09)	(2.14)	(2.37)
Treat×Common$_{neg}$	-0.433***	-0.778***	-0.995***	-1.026**
	(-3.52)	(-3.21)	(-2.81)	(-2.06)
Treat×Additional$_{pos}$	0.415**	0.666*	0.950**	1.190*
	(2.53)	(1.91)	(2.23)	(1.88)
Treat×Additional$_{neg}$	-0.790***	-1.663***	-2.502***	-2.544***
	(-4.45)	(-5.20)	(-5.42)	(-4.83)
Treat×Irrelevant$_{pos}$	-0.051	-0.111	0.359	1.189
	(-0.16)	(-0.18)	(0.34)	(0.48)
Treat×Irrelevant$_{neg}$	-0.335	-0.813	-1.045	-1.117
	(-0.89)	(-1.06)	(-0.88)	(-0.68)
Treat×8K$_{pos}$	1.735**	4.898**	7.282**	11.911**
	(2.06)	(2.48)	(2.35)	(2.08)
Treat×8K$_{neg}$	-0.307	-0.265	0.394	0.180
	(-0.35)	(-0.15)	(0.14)	(0.05)
Controls	Y	Y	Y	Y
Industry FE	Y	Y	Y	Y
Year FE	Y	Y	Y	Y
Observations	16,398	15,937	15,431	14,901
R-squared	0.512	0.517	0.522	0.511

Table 6. Heterogeneity in the impact of WSJ sentiment on stock performance

This table reports the long-term effects of WSJ sentiment on stock performance, with results sorted by firm size (Panel A), analyst coverage (Panel B), and liquidity (Panel C). In each panel, the matched sample is divided into two groups based on the median values of $\log(ME)$, analyst coverage, and stock liquidity, respectively. The tests are conducted at the filing level using the matched sample from 2012 to 2021. The treated filings consist of those with disclosed corporate events covered by WSJ articles, while the control filings are a subset of non-treated filings selected to closely match the treated filings based on a set of firm characteristics and 8-K items. The dependent variable $CAR[-5,k]$ is the cumulative abnormal return from the five trading days prior to the event date to the k trading days following the event date relative to Carhart (1997) four-factor asset pricing model. WSJ_{pos} (WSJ_{neg}) denotes the interaction between the absolute value of the WSJ FinBERT sentiment score and a positive (negative) sentiment dummy, where the dummy equals one if the WSJ sentiment is classified as positive (negative). $8K_{pos}$ ($8K_{neg}$) is defined as the interaction between the absolute value of the 8-K FinBERT sentiment score and a positive (negative) sentiment dummy, with the dummy equal to one if the 8-K sentiment is positive (negative). All regressions include controls for 8-K sentiment, filing items, and firm characteristics. All variables are defined in Appendix A. Standard errors are adjusted for heteroskedasticity and clustered at the firm level, with t-statistics reported in parentheses below the coefficients. Coefficients marked with ***, **, and * are significant at the 1%, 5%, and 10% levels, respectively.

Panel A. log(ME)

VARIABLES	(1) CAR [-5, 5]	(2) CAR [-5, 10]	(3) CAR [-5, 21]	(4) CAR [-5, 42]	(5) CAR [-5, 63]	(6) CAR [-5, 126]	(7) CAR [-5, 252]
Low log(ME)							
Treat×WSJ$_{Pos}$	14.394***	14.267***	13.833***	15.601***	15.216***	18.518***	22.560***
	(4.56)	(4.29)	(3.98)	(4.02)	(3.80)	(3.57)	(3.21)
Treat×WSJ$_{Neg}$	-15.700***	-15.020***	-11.836***	-12.959***	-10.851**	-4.894	0.306
	(-6.97)	(-6.22)	(-4.77)	(-3.77)	(-2.39)	(-0.68)	(0.03)
High log(ME)							
Treat×WSJ$_{Pos}$	4.649***	4.788***	5.394***	5.398***	4.818***	2.328	3.350
	(5.45)	(4.95)	(5.48)	(4.10)	(3.16)	(0.92)	(0.95)
Treat×WSJ$_{Neg}$	-5.447***	-5.234***	-5.460***	-7.496***	-8.017***	-7.753**	-2.886
	(-5.16)	(-4.49)	(-3.51)	(-3.17)	(-3.21)	(-2.14)	(-0.66)
Controls	Y	Y	Y	Y	Y	Y	Y
Industry FE	Y	Y	Y	Y	Y	Y	Y
Year FE	Y	Y	Y	Y	Y	Y	Y
Diff in coefficient on Treat×WSJ$_{Pos}$	9.835***	9.570***	8.472**	10.215**	10.355**	16.063**	19.261**
	(3.01)	(2.77)	(2.35)	(2.48)	(2.41)	(2.78)	(2.45)
Diff in coefficient on Treat×WSJ$_{Neg}$	-10.186***	-9.700***	-6.446**	-5.443	18.804***	2.700	3.189
	(-4.10)	(-3.62)	(-2.20)	(-1.30)	(3.63)	(0.34)	(0.25)

Panel B. Analyst Coverage

VARIABLES	(1) CAR [-5, 5]	(2) CAR [-5, 10]	(3) CAR [-5, 21]	(4) CAR [-5, 42]	(5) CAR [-5, 63]	(6) CAR [-5, 126]	(7) CAR [-5, 252]
Low Analyst Coverage							
Treat×WSJ_{Pos}	13.805***	13.524***	15.269***	16.668***	15.601***	19.325***	22.349***
	(4.35)	(4.08)	(4.51)	(4.45)	(4.20)	(4.09)	(3.79)
Treat×WSJ_{Neg}	-13.634***	-12.176***	-10.114***	-8.645**	-9.756**	-6.247	-0.402
	(-6.51)	(-5.48)	(-4.31)	(-2.35)	(-2.15)	(-0.97)	(-0.05)
High Analyst Coverage							
Treat×WSJ_{Pos}	6.324***	6.428***	5.583***	6.070***	6.023***	4.177	7.608*
	(4.99)	(4.63)	(3.54)	(3.25)	(2.88)	(1.34)	(1.71)
Treat×WSJ_{Neg}	-8.904***	-9.271***	-8.202***	-12.171***	-10.771***	-7.227	-4.600
	(-5.76)	(-5.55)	(-4.37)	(-4.63)	(-3.32)	(-1.49)	(-0.58)
Controls	Y	Y	Y	Y	Y	Y	Y
Industry FE	Y	Y	Y	Y	Y	Y	Y
Year FE	Y	Y	Y	Y	Y	Y	Y
Diff in coefficient on Treat×WSJ_{Pos}	7.529**	7.098**	9.653***	10.559**	9.616**	15.128***	14.763**
	(2.20)	(1.98)	(2.58)	(2.52)	(2.26)	(2.67)	(2.00)
Diff in coefficient on Treat×WSJ_{Neg}	-4.675*	-2.932	-2.021	3.451	1.012	0.866	4.122
	(-1.80)	(-1.06)	(-0.67)	(0.76)	(0.18)	(0.11)	(0.35)

Panel C. Liquidity

VARIABLES	(1) CAR [-5, 5]	(2) CAR [-5, 10]	(3) CAR [-5, 21]	(4) CAR [-5, 42]	(5) CAR [-5, 63]	(6) CAR [-5, 126]	(7) CAR [-5, 252]
Low Liquidity							
Trea×WSJ_{Pos}	14.791***	14.837***	14.436***	15.897***	15.346***	19.473***	23.301***
	(4.62)	(4.40)	(4.13)	(4.09)	(3.86)	(3.74)	(3.32)
Treat×WSJ_{Neg}	-14.367***	-13.453***	-11.310***	-13.024***	-10.416**	-3.986	-2.215
	(-6.48)	(-5.57)	(-4.72)	(-4.16)	(-2.52)	(-0.60)	(-0.19)
High Liquidity							
Treat×WSJ_{Pos}	4.381***	4.341***	4.786***	4.818***	4.248***	0.712	2.206
	(5.09)	(4.43)	(4.68)	(3.48)	(2.65)	(0.27)	(0.58)
Treat×WSJ_{Neg}	-6.906***	-6.870***	-6.127***	-7.660***	-9.086***	-9.219**	-0.994
	(-6.02)	(-5.58)	(-3.20)	(-2.63)	(-3.19)	(-2.53)	(-0.22)
Controls	Y	Y	Y	Y	Y	Y	Y
Industry FE	Y	Y	Y	Y	Y	Y	Y
Year FE	Y	Y	Y	Y	Y	Y	Y
Diff in coefficient on Treat×WSJ_{Pos}	10.408***	10.501***	9.745***	11.216***	11.278***	18.804***	21.218***
	(3.14)	(2.99)	(2.67)	(2.70)	(2.62)	(3.21)	(2.65)
Diff in coefficient on Treat×WSJ_{Neg}	-7.417***	-6.542**	-5.131*	-5.236	-1.177	5.235	-1.101
	(-2.97)	(-2.41)	(-1.67)	(-1.22)	(-0.23)	(0.69)	(-0.09)

Table 7. Heterogeneity in the decomposed impact of WSJ sentiment on stock performance

This table reports the heterogeneity in the decomposed effects of WSJ sentiment across information categories on firms' stock performance, with results sorted by firm size (Panel A), analyst coverage (Panel B), and liquidity (Panel C). In each panel, the matched sample is divided into two groups based on the median values of $\log(ME)$, analyst coverage, and stock liquidity, respectively. Each WSJ article in the matched sample is segmented using ChatGPT into three sections: common information, additional information, and irrelevant information. Sentiment for each section is measured using the FinBERT model, with section-level sentiment scores calculated as the average across all sentences within the section. The dependent variable is $CAR[-5, k]$, the cumulative abnormal return from five trading days prior to the filing date through k rading days after the filing date, estimated relative to the four-factor asset pricing model of Carhart (1997). $Treat$ is a dummy equal to one if the 8-K filing discloses event(s) covered by the WSJ. $Common_{pos}$ ($Common_{neg}$) is the interaction between the absolute value of the FinBERT sentiment score for the WSJ article's common information section and a positive (negative) sentiment dummy, with the dummy equal to one if that section is classified as positive (negative). $Additional_{pos}$ ($Additional_{neg}$) and $Irrelevant_{pos}$ ($Irrelevant_{neg}$) are the interactions between the absolute value of the FinBERT sentiment score for the additional information and irrelevant information sections, respectively, and their corresponding positive (negative) sentiment dummies. All regressions include controls for 8-K sentiment, filing items, and firm characteristics. All variables are defined in the Appendix A. Standard errors are adjusted for heteroskedasticity and clustered at the firm level, with t-statistics reported in parentheses below the coefficients. Coefficients marked with ***, **, and * are significant at the 1%, 5%, and 10% levels, respectively.

Panel A. log(ME)

VARIABLES	(1) CAR [-5, 5]	(2) CAR [-5, 10]	(3) CAR [-5, 21]	(4) CAR [-5, 42]	(5) CAR [-5, 63]	(6) CAR [-5, 126]	(7) CAR [-5, 252]
Low log(ME)							
Treat×Common_{pos}	3.755***	3.589***	3.981***	3.717**	3.090*	4.762**	6.311**
	(3.05)	(2.81)	(2.89)	(2.40)	(1.95)	(2.33)	(2.45)
Treat×Common_{neg}	-5.340***	-4.954***	-4.538***	-5.036***	-4.542***	-2.867*	-3.389
	(-6.60)	(-6.20)	(-5.24)	(-4.42)	(-3.22)	(-1.69)	(-1.16)
Treat×Additional_{pos}	5.800***	6.011***	4.987**	6.339**	7.479**	6.284*	8.254*
	(3.12)	(2.95)	(2.50)	(2.25)	(2.47)	(1.65)	(1.72)
Treat×Additional_{neg}	-3.980***	-4.145***	-3.084**	-4.042**	-2.257	-3.623	1.138
	(-2.72)	(-2.93)	(-1.98)	(-2.23)	(-0.97)	(-1.01)	(0.18)
Treat×Irrelevant_{pos}	4.712	4.314	8.070	7.079	9.272	10.568	8.820
	(1.04)	(0.90)	(1.35)	(0.93)	(1.03)	(0.99)	(0.48)
Treat×Irrelevant_{neg}	-3.118	-5.256*	-5.134*	-7.950**	-8.364	-6.653	-0.922
	(-1.10)	(-1.73)	(-1.73)	(-2.22)	(-1.58)	(-0.99)	(-0.10)
High log(ME)							
Treat×Common_{pos}	2.272***	2.395***	2.047***	1.725***	1.444**	1.459	1.165
	(6.85)	(6.49)	(5.09)	(3.57)	(2.56)	(1.52)	(0.85)
Treat×Common_{neg}	-1.423***	-1.188**	-1.364**	-1.495*	-2.839***	-3.319**	-2.428
	(-2.97)	(-2.09)	(-2.18)	(-1.80)	(-2.97)	(-2.35)	(-1.19)
Treat×Additional_{pos}	1.193**	1.275**	1.431**	2.038**	2.255**	0.701	1.088
	(2.33)	(2.10)	(2.13)	(2.44)	(2.09)	(0.48)	(0.49)
Treat×Additional_{neg}	-1.975***	-1.826***	-2.228***	-3.277***	-2.672**	-0.730	4.485*
	(-3.71)	(-2.95)	(-2.72)	(-3.08)	(-2.00)	(-0.37)	(1.95)
Treat×Irrelevant_{pos}	2.153	1.673	3.934**	2.928	3.374*	4.466	2.636
	(1.46)	(1.02)	(2.50)	(1.64)	(1.94)	(1.45)	(0.63)
Treat×Irrelevant_{neg}	-1.025	-1.442	0.178	3.894*	0.999	1.069	-2.353
	(-0.77)	(-1.03)	(0.12)	(1.86)	(0.49)	(0.32)	(-0.57)
Diff							
Treat×Common_{pos}	1.483	1.195	1.934	1.992	1.646	3.304	5.146*
	(1.16)	(0.90)	(1.35)	(1.23)	(0.98)	(1.46)	(1.77)
Treat×Common_{neg}	-3.916***	-3.767***	-3.174***	-3.541**	-1.703	0.451	-0.961
	(-4.16)	(-3.84)	(-2.97)	(-2.52)	(-1.00)	(0.20)	(-0.27)
Treat×Additional_{pos}	4.607**	4.736**	3.556*	4.301	5.223	5.583	7.165
	(2.39)	(2.23)	(1.69)	(1.46)	(1.63)	(1.37)	(1.36)
Treat×Additional_{neg}	-2.005	-2.319	-0.857	-0.765	0.414	-2.893	-3.346
	(-1.29)	(-1.50)	(-0.49)	(-0.36)	(0.15)	(-0.70)	(-0.50)
Treat×Irrelevant_{pos}	2.559	2.641	4.135	4.15	5.898	6.103	6.184
	(0.54)	(0.52)	(0.67)	(0.53)	(0.64)	(0.55)	(0.33)
Treat×Irrelevant_{neg}	-2.093	-3.814	-5.312	-11.844***	-9.363*	-7.722	1.431
	(-0.67)	(-1.14)	(-1.61)	(-2.85)	(-1.65)	(-1.02)	(0.14)

Panel B. Analyst Coverage

VARIABLES	(1) CAR [-5, 5]	(2) CAR [-5, 10]	(3) CAR [-5, 21]	(4) CAR [-5, 42]	(5) CAR [-5, 63]	(6) CAR [-5, 126]	(7) CAR [-5, 252]
Low Coverage							
Treat×Common$_{pos}$	3.069***	2.816***	3.181***	3.427***	3.146***	3.505**	5.022**
	(4.33)	(3.72)	(3.82)	(3.44)	(2.97)	(2.47)	(2.55)
Treat×Common$_{neg}$	-4.181***	-3.802***	-4.373***	-4.071***	-4.780***	-4.078***	-4.068
	(-5.09)	(-4.79)	(-5.16)	(-3.48)	(-3.48)	(-2.63)	(-1.55)
Treat×Additional$_{pos}$	5.115***	4.981**	4.525**	5.490**	6.195**	6.652*	6.323
	(2.86)	(2.57)	(2.41)	(2.03)	(2.15)	(1.90)	(1.57)
Treat×Additional$_{neg}$	-4.710***	-5.223***	-3.993***	-4.787***	-2.694	-3.381	1.072
	(-3.85)	(-4.92)	(-3.18)	(-3.06)	(-1.31)	(-1.10)	(0.25)
Treat×Irrelevant$_{pos}$	6.716	5.049	6.445	3.117	2.159	6.153	11.229
	(1.49)	(1.07)	(1.13)	(0.42)	(0.25)	(0.63)	(0.81)
Treat×Irrelevant$_{neg}$	-5.011**	-6.570***	-4.279*	-4.547	-5.680	-3.967	-0.374
	(-2.53)	(-3.09)	(-1.80)	(-1.38)	(-1.28)	(-0.77)	(-0.05)
High Coverage							
Treat×Common$_{pos}$	2.917***	3.104***	2.821***	2.050*	1.386	2.117	1.606
	(3.28)	(3.44)	(2.94)	(1.94)	(1.25)	(1.48)	(0.84)
Treat×Common$_{neg}$	-2.906***	-2.633***	-2.048***	-2.686***	-2.780***	-2.275	-2.532
	(-4.80)	(-3.99)	(-2.85)	(-2.99)	(-2.64)	(-1.56)	(-1.08)
Treat×Additional$_{pos}$	1.993**	2.393***	2.102**	3.287***	3.936***	1.338	4.648
	(2.57)	(2.76)	(2.13)	(2.77)	(2.65)	(0.62)	(1.46)
Treat×Additional$_{neg}$	-1.995*	-1.537	-1.765	-3.055*	-2.705	-1.836	2.574
	(-1.69)	(-1.23)	(-1.26)	(-1.89)	(-1.36)	(-0.62)	(0.54)
Treat×Irrelevant$_{pos}$	0.557	0.853	4.518**	4.653**	7.180***	6.880*	0.071
	(0.35)	(0.48)	(2.32)	(2.07)	(2.76)	(1.71)	(0.01)
Treat×Irrelevant$_{neg}$	1.755	0.732	0.534	2.990	0.078	-0.198	-3.897
	(0.98)	(0.36)	(0.25)	(1.04)	(0.03)	(-0.05)	(-0.65)
Diff							
Treat×Common$_{pos}$	0.152	-0.288	0.360	1.377	1.760	1.388	3.416
	(0.13)	(-0.24)	(0.28)	(0.95)	(1.15)	(0.69)	(1.25)
Treat×Common$_{neg}$	-1.275	-1.169	-2.325**	-1.384	-2.000	-1.803	-1.535
	(-1.25)	(-1.13)	(-2.09)	(-0.94)	(-1.16)	(-0.85)	(-0.44)
Treat×Additional$_{pos}$	3.122	2.588	2.424	2.203	2.260	5.314	1.675
	(1.60)	(1.22)	(1.14)	(0.74)	(0.70)	(1.30)	(0.33)
Treat×Additional$_{neg}$	-2.715	-3.687**	-2.228	-1.732	0.011	-1.545	-1.503
	(-1.60)	(-2.24)	(-1.19)	(-0.77)	(0.00)	(-0.36)	(-0.23)
Treat×Irrelevant$_{pos}$	6.160	4.196	1.927	-1.536	-5.021	-0.727	11.158
	(1.29)	(0.83)	(0.32)	(-0.20)	(-0.56)	(-0.07)	(0.72)
Treat×Irrelevant$_{neg}$	-6.766**	-7.302**	-4.813	-7.537*	-5.758	-3.769	3.522
	(-2.53)	(-2.47)	(-1.51)	(-1.72)	(-1.07)	(-0.57)	(0.38)

Panel C. Liquidity

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	CAR	CAR	CAR	CAR	CAR	CAR	CAR
VARIABLES	[-5, 5]	[-5, 10]	[-5, 21]	[-5, 42]	[-5, 63]	[-5, 126]	[-5, 252]
Low Liquidity							
Treat×Common$_{pos}$	3.721***	3.674***	4.008***	3.821**	3.153**	4.542**	6.178**
	(3.03)	(2.93)	(3.00)	(2.57)	(2.08)	(2.33)	(2.40)
Treat×Common$_{neg}$	-5.103***	-4.711***	-4.332***	-4.720***	-4.508***	-3.332*	-4.946
	(-6.20)	(-5.79)	(-4.99)	(-4.12)	(-3.08)	(-1.96)	(-1.62)
Treat×Additional$_{pos}$	5.965***	5.951***	4.942**	6.085**	6.774**	6.111	7.254
	(3.18)	(2.90)	(2.46)	(2.14)	(2.22)	(1.60)	(1.52)
Treat×Additional$_{neg}$	-4.064***	-4.178***	-3.267**	-4.398**	-2.348	-2.712	0.546
	(-2.73)	(-2.94)	(-2.09)	(-2.38)	(-1.00)	(-0.77)	(0.09)
Treat×Irrelevant$_{pos}$	4.999	4.820	8.211	6.847	10.354	13.214	13.541
	(1.18)	(1.09)	(1.51)	(0.97)	(1.25)	(1.38)	(0.86)
Treat×Irrelevant$_{neg}$	-1.571	-4.313	-5.517*	-7.663**	-8.420*	-8.091	-6.371
	(-0.57)	(-1.43)	(-1.95)	(-2.25)	(-1.67)	(-1.21)	(-0.72)
High Liquidity							
Trea×Common$_{pos}$	2.236***	2.273***	1.956***	1.491***	1.220**	1.389	1.292
	(6.62)	(5.86)	(4.51)	(2.93)	(2.09)	(1.50)	(0.98)
Treat×Common$_{neg}$	-1.872***	-1.587***	-1.764**	-2.110**	-3.160***	-3.003**	-0.644
	(-3.72)	(-2.66)	(-2.44)	(-2.28)	(-3.30)	(-2.13)	(-0.36)
Treat×Additional$_{pos}$	1.151**	1.422**	1.505**	2.249***	2.822***	0.704	1.546
	(2.22)	(2.34)	(2.30)	(2.76)	(2.64)	(0.47)	(0.65)
Treat×Additional$_{neg}$	-1.794**	-1.604**	-1.932**	-2.686**	-2.547*	-1.836	4.176*
	(-2.43)	(-2.08)	(-2.04)	(-2.12)	(-1.83)	(-0.91)	(1.79)
Treat×Irrelevant$_{pos}$	1.509	0.950	3.269*	2.144	1.737	1.653	-1.556
	(0.99)	(0.55)	(1.94)	(1.10)	(0.95)	(0.56)	(-0.34)
Treat×Irrelevant$_{neg}$	-2.088	-2.064	0.259	3.499	0.871	2.047	1.236
	(-1.35)	(-1.23)	(0.17)	(1.48)	(0.39)	(0.63)	(0.29)
Diff							
Treat×Common$_{pos}$	1.485	1.401	2.052	2.330	1.933	3.153	4.885*
	(1.17)	(1.07)	(1.46)	(1.48)	(1.19)	(1.46)	(1.69)
Treat×Common$_{neg}$	-3.231***	-3.124***	-2.569**	-2.610*	-1.347	-0.329	-4.302
	(-3.35)	(-3.10)	(-2.27)	(-1.77)	(-0.77)	(-0.15)	(-1.22)
Treat×Additional$_{pos}$	4.815**	4.530**	3.437	3.836	3.952	5.407	5.707
	(2.47)	(2.12)	(1.63)	(1.30)	(1.22)	(1.32)	(1.07)
Treat×Additional$_{neg}$	-2.270	-2.573	-1.334	-1.711	0.199	-0.876	-3.629
	(-1.37)	(-1.59)	(-0.73)	(-0.76)	(0.07)	(-0.21)	(-0.55)
Treat×Irrelevant$_{pos}$	3.490	3.870	4.942	4.703	8.617	11.561	15.097
	(0.78)	(0.81)	(0.87)	(0.64)	(1.01)	(1.15)	(0.92)
Treat×Irrelevant$_{neg}$	0.516	-2.249	-5.776*	-11.162***	-9.291*	-10.138	-7.607
	(0.16)	(-0.65)	(-1.80)	(-2.69)	(-1.69)	(-1.37)	(-0.77)

Table 8. Impact of WSJ sentiment on stock performance: ChatGPT-based sentiment

This table reports the effects of WSJ sentiment on stock returns over a one-year horizon at the filing level, using the matched sample from 2012 to 2021. Sentiment is identified with ChatGPT based on the prompt provided in Table B.3 in Appendix B. The treated filings consist of those with disclosed corporate events covered by WSJ articles, while the control filings are a subset of non-treated filings selected to closely match the treated filings based on a set of firm characteristics and 8-K items. The dependent variable $CAR[-5, k]$ is the cumulative abnormal return from the five trading days prior to the event date to the k trading days following the event date relative to Carhart (1997) four-factor asset pricing model. WSJ_{pos} (WSJ_{neg}) denotes the interaction between the absolute value of the WSJ ChatGPT sentiment score and a positive (negative) sentiment dummy, where the dummy equals one if the WSJ sentiment is classified as positive (negative). $8K_{pos}$ ($8K_{neg}$) is defined as the interaction between the absolute value of the 8-K ChatGPT sentiment score and a positive (negative) sentiment dummy, with the dummy equal to one if the 8-K sentiment is positive (negative). All regressions include controls for 8-K sentiment, filing items, and firm characteristics. All variables are defined in the Appendix A. Standard errors are adjusted for heteroskedasticity and clustered at the firm level, with t-statistics reported in parentheses below the coefficients. Coefficients marked with ***, **, and * are significant at the 1%, 5%, and 10% levels, respectively.

VARIABLES	(1) CAR [-5, 5]	(2) CAR [-5, 10]	(3) CAR [-5, 21]	(4) CAR [-5, 42]	(5) CAR [-5, 63]	(6) CAR [-5, 126]	(7) CAR [-5, 252]
Treat×WSJ_{Pos}	4.059***	4.150***	3.909***	3.715***	3.579***	3.692***	5.052***
	(6.57)	(6.26)	(5.40)	(4.39)	(3.91)	(3.12)	(2.97)
Treat×WSJ_{Neg}	-7.505***	-7.338***	-6.382***	-7.382***	-7.726***	-5.911***	-2.690
	(-9.07)	(-8.71)	(-7.07)	(-6.53)	(-5.83)	(-2.96)	(-0.88)
Treat×$8K_{Pos}$	1.432**	1.504**	1.565**	2.188*	2.399	3.951*	7.696**
	(2.36)	(2.12)	(2.05)	(1.80)	(1.61)	(1.76)	(2.15)
Treat×$8K_{Neg}$	-0.749	0.557	0.834	0.349	-1.138	-4.037	0.708
	(-0.60)	(0.42)	(0.53)	(0.20)	(-0.53)	(-1.24)	(0.17)
$8K_{Pos}$	0.246	0.197	0.322	0.632	0.637	-0.426	-1.777
	(0.80)	(0.52)	(0.71)	(0.94)	(0.79)	(-0.34)	(-0.98)
$8K_{Neg}$	-1.870**	-3.093***	-1.657	-1.678	0.609	2.249	-2.959
	(-2.15)	(-3.28)	(-1.41)	(-1.28)	(0.36)	(0.84)	(-0.91)
Controls	Y	Y	Y	Y	Y	Y	Y
Industry FE	Y	Y	Y	Y	Y	Y	Y
Year FE	Y	Y	Y	Y	Y	Y	Y
Observations	16,889	16,870	16,817	16,676	16,550	16,144	15,162
R-squared	0.339	0.294	0.243	0.183	0.156	0.119	0.074

Table 9. Heterogeneity in the impact of WSJ sentiment on stock performance: ChatGPT-based sentiment

This table reports the effects of WSJ sentiment on stock performance, with results sorted by firm size (Panel A), analyst coverage (Panel B), and liquidity (Panel C). Sentiment is identified with ChatGPT based on the prompt provided in Table B.3 in Appendix B. The tests are conducted at the filing level using the matched sample from 2012 to 2021. The treated filings consist of those with disclosed corporate events covered by WSJ articles, while the control filings are a subset of non-treated filings selected to closely match the treated filings based on a set of firm characteristics and 8-K items. In each panel, the matched sample is divided into two groups based on the median values of $\log(ME)$, analyst coverage, and stock liquidity, respectively. The dependent variable $CAR[-5, k]$ is the cumulative abnormal return from the five trading days prior to the filing date to the k trading days following the filing date. WSJ_{pos} (WSJ_{neg}) denotes the interaction between the absolute value of the WSJ ChatGPT sentiment score and a positive (negative) sentiment dummy, where the dummy equals one if the WSJ sentiment is classified as positive (negative). $8K_{pos}$ ($8K_{neg}$) is defined as the interaction between the absolute value of the 8-K ChatGPT sentiment score and a positive (negative) sentiment dummy, with the dummy equal to one if the 8-K sentiment is positive (negative). All regressions include controls for 8-K sentiment, filing items, and firm characteristics. All variables are defined in Appendix A. Standard errors are adjusted for heteroskedasticity and clustered at the firm level, with t-statistics reported in parentheses below the coefficients. Coefficients marked with ***, **, and * are significant at the 1%, 5%, and 10% levels, respectively.

Panel A. log(ME)

VARIABLES	(1) CAR [-5, 5]	(2) CAR [-5, 10]	(3) CAR [-5, 21]	(4) CAR [-5, 42]	(5) CAR [-5, 63]	(6) CAR [-5, 126]	(7) CAR [-5, 252]
Low log(ME)							
Treat×WSJ$_{Pos}$	5.980***	6.051***	5.480***	5.549***	5.336***	5.901***	8.693***
	(5.32)	(5.06)	(4.29)	(3.74)	(3.35)	(2.99)	(3.10)
Treat×WSJ$_{Neg}$	-10.434***	-10.351***	-8.735***	-10.664***	-10.774***	-7.259**	-2.073
	(-7.87)	(-7.63)	(-5.98)	(-5.94)	(-5.24)	(-2.21)	(-0.39)
High log(ME)							
Treat×WSJ$_{Pos}$	1.843***	1.883***	2.004***	1.542***	1.529**	1.219	0.957
	(5.69)	(5.22)	(4.57)	(3.03)	(2.56)	(1.23)	(0.68)
Treat×WSJ$_{Neg}$	-3.676***	-3.580***	-3.355***	-3.238***	-3.689***	-3.730*	-1.585
	(-6.17)	(-5.71)	(-4.41)	(-3.00)	(-2.87)	(-1.93)	(-0.73)
Controls	Y	Y	Y	Y	Y	Y	Y
Industry FE	Y	Y	Y	Y	Y	Y	Y
Year FE	Y	Y	Y	Y	Y	Y	Y
Diff							
Treat×WSJ$_{Pos}$	4.137***	4.167***	3.476**	4.007**	3.807**	4.681**	7.736**
	(3.54)	(3.34)	(2.57)	(2.56)	(2.24)	(2.12)	(2.47)
Treat×WSJ$_{Neg}$	-6.758***	-6.771***	-5.380***	-7.426***	-7.085***	-3.529	-0.488
	(-4.65)	(-4.53)	(-3.27)	(-3.55)	(-2.92)	(-0.93)	(-0.09)

Panel B. Analyst Coverage

VARIABLES	(1) CAR [-5, 5]	(2) CAR [-5, 10]	(3) CAR [-5, 21]	(4) CAR [-5, 42]	(5) CAR [-5, 63]	(6) CAR [-5, 126]	(7) CAR [-5, 252]
Low Analyst Coverage							
Treat×WSJ_{Pos}	4.880***	4.811***	4.885***	5.265***	5.093***	5.429***	6.418***
	(5.44)	(5.04)	(4.75)	(4.22)	(3.79)	(3.20)	(3.00)
Treat×WSJ_{Neg}	-8.170***	-7.995***	-6.731***	-7.799***	-7.607***	-5.350*	0.804
	(-5.99)	(-6.27)	(-5.24)	(-4.47)	(-3.99)	(-1.65)	(0.20)
High Analyst Coverage							
Treat×WSJ_{Pos}	3.342***	3.451***	3.004***	2.398**	2.314**	2.300	4.379*
	(4.17)	(4.09)	(3.22)	(2.28)	(2.00)	(1.49)	(1.78)
Treat×WSJ_{Neg}	-6.758***	-6.668***	-5.919***	-6.725***	-7.621***	-6.005**	-5.134
	(-7.82)	(-6.76)	(-5.06)	(-4.78)	(-4.48)	(-2.52)	(-1.44)
Controls	Y	Y	Y	Y	Y	Y	Y
Industry FE	Y	Y	Y	Y	Y	Y	Y
Year FE	Y	Y	Y	Y	Y	Y	Y
Diff							
Treat×WSJ_{Pos}	1.538	1.360	1.881	2.867*	2.779	3.129	2.038
	(1.28)	(1.07)	(1.35)	(1.76)	(1.57)	(1.36)	(0.63)
Treat×WSJ_{Neg}	-1.413	-1.326	-0.812	-1.074	0.013	0.655	5.939
	(-0.88)	(-0.82)	(-0.47)	(-0.48)	(0.01)	(0.16)	(1.12)

Panel C. Liquidity

VARIABLES	(1) CAR [-5, 5]	(2) CAR [-5, 10]	(3) CAR [-5, 21]	(4) CAR [-5, 42]	(5) CAR [-5, 63]	(6) CAR [-5, 126]	(7) CAR [-5, 252]
Low Liquidity							
Treat×WSJ$_{Pos}$	6.056***	6.074***	5.578***	5.585***	5.235***	6.132***	8.885***
	(5.32)	(5.01)	(4.31)	(3.73)	(3.27)	(3.07)	(3.17)
Treat×WSJ$_{Neg}$	-9.621***	-9.416***	-7.645***	-9.520***	-9.761***	-6.103*	-2.132
	(-7.47)	(-7.02)	(-5.18)	(-5.28)	(-4.57)	(-1.86)	(-0.40)
High Liquidity							
Treat×WSJ$_{Pos}$	1.668***	1.791***	1.790***	1.376***	1.447**	0.748	0.804
	(5.77)	(5.42)	(4.39)	(2.78)	(2.52)	(0.78)	(0.57)
Treat×WSJ$_{Neg}$	-5.059***	-5.003***	-5.038***	-4.978***	-5.363***	-5.382***	-2.112
	(-7.20)	(-7.19)	(-6.10)	(-4.01)	(-3.85)	(-2.67)	(-0.97)
Controls	Y	Y	Y	Y	Y	Y	Y
Industry FE	Y	Y	Y	Y	Y	Y	Y
Year FE	Y	Y	Y	Y	Y	Y	Y
Diff							
Treat×WSJ$_{Pos}$	4.389***	4.282***	3.787***	4.209***	3.789**	5.385**	8.081**
	(3.73)	(3.41)	(2.79)	(2.67)	(2.23)	(2.43)	(2.57)
Treat×WSJ$_{Neg}$	-4.562***	-4.413***	-2.608	-4.542**	-4.398*	-0.721	-0.019
	(-3.11)	(-2.92)	(-1.54)	(-2.08)	(-1.73)	(-0.19)	(0.00)

Appendix A: Definitions of Variables

Variable	Description
Sentiment measures	
Bert Sentiment	We use FinBERT model to categorize the sentiment expressed in each individual sentence into positive, negative, and neutural. For the sentences classified as positive,negative, or neutural, we assign a sentiment score of 1, -1, or 0, respectively. The filing level or article level sentiment score is obtained by taking the average of the sentiment scores across all sentinces in the filing or article.
GPT Sentiment	We ask ChatGPT to to categorize the sentiment expressed in each text into one of the five sentiment levels: -1, -0.5, 0, 0.5, 1. -1 indicates very negative sentiment, -0.5 indicates mildly negative sentiment, 0 indicates neutral sentiment, 0.5 indicates mildly positive sentiment, and 1 indicates very positive sentiment.
Dependent variables	
CAR [-5, t]	Cumulative abnormal return over trading days -5 to t around the event day estimated using the Carhart (1997) four-factor model. Day 0 in event time is identified as 8-K filing date or WSJ article publish date.
ROA [t+1, t+n]	The cumulative EBITDA for the fiscal quarter following the corporate event, $t+1$, through the n-th fiscal quarter, $t+n$, scaled by total assets in the fiscal quarter immediately after the corporate event.
Firm characteristics	
log(ME)	Natural logarithm of market capitalization at the end of the year.
Book-to-Market	Book assets divided by (book assets −book equity + market equity) at the end of the year.
Z Score	Calculated as 3.3 × Operating Income Before Depreciation + Sales/Turnover + 1.4 × Retained Earnings + 1.2 × (Current Assets − Current Liabilities)/Assets at the end of the year.
Market Leverage	The sum of long-tem debt (*dlttq*) and short-term debt (*dlcq*) divided by the sum of long-term debt and short-term debt plus the market value of equity (*cshoq* × *prccq*) at the end of the year.
Institutional Ownership	The ratio between shares held by institutional investors from the Thomson-Reuters Institutional Holdings (13F) database and the total shares outstanding from the CRSP database in basis points at the end of the quarter.
S&P 1500 = 1	A dummy variable that takes the value of 1 if the stock is S&P 1500 stock and 0 otherwise.
ROA	EBITA divided by total assets at the end of the year.

Variable	Description
Investment-to-assets	The sum of capital expenditure (*capx*), acquisitions (*aqc*), and increase in investments (*ivch*) divided by total assets at the end of the year.
Tangibility	Net property, plant, and equipment divided by total assets at the end of the year.
Momentum	Past 12-month cumulative stock return.
Fluidity	Product Market Fuidity, a proxy to capture changes in rival firms' products in relation to the firm's products. See Hoberg, Phillips, and Prabhala (2014) for details on how they compute this variable using textual analysis.
log(Illiq)	Natural logarithm of the illiquidity measure of Amihud (2002) calculated as the square root of the absolute value of the daily return over daily dollar volume.
Analyst Coverage	Number of analysts covering a stock from I/B/E/S.
CAR[-3,-1]	Cumulative abnormal return over the three trading days to one trading day prior to the publication date of the WSJ article, estimated using the Carhart (1997) four-factor model.
Other variables	
Item 2.02	A dummy variable that equals 1 if the 8-K filing contains Item 2.02 and 0 otherwise.
Item 8.01	A dummy variable that equals 1 if the 8-K filing contains Item 8.01 and 0 otherwise.
Item 7.01	A dummy variable that equals 1 if the 8-K filing contains Item 7.01 and 0 otherwise.
Item 5.02	A dummy variable that equals 1 if the 8-K filing contains Item 5.02 and 0 otherwise.
Item 1.01	A dummy variable that equals 1 if the 8-K filing contains Item 1.01 and 0 otherwise.
Item 5.07	A dummy variable that equals 1 if the 8-K filing contains Item 5.07 and 0 otherwise.
Item 2.03	A dummy variable that equals 1 if the 8-K filing contains Item 2.03 and 0 otherwise.
Item 5.03	A dummy variable that equals 1 if the 8-K filing contains Item 5.03 and 0 otherwise.

Appendix B: Prompts and Examples

Table B.1. Prompt for identifying event and matching 8-K filing and WSJ article

This table presents the instructions given to ChatGPT for analyzing 8-K filings and WSJ articles to identify overlapping corporate events. *Text1* refers to the body of the 8-K filing, and *Text2* refers to the WSJ article. Each *Text1–Text2* pair meets two criteria: (1) the WSJ article references the same firm that filed the 8-K, and (2) the article's publication date falls within eight business days of the filing date. The prompt first summarizes the corporate events discussed in both texts, then determines whether they describe the same event.

Text1 and Text2 represent two discussions on the same corporate event. Acting as a financial expert, please analyze the two texts. Note that each text may discuss multiple events of different firms, and please only focus on the overlapping events. Your final goal is providing your answers to the following seven questions in a JSON format:

1. 'Text1': Provide a one-sentence summary of how Text1 addresses this corporate event.
2. 'Text2': Provide a one-sentence summary of how Text2 addresses this corporate event.
3. 'Same Event': Are Text1 and Text2 discussing at least one same event?

Structure the output in a JSON format:

```
{
  "Text1": "Text1 summarizes the event by...",
  "Text2": "Text2 summarizes the event by...",
  "Same Event": "Yes/No"
}
```

Table B.2. Prompt for decomposing WSJ article

This table presents the detailed instructions provided to ChatGPT for identifying the common corporate event discussed in 8-K filings and WSJ articles and for analyzing the information contained in the articles. *Text1* refers to the body of the 8-K filing, and *Text2* refers to the corresponding WSJ article. Each *Text1*–*Text2* pair covers at least one shared underlying corporate event. The prompt first identifies the event discussed in both texts and then determines whether the WSJ article focuses on that event. If the article does focus on the event, the prompt partitions its content into two categories: Common information and additional information. If the article does not primarily focus on the event, the prompt classifies its content into three categories: Common information, Additional information, and Irrelevant information.

> You are a financial expert tasked with analyzing corporate events. You will be provided with two texts: **Text 1**, the firm's official disclosure, and **Text 2**, a news article that may report on the same event(s). **Text 2** has been pre-tokenized into sentences, with each sentence assigned a corresponding number. Your job is to extract, interpret, and categorize the information from these texts step by step.
>
> 1. Identify the Overlapping Event and Firm:
>   - Examine **Text 1** and **Text 2** to identify the **overlapping event** and the **firm** mentioned in both texts.
>   - Provide a concise summary of the overlapping event and the firm involved.
>   - If no overlapping event or firm is found, set the event description to **"N/A."**
> 2. Determine the Focus of Text 2:
>   - Text 2 Length: Count the total number of sentences in **Text 2**.
>   - Text 2 Focus Length: Count the number of sentences in **Text 2** specifically related to the overlapping event/firm or factors affecting the firm's performance.
>   - Text 2 Percentage: Calculate the percentage of sentences in **Text 2** that focus on the overlapping event/firm.
>   - Text 2 Focus:
>   – Assign **1** if **Text 2 Percentage** is **30% or higher**.
>   – Assign **0** if **Text 2 Percentage** is **less than 30%**.
> 3. Analyze Relevant Information in Text 2:
>   - If **Text 2 Focus = 0**: Divide **Text 2** into three sections and output the sentence numbers in each section:
>   (a) Common Information: Sentences that match or restate the disclosure in **Text 1**.
>   (b) Additional Information: Sentences providing new insights or details related to the overlapping event/firm or factors affecting the firm's performance.
>   (c) Irrelevant Information: Sentences unrelated to the overlapping event or firm, or not fitting into the Common Information or Additional Information categories.

- If **Text 2 Focus = 1**: Divide **Text 2** into two sections and output the sentence numbers in each section:
  (a) Common Information: Sentences that match or restate the disclosure in **Text 1**.
  (b) Additional Information: All other sentences from **Text 2** not categorized as Common Information.
  (c) Mark **Irrelevant Information** as **"N/A."**
- Validation: Ensure that the total number of sentences in **Text 2** equals the sum of sentences across all categorized sections. No sentences should be omitted or duplicated.

4. Double-Check Accuracy: Cross-check the categorized sections with the original content of **Text 2** to ensure no sentences are missing or misclassified.

5. Assess Sentiment: Provide the overall sentiment of **Text 2** using the following sentiment scale:

   - **-1**: Very negative sentiment
   - **-0.5**: Mildly negative sentiment
   - **0**: Neutral sentiment
   - **0.5**: Mildly positive sentiment
   - **1**: Very positive sentiment

Expected JSON Output:

```
{
    "event_description": "A brief summary of the overlapping event and the
    firm involved, or 'N/A' if no overlap exists.",
    "text2_length": "Total number of sentences in Text 2.",
    "text2_percentage": "Percentage of Text 2 focusing on the overlapping
    event/firm.",
    "text2_focus": "1 or 0",
    "common_information": {
        "section": ["1", "2", ...],
        "length": "Number of sentences in the Common Information section."
    },
    "additional_information": {
        "section": ["3", "4", ...],
        "length": "Number of sentences in the Additional Information section."
    },
    "irrelevant_information": {
        "section": ["5", "6", ...] or "N/A",
        "length": "Number of sentences in the Irrelevant Information section,
        or 'N/A'."
    },
    "text2_sentiment": "-1 or -0.5 or 0 or 0.5 or 1"
}
```

Table B.3. Prompts for identifying sentiment in 8-K filings and WSJ article

This table provides the detailed prompts given to ChatGPT to identify sentiment in 8-K filings and in WSJ articles. Panel A outlines the instructions for classifying sentiment in 8-K filings, while Panel B presents the instructions for identifying sentiment in WSJ articles and sentiment within the common, additional, and irrelevant segments of WSJ articles.

Panel A. Prompt for identifying 8-K filing sentiment

You are a financial expert tasked with analyzing corporate disclosures. Your goal is to assess the sentiment expressed in the following text and classify it into one of five sentiment levels:

1. **-1**: Very negative sentiment
2. **-0.5**: Mildly negative sentiment
3. **0**: Neutral sentiment
4. **0.5**: Mildly positive sentiment
5. **1**: Very positive sentiment

Text to Analyze:"{text}"

Task:
Based on the sentiment levels described above, provide the sentiment category as a single number (e.g., -1, -0.5, 0, 0.5, or 1). Your response should contain only the sentiment value.

Panel B. Prompt for identifying WSJ article sentiment

You are a financial expert tasked with analyzing corporate events. You will be provided with find a segment of a news article related to a corporate event. Your task is to assess the sentiment expressed in the provided text and classify it into one of five sentiment levels:

1. **-1**: Very negative sentiment
2. **-0.5**: Mildly negative sentiment
3. **0**: Neutral sentiment
4. **0.5**: Mildly positive sentiment
5. **1**: Very positive sentiment

Text to Analyze:"{text}"

Task:
Based on the sentiment levels described above, provide the sentiment category as a single number (e.g., -1, -0.5, 0, 0.5, or 1). Your response should contain only the sentiment value.

Table B.4. Prompts for topic analysis of 8-K items and WSJ additional information

This table presents the detailed prompts provided to ChatGPT to analyze topics from 8-K filings and WSJ segments related to corporate events, which are then used to construct the Sankey diagram. Panel A provides the instructions for identifying five topics for each 8-K item using a random subsample of 5,000 items. Panel B provides the instructions for identifying ten key topics in the WSJ additional information segments, using a random subsample of 2,000 segments. Panel C provides the instructions for consolidating the identified topics into broader corporate event categories, which are applied to both 8-K topics and WSJ additional information topics. Panel D provides the instructions for classifying each 8-K item into one of eight event categories, while Panel E provides the instructions for classifying the WSJ additional information into one of ten event categories.

Panel A. Prompt to identify 8-K item topic

You are a financial expert specializing in analyzing corporate events. You will be provided with a text segment from a corporate disclosure filed by a company regarding a corporate event. Your task is to thoroughly analyze the text and identify key topics discussed in it.

Requirements:

1. Identify approximately **5 key topics** discussed in the text. If fewer than 5 topics are present, list only the available topics.
2. The topics should be expressed as **concise, generalized phrases** that are broad enough to apply to other disclosures or firms while accurately capturing the content.
3. **Avoid redundancy** by not including repeated or overly specific details. Keep the phrases succinct and clear.

Expected JSON Output:

```
{
    "text_topics": ["topic 1", "topic 2", ...] or "N/A"
}
```

Inputs:
- **Text**: {text}

Panel B. Prompt to identify WSJ topic

You are a financial expert specializing in analyzing corporate events. You will be provided with two text segments: **Text 1**, the segment discussing the corporate event(s), and **Text 2**, the segment providing additional discussions on the event(s). Both extracted from the same news article. Your task is to thoroughly analyze each text and identify approximately 10 key topics discussed in each.

Requirements:

1. The topics should be expressed as concise, generalized phrases that are broad enough to apply to other news articles or other firms but still accurately capture the essence of the content.
2. If fewer than 10 topics are present, include only the available topics.
3. Avoid overly specific details or repeated topics; keep the phrases succinct and clear.

Expected JSON Output:

```
{
    "text1_topics": ["topic 1", "topic 2", ...] or "N/A",
    "text2_topics": ["topic 1", "topic 2", ...] or "N/A"
}
```

Inputs:
- **Text 1**: {text1}
- **Text 2**: {text2}

Panel C. Prompt to summarize topics

The following list contains topics discussing corporate events. As a financial expert, your task is to summarize these topics into 10 unique categories based on their frequency:

', '.join(text)

Summarized Categories:

Panel D. Prompt to categorize 8-K item information

You are a financial expert tasked with analyzing corporate events. Your goal is to classify the provided text segment into **one** predefined corporate event category from the list below. The text should be assigned to only **one category**. If the text clearly belongs to multiple categories, select the one that is most relevant or has the most content. If the text does not fit into any of the listed categories, respond with "N/A".

1. **Mergers and Acquisitions**: This category includes discussions related to merger agreement details, acquisition financing, regulatory approvals, shareholder considerations, and merger-related compliance issues.

2. **Debt Financing and Restructuring**: Topics under this category encompass various aspects of debt issuance, including loan agreements, credit facilities, amendments to credit terms, debt restructuring processes, and securities regulations compliance.

3. **Corporate Governance**: This category involves topics related to corporate governance structures, board composition, shareholder voting processes, and related party transactions, especially following significant corporate events.

4. **Equity Transactions and Stock Issues**: Discussions in this category cover stock purchase agreements, common and preferred stock issuance, stock buybacks, dividend distributions, and equity incentive plans, including any impacts on shareholder rights and corporate structures.

5. **Financial Reporting and Disclosure**: Focused on the obligations for financial transparency, this category encapsulates topics related to quarterly and annual reports, SEC filings, disclosures of financial performance, and compliance with accounting standards.

6. **Investor Relations and Communication**: This category encompasses thorough discussions on investor communication strategies, shareholder meeting outcomes, proxy solicitations, and corporate publicity efforts surrounding financial results.

7. **Regulatory Approvals and Legal Proceedings**: Topics within this category touch on regulatory approvals necessary for mergers, acquires, litigations involving the company, compliance with regulations or securities laws, and legal issues arising from corporate restructuring or commercial disputes.

8. **Executive Management and Compensation**: This category encompasses the presence of topics related to executive appointments, changes in management, compensation packages for leaders, severance arrangements, and overall leadership transitions.

Expected JSON Output:

```
{
    "8kitem_categories": "1" or "2" or "3" or "4" or "5" or "6" or "7" or "8"
    or "N/A"
}
```

{text}

Panel E. Prompt to categorize WSJ additional information

You are a financial expert tasked with analyzing corporate events. Your goal is to classify the given text segment into one or more predefined corporate event categories. Below are the definitions for the 10 major corporate event categories:

1. **Corporate Governance and Leadership Changes**: Includes discussions on board composition, CEO transitions, executive compensation, shareholder activism, corporate governance practices, and the impact of leadership changes on company performance.
2. **Mergers and Acquisitions (M&A)**: Focuses on merger negotiations, acquisition strategies, divestitures, market reactions to M&A, regulatory scrutiny, competition analysis, and historical context of major acquisitions.
3. **Financial Performance and Market Reactions**: Covers earnings reports, revenue forecasts, profit margins, stock price movements, investor sentiment, performance metrics, and analysis of key financial indicators.
4. **Investment Strategies and Market Trends**: Addresses investment trends, asset management, capital allocation, market volatility, hedge fund activities, ETF dynamics, and strategies for navigating challenging economic environments.
5. **Consumer Behavior and Market Dynamics**: Examines shifts in consumer preferences, retail performance, purchasing trends, e-commerce growth, and the impact of economic conditions on consumer spending.
6. **Regulatory and Legal Challenges**: Discusses the implications of regulatory changes, compliance issues, lawsuits, and legal disputes affecting corporate actions and market positioning.
7. **Sector-Specific Challenges and Opportunities**: Focuses on particular industries (e.g., automotive, pharmaceuticals, energy, technology), exploring challenges, growth potential, market share dynamics, and competitive positioning within these sectors.
8. **Corporate Social Responsibility and Sustainability**: Involves discussions on environmental impacts, community engagement, social equity, sustainability initiatives, and corporate responsibility strategies.
9. **Technological Innovations and Market Adaptations**: Highlights the role of technology in business strategies, shifts towards digital solutions, innovation in product offerings, and responses to technological disruptions.
10. **Economic Conditions and Global Market Influences**: Covers the impact of macroeconomic factors such as inflation, interest rates, global trade tensions, and geopolitical events on financial markets and corporate performance.

Your task is to analyze the text provided and assign it to one or more categories from the list above. If the text does not fit into any category, respond with "N/A."
Expected JSON Output:

```
{
    "additional_categories": ["1", "2", "..."] or "N/A"
}
```

Inputs: {text}

Table B.5. Example of matched 8-K filing and WSJ article

This table provides an example of a corporate event covered in both an 8-K filing and a WSJ article. Panel A presents the main body of the 8-K filing filed by Moderna. Panel B shows the original WSJ article titled *Moderna Gets U.S. Funding for Development, Manufacturing of Experimental Coronavirus Vaccine*. Panel C shows the text segments that ChatGPT extracted from the WSJ article, distinguishing between *Common Information* and *Additional Information*, where the former contains information common to both sources and the latter additional details beyond the filing.

Panel A. Moderna's Form 8-K Filing

Item 1.01. Entry into a Material Definitive Agreement.
On April 16, 2020, Moderna, Inc. (the "Company") entered into a contract (the "BARDA Contract") with the Biomedical Advanced Research and Development Authority ("BARDA"), a division of the U.S. Department of Health and Human Services' Office of the Assistant Secretary for Preparedness and Response. Under the BARDA Contract, the Company will receive an award of up to $483 million to accelerate the development of the Company's mRNA vaccine candidate (currently mRNA-1273) against the novel coronavirus (SARS-CoV-2). BARDA will fund the advancement of the mRNA vaccine candidate to potential licensure, including clinical studies of the candidate and the manufacture of the candidate for use in these studies.

The BARDA Contract could result in payments to the Company of up to approximately $483 million, and consists of an approximately two-year base period-of-performance and a total contract period-of-performance (base period plus option exercises) of up to approximately five years and six months (if necessary). Under the base period-of-performance, the Company will conduct activities intended to obtain licensure of mRNA-1273 through a biologics license application submission for mRNA-1273.

The BARDA Contract contains terms and conditions that are customary for government contracts of this nature, including provisions giving the government the right to terminate the contract at any time for its convenience.

The foregoing is a brief description of the material terms of the BARDA Contract and does not purport to be a complete description of the rights and obligations of the parties thereunder. The foregoing description is qualified in its entirety by reference to the BARDA Contract, which will be filed as an exhibit to the Company's Quarterly Report on Form 10-Q for the period ending June 30, 2020.

Item 7.01. Regulation FD Disclosure.
On April 16, 2020, the Company issued a press release announcing its entry into the BARDA Contract, a copy of which is furnished as Exhibit 99.1 to this Current Report on Form 8-K.

The information in this Item 7.01 to this Current Report on Form 8-K, and in Exhibit 99.1 furnished herewith, shall not be deemed to be "filed" for purposes of Section 18 of the Securities Exchange Act of 1934, as amended (the "Exchange Act"), or otherwise subject to the liabilities of that section, nor shall such information be deemed incorporated by reference in any filing under the Securities Act of 1933, as amended, or the Exchange Act, except as expressly set forth by specific reference in such a filing.

Item 9.01. Financial Statements and Exhibits.
(d) Exhibits

Panel B. WSJ News Article

Moderna Gets U.S. Funding for Development, Manufacturing of Experimental Coronavirus Vaccine

Funding of up to $483 million will help the company test the vaccine further and scale up manufacturing capabilities

By Peter Loftus
Updated April 16, 2020 7:23 pm ET

Moderna Inc. said Thursday it reached an agreement to receive as much as $483 million in funding from a federal agency to accelerate the development and production of its closely watched experimental vaccine against the new coronavirus.

The federal funding will cover advancing the vaccine through a series of studies to potential approval by the U.S. Food and Drug Administration.

The Cambridge, Mass., company also said it plans to hire up to 150 new employees to scale up manufacturing of doses for the studies, and to develop processes for large-scale manufacturing if the studies succeed.

"This grant is enabling us to aggressively fund the best and largest clinical studies that we can do," Moderna Chief Executive Stephane Bancel said in an interview. "We can fund the manufacturing process so we can make as much product as we can."

The announcement shows the vaccine's development has moved far along enough that preparations are under way to test it further and to expand manufacturing, but the commitments don't guarantee the vaccine will prove to work safely.

Moderna's vaccine is among the most advanced coronavirus vaccine candidates. The first study of the vaccine in people started in March, though U.S. government researchers say the soonest it could finish testing and be ready for widespread use is early next year.

The first study is being expanded to include an older group of patients than those originally enrolled in the trial, Moderna said. The company plans to start a second-phase study of the vaccine in the second quarter, and could begin a late-stage trial in the fall.

The U.S. funding will come from the Biomedical Advanced Research and Development Authority, a part of the U.S. Health and Human Services department, Moderna said.

Moderna developed the vaccine candidate working with the National Institute of Allergy and Infectious Diseases. Results from the first trial could come this summer, according to NIAID, though further testing is needed to ensure it works safely and effectively.

Among other companies with advanced vaccine candidates is China's CanSino Biologics Inc., which last week said its vaccine candidate had proved safe in initial human testing and would soon start the second phase of testing.

Johnson & Johnson has said it plans to start human testing on a vaccine by September, and it could be available on an emergency basis early next year.

Moderna's vaccine uses genetic code, known as messenger RNA, to prepare a person's immune system to fight off the new coronavirus. Messenger RNA, or mRNA, carries the instructions encoded in DNA for making proteins.

The company's coronavirus vaccine mRNA is based on a part of the coronavirus. The vaccine aims to instruct the body's cells to manufacture a protein resembling the coronavirus part. If exposed to the virus, the vaccinated person's immune system should then be trained to fight the infection.

Moderna has manufactured initial batches of the vaccine for human testing at a plant in Norwood, Mass., outside Boston, and delivered the first batch in February.

Moderna said it would expand production to three shifts a day, seven days a week from two shifts a day five days a week.

The expansion of its manufacturing capacity, Moderna said, could enable the company to supply millions of doses a month in 2020, and with further investment, tens of millions a month in 2021, if the vaccine is successful in studies.

Panel C. Text segments extracted from WSJ article

Common Information

- Moderna Inc. said Thursday it reached an agreement to receive as much as $483 million in funding from a federal agency to accelerate the development and production of its closely watched experimental vaccine against the new coronavirus.
- The federal funding will cover advancing the vaccine through a series of studies to potential approval by the U.S. Food and Drug Administration.
- The U.S. funding will come from the Biomedical Advanced Research and Development Authority, a part of the U.S. Health and Human Services department, Moderna said.

Additional Information

- The Cambridge, Mass., company also said it plans to hire up to 150 new employees to scale up manufacturing of doses for the studies, and to develop processes for large-scale manufacturing if the studies succeed.
- "This grant is enabling us to aggressively fund the best and largest clinical studies that we can do," Moderna Chief Executive Stephane Bancel said in an interview. "We can fund the manufacturing process so we can make as much product as we can."
- The announcement shows the vaccine's development has moved far along enough that preparations are under way to test it further and to expand manufacturing, but the commitments don't guarantee the vaccine will prove to work safely.
- Moderna's vaccine is among the most advanced coronavirus vaccine candidates. The first study of the vaccine in people started in March, though U.S. government researchers say the soonest it could finish testing and be ready for widespread use is early next year.
- The first study is being expanded to include an older group of patients than those originally enrolled in the trial, Moderna said. The company plans to start a second-phase study of the vaccine in the second quarter, and could begin a late-stage trial in the fall.
- Moderna developed the vaccine candidate working with the National Institute of Allergy and Infectious Diseases. Results from the first trial could come this summer, according to NIAID, though further testing is needed to ensure it works safely and effectively.
- Among other companies with advanced vaccine candidates is China's CanSino Biologics Inc., which last week said its vaccine candidate had proved safe in initial human testing and would soon start the second phase of testing.
- Johnson & Johnson has said it plans to start human testing on a vaccine by September, and it could be available on an emergency basis early next year.
- Moderna's vaccine uses genetic code, known as messenger RNA, to prepare a person's immune system to fight off the new coronavirus. Messenger RNA, or mRNA, carries the instructions encoded in DNA for making proteins.

- The company's coronavirus vaccine mRNA is based on a part of the coronavirus. The vaccine aims to instruct the body's cells to manufacture a protein resembling the coronavirus part. If exposed to the virus, the vaccinated person's immune system should then be trained to fight the infection.
- Moderna has manufactured initial batches of the vaccine for human testing at a plant in Norwood, Mass., outside Boston, and delivered the first batch in February.
- Moderna said it would expand production to three shifts a day, seven days a week from two shifts a day five days a week.
- The expansion of its manufacturing capacity, Moderna said, could enable the company to supply millions of doses a month in 2020, and with further investment, tens of millions a month in 2021, if the vaccine is successful in studies

Appendix C: 8-K items, definitions and propotions.

Table C.1. Section definitions in 8-K

This table presents the items (or schedules) and their corresponding definitions for each item in 8-K reports, along with the proportion of observations for each item relative to the total number of observations in our sample. When calculating this proportion, we exclude Section 9.

Item	Definition	Proportion
Section 1	Registrant's Business and Operations	14.6%
Item 1.01	Entry into a Material Definitive Agreement	13.1%
Item 1.02	Termination of a Material Definitive Agreement	1.3%
Item 1.03	Bankruptcy or Receivership	0.2%
Item 1.04	Mine Safety - Reporting of Shutdowns and Patterns of Violations	0.0%
Section 2	Financial Information	29.1%
Item 2.01	Completion of Acquisition or Disposition of Assets	1.9%
Item 2.02	Results of Operations and Financial Condition	21.3%
Item 2.03	Creation of a Direct Financial Obligation or an Obligation under an Off-Balance Sheet Arrangement of a Registrant	5.5%
Item 2.04	Triggering Events That Accelerate or Increase a Direct Financial Obligation or an Obligation under an Off-Balance Sheet Arrangement	0.2%
Item 2.05	Costs Associated with Exit or Disposal Activities	0.1%
Item 2.06	Material Impairments	0.1%
Section 3	Securities and Trading Markets	5.8%
Item 3.01	Notice of Delisting or Failure to Satisfy a Continued Listing Rule or Standard; Transfer of Listing	1.1%
Item 3.02	Unregistered Sales of Equity Securities	3.6%
Item 3.03	Material Modification to Rights of Security Holders	1.1%
Section 4	Matters Related to Accountants and Financial Statements	1.2%
Item 4.01	Changes in Registrant's Certifying Accountant	0.9%
Item 4.02	Non-Reliance on Previously Issued Financial Statements or a Related Audit Report or Completed Interim Review	0.3%
Section 5	Corporate Governance and Management	11.3%
Item 5.01	Changes in Control of Registrant	0.7%
Item 5.02	Departure of Directors or Certain Officers; Election of Directors; Appointment of Certain Officers; Compensatory Arrangements of Certain Officers	1.5%
Item 5.03	Amendments to Articles of Incorporation or Bylaws; Change in Fiscal Year	3.0%
Item 5.04	Temporary Suspension of Trading Under Registrant's Employee Benefit Plans	0.0%
Item 5.05	Amendment to Registrant's Code of Ethics, or Waiver of a Provision of the Code of Ethics	0.0%
Item 5.06	Change in Shell Company Status	0.1%
Item 5.07	Submission of Matters to a Vote of Security Holders	5.8%
Item 5.08	Shareholder Director Nominations	0.0%

Item	**Definition**	**Proportion**
Section 6	Asset-Backed Securities	0.1%
Item 6.01	ABS Informational and Computational Material	0.0%
Item 6.02	Change of Servicer or Trustee	0.1%
Item 6.03	Change in Credit Enhancement or Other External Support	0.0%
Item 6.04	Failure to Make a Required Distribution	0.0%
Item 6.05	Securities Act Updating Disclosure	0.0%
Section 7	Regulation FD	16.9%
Item 7.01	Regulation FD Disclosure	16.9%
Section 8	Other Events	21.1%
Item 8.01	Other Events (The registrant can use this Item to report events that are not specifically called for by Form 8-K, that the registrant considers to be of importance to security holders.)	21.1%
Section 9	Financial Statements and Exhibits	NA
Item 9.01	Financial Statements and Exhibits	NA

Table C.2. Determinants of WSJ coverage on corporate event

This table presents the results of a probit regression that examines the determinants of WSJ coverage for corporate events disclosed in Form 8-K filings. The dependent variable *Treat=1* equals one if the event reported in an 8-K was covered by the WSJ, and zero otherwise. *S&P 1500=1* is an indicator equal to one if the firm belongs to the S&P 1500 index. *Items* denotes the specific 8-K filing items. Definitions of the remaining independent variables are provided in Appendix A. Firm characteristics are measured as of the fiscal year prior to the filing date. Standard errors are heteroskedasticity-robust and clustered at the firm level, with z-statistics reported in parentheses. Coefficients marked with ***, **, and * are significant at the 1%, 5%, and 10% level, respecitively.

VARIABLES	(1) Treat=1
log(ME)	0.207***
	(25.09)
Book-to-Market	-0.016
	(-1.59)
Z Score	0.038***
	(7.80)
Market Leverage	0.348***
	(11.08)
IO	-0.268***
	(-11.67)
S&P 1500 = 1	-0.080***
	(-5.50)
ROA	-0.343***
	(-6.46)
Investment-to-assets	0.128***
	(5.11)
Tangibility	0.012
	(0.50)
Momentum	0.037***
	(3.43)
Fluidity	-0.019***
	(-10.43)
log(Illiq)	-0.139***
	(-19.57)
Items	-0.000***
	(-11.66)
Observations	288,767