News, Sentiment and Trading *

Luca Barbaglia†
Mingyuan Kong‡
Sebastiano Manzan§

Preliminary and Incomplete - Please Do Not Quote

Abstract

We conduct a sentence-level sentiment analysis on idiosyncratic firm news texts and study news implied article-level sentiment variations. News implied sentiment variation has impacts on stock market trading where it increases turnover and volatility. News exposure positively affects trading activities but negatively affects returns. We also find that the impacts from sentiment, sentiment variations and news exposure are stronger for small firms, while influences are weaker in large firms. Our results indicate that the impact of sentiment on trading is not linear that neutral sentiment has stronger impact compared to tail sentiments. Neutral sentiment creates more space for divergent expectations and different interpretations of information. The findings highlight the importance of considering text structure, firm size, and neutral news in understanding the impacts of news.

JEL classification: G12, G14, G40, G41.

Keywords: sentiment, textual analysis, sentiment variation, neutral sentiment.

*We would like to thank Lin Peng, Jun Wang, Dexin Zhou, Yijun Zhou and seminar participants at Baruch College Brownbag seminar.

†European Commission - Joint Research Centre (JRC). Email: luca.barbaglia@ec.europa.eu
‡Baruch College, CUNY. Email: Mingyuan.Kong@baruch.cuny.edu
§Baruch College, CUNY. Email: Sebastiano.Manzan@baruch.cuny.edu

1. Introduction

In a rational setting of traditional finance theory, sentiment does not play a role. However, existing literature documents evidence that sentiment is affecting securities market. Realizing the importance of sentiment, researchers first seek proxies and constructs measures to evaluate sentiment. Since sentiment is regarded as a belief, early research is working on capturing sentiment from market data. Later on, the access to a huge amount of news data gives researchers a way to analyze the sentiment from information at hands. At early stage, researchers apply dictionary-based method to measure sentiment of text. With the advances in statistical tools and machine learning methods, it is possible to work with big data to measure sentiment in an efficient and accurate way.

Our paper provides new insights into textual analysis and news sentiment. First, we contribute to the area of news impact on stock market. Existing literature shows that positive sentiment motivates trading, while we investigate further in this and elaborate the relationship between news sentiment and stock trading. Our hypothesis is that neutral news without clear positive or negative signals allows investors to interpret information in heterogeneous ways and this motivates more trading. Second, we pay attention to text structures in our textual analysis. Unlike most works that take a whole article as one object, we start with sentences of each article and consider the sentiment flows within one article. We hypothesize that sentiment variations, which are implied by text structures, affect trading. High sentiment variation indicates more uncertainty of information for investors. Third, we emphasize the importance of firm size. Small firms are usually not well-exposed to media, and because of this, investors' attentions are captured by large firms. Once small firms are mentioned in news, investors tend to have stronger reactions. We investigate the magnitude of impact from text sentiment.

There has been literature studying the impact of sentiment. Tetlock (2007) measure the interactions between The Wall Street Journal news and the stock market. It concludes that high media pessimism predicts downward pressure on market prices, which is followed by a reversion to fundamentals. Also, unusually high or low pessimism predicts high market trading volume. Ke, Kelly, and Xiu (2019) construct a text-mining methodology to extract sentiment information from news articles to predict asset returns. Their model has few overlapped words with the popular dictionary in Loughran and McDonald (2011). Schwenkler and Zheng (2019) document the impact of news on distressed firm links and the news-implied relationship, while Tao, Yim, and Han (2020) also study the news-based links of news cocoverage which investors fail to recognize.

Positive and negative news are regarded as the main sources that drive stock market movements. However, most news articles are not delivering clear positive or negative signals, they usually transit neutral sentiment to investors. This raises the question: does neutral sentiment affect the market? If yes, how does it affect the market? Neutral news contains information as positive or negative news, and due to the fact that it lacks clear directional sentiment signal, investors may interpret it from various perspectives resulting in different trading decisions. Most existing literature takes an article as a whole and label it with a sentiment score. But even articles with the same sentiment level can vary a lot in their structures. For example, it is possible that one article with six neutral sentences generates a neutral sentiment, while the other article with three positive sentences and three negative sentences generates a neutral sentiment as well. Investors will definitely have different feelings when reading these articles. The structure matters. Also, existing literature often pays attention to large firms which are better-exposed to media. But small firms can be greatly affected by news since investors can even forget the existence of the small firm until it is mentioned by news. Due to our large dataset, we are able to conduct our study on both large and small firms. This paper seeks to shed light on the area of neutral news, news structure, and news sentiment impact on small firms.

We see multiple methodologies in this area of study. There are dictionaries that are widely used in existing studies (see, e.g., Stone, Dunphy, and Smith, 1966; Hart, 2000; Loughran and McDonald, 2011). Because of the development in machine learning and other statistical tools, new methodologies are used now. Ke et al. (2019) apply a supervised learning framework to extract sentiment information from news. Saurabh and Dey (2020) apply Artificial Neural Network, Granger-causality, and Vector Auto Regression. Zhu, Wu, and Wells (2023) propose News Embedding UMAP Sparse Selection (NEUSS) model and News Sparse Encoder with Rationale (INSER) model to predict stock returns.

Motivated by these studies, we apply VADER (Valence Aware Dictionary for Sentiment Reasoning) in our work. It is a natural language model that is trained by Twitter data. The model maps lexical characteristics with human emotions, generating a sentiment score for each input. Unlike most existing literature that analyzing articles as a whole. We split texts into sentences and conduct sentiment analysis at sentence level. By doing this, we are able to study the structure of texts. Because Twitter has a word limit which makes every post short and we conduct our analysis at sentence level, the our sample input size and VADER training object size match so that VADER is able to detect the sentiment at sentence level. Then we consider the sentence sentiment flows in each article and evaluate its variation using standard deviation, we get the news implied sentiment variation and further study its impact.

Different datasets have been studied comprehensively. Manela and Moreira (2017) study

Wall Street Journal front-page articles. Ke et al. (2019) select Dow Jones Newswire articles. Chen, Després, Guo, and Renault (2019) use StockTwits and Reddit rather than traditional news sources. Li (2019) focus Asian-pacific countries and use China Investor's Sentiment Index Research Database. Tetlock (2007) take The Wall Street Journal news as the study object.

We obtain the news data from Dow Jones Data, News and Analytics (DNA) database, ranging from January 1, 2005 to June 30, 2021. Our data covers publishers including The Wall Street Journal (J), The New York Times (NYTF), Reuters News (LBA), The Washington Post (WP), The Dallas Morning News (DAL), Chicago Sun-Times (CHI), and The San Francisco Chronicle (SFC). Compared to most existing studies, our sample is larger and more comprehensive that we have news sources from the west coast to the east coast. Main publishers are all covered and we have both large firms and small firms in our sample. With this large dataset, we are able to achieve our goal of study.

Our paper provides the following findings. First, we document that the impact on trading is not linear that neutral news allows investors to interpret signals in heterogeneous ways and this motivates more trading in the market. The impact from neutral news is stronger than impacts from positive or negative sentiments which are clear to investors. We emphasize the importance of neutral news. Second, our result suggests that text structure matters that news implied sentiment variation motivates trading. When the sentiment variation is high, it implies a high level of uncertainty and ambiguity that leads to divergent expectations and interpretations, resulting in a high trading volume and high volatility. Third, we find that firm size determines the magnitude of impact from text sentiment. Large firms are less likely to be affected by news. Because of data availability and sufficient media coverage on large firms, most existing literature focuses on large firms. Our results show that large firm size weakens the impact and smaller firms are greatly affected by news. Fourth, we find that news exposure positively affects trading activities but negatively affects returns, and impact is more prominent in small firms. Small firms are worth more attention. Lastly, we provide more evidence that high idiosyncratic news sentiment motivates trading. This supports the findings in existing literature that when sentiment is positive, there are more trading activities. News sentiment has positive and significant impact on turnover.

The remainder of the paper is organized as follows. Section 2 provides reviews of related literature. Section 3 describes the data and methodologies. Section 4 investigates how news sentiments affect trading. Section 5 concludes.

2. Related Literature

Classical finance theory ignores the impact of sentiment. Because of the existence of arbitragers in the market, irrational investments are traded against arbitrage so that the mispriced securities are driven to the fundamental values. Kyle (1985) and Black (1986) call these irrational investors noise traders. De Long, Shleifer, Summers, and Waldmann (1990) examine the behavior of noise trader and propose the assumption that investors are subject to sentiment. They define sentiment as '...a belief about future cash flows and investment risks that is not justified by the facts at hand'. Strategies that trade against noise traders are actually requiring the mean-reverting pattern of sentiment.

The first challenge is to define and measure sentiment. Baker and Wurgler (2006) propose one possible definition of investor sentiment: the propensity to speculate. This definition motivates them to study sentiment in a cross-sectional setting, considering that even if arbitrage forces are the same across stocks, there are cross-sectional effects. Baker and Wurgler (2006) investigate six proxies for sentiment and form a composite index SENTIMENT. They find that when beginning-of-period proxies for sentiment are low, subsequent returns are relatively high for small stocks, young stocks, high volatility stocks, un-profitable stocks, non-dividend-paying stocks, extreme growth stocks, and distressed stocks. Baker and Wurgler (2007) study more proxies for sentiment and they conduct a 'top down' approach which focuses on the measurement of reduced-form, aggregate sentiment and traces its effects to market returns and individual stocks. They show that it is possible to measure sentiment and stocks that are difficult to arbitrage or to value are most affected by sentiment. Baker, Wurgler, and Yuan (2012) construct a quantitative sentiment indices in six stock markets and find that global sentiment is a statistically and economically significant contrarian predictor of market returns. Huang, Jiang, Tu, and Zhou (2015) hold a different view with Baker and Wurgler (2006) and Baker and Wurgler (2007), and they extract the most relevant common component from the proxies. After applying a two-step partial least squares regression, it is better extracting the common components with good filter to the information which is irrelevant to the variable of interest.

Yu and Yuan (2011) analyze whether investor sentiment influences the mean-variance relation and explores whether sentiment attenuates the link between the conditional mean and variance of returns. Asset pricing under rational settings suggest the positive relation between risk and return (Merton (1980)). This paper finds that the stock market's expected excess return is positively related to the market's conditional variance in low-sentiment periods but unrelated to variance in high-sentiment periods. They conclude a stronger negative correlation between returns and contemporaneous volatility innovations in the low-sentiment periods. Stambaugh, Yu, and Yuan (2012) study the role of investor sentiment in the asset pricing framework with concentration on anomalies. They find that long-short strategies for a broad set of anomalies in cross-sectional returns exhibit empirical properties consistent with a combination of short-sale impediments and market-wide sentiment. Stambaugh, Yu, and Yuan (2014) conduct simulation analysis and the result supports Stambaugh et al. (2012).

The literature above works on market sentiment, and the sentiment is really aligned to the definition in De Long et al. (1990). Researchers extract sentiment proxies and index from traders' behaviors in the market, and even more indirectly from the outcomes of their actions such as returns and turnovers. Another method relies on information at hand, and it is called text-based or textual sentiment. The fundamental difference between investor sentiment and textual sentiment is that the former focuses on the subjective judgments and behavioral characteristics of investors, while the latter includes the more objective reflection of conditions within firms, institutions and markets. Because of the increasing collection of financial closures and news data, researchers start to work on extracting sentiment from financial text data.

Traditional analysis is a dictionary-based method, in which the original text are matching a dictionary with positive and negative words and then calculate the overall 'tone' of the text. There are some popular dictionaries used in finance research. General Inquirer (GI) is a built-in dictionary developed and used in Stone et al. (1966). Most of the lists are from Harvard IV-4 dictionaries. DICTION is another popular dictionary developed by Hart (2000). Das and Chen (2007) first apply textual analysis in finance, in which they develop a methodology to extract investor sentiment from messages.Loughran and McDonald (2011) study the dictionaries in the finance framework and find that the overall negative words are not exactly negative in financial texts. They develop a negative list and five other word lists which are suitable for financial text. But the dictionary-based method is not accurate, and the word selection is quite subjective. Later on, with the development of machine learning methods, researchers start to analyze text sentiment through statistical techniques and machine learning tools, in which there is a training set labeled by tones and then applying the sophisticated parameters in the test set. Ke et al. (2019) construct a supervised learning framework, in which a text-mining method is applied to extract sentiment information from Dow Jones Newswire news articles. After comparing with performance of simple trading strategies that relies on traditional sentiment scores method used in the industry, strategies based on their scoring method ourperfom. Saurabh and Dey (2020) apply Artificial Neural Network, Granger-causality, and Vector Auto Regression to detect the relation between social moods-dimension and stock market. They don't calculate sentiment scores for text data, but words of particular emotions with synonymous words instead. Social moods-dimension significantly impacts the market return at the aggregate level.

There are multiple sources of news such as newswires and some forums. Manela and Moreira (2017) study Wall Street Journal front-page articles and construct a news implied volatility (NVIX). They are the first to extract information about aggregate uncertainty from news coverage using machine learning techniques. Baker, Bloom, Davis, and Sammon (2021) study next-day newspaper articles that explain stock market jumps from the views of proximate cause, clarity of explanation, and geographic impact. They find that policy news such as monetary policy and government spending is the main trigger of upward jumps and this type of jumps is inversely related to stock market performance and lowers market volatility. Tetlock (2007) measure the interactions between The Wall Street Journal news and the stock market. They conclude that high media pessimism predicts downward pressure on market prices, which is followed by a reversion to fundamentals. Shapiro, Sudhof, and Wilson (2020) demonstrates text sentiment analysis tools on economic sentiment derived from economic and financial newspaper articles. They test dictionary based tools and natural languages process tools. Also, they combine different tools to analyze the comprehensive techniques. Their results illustrate the gains from combining existing lexicons and from accounting for negation.Chen et al. (2019) use two special datasets: StockTwits and Reddit. They study the sentiment of messages posted on StockTwits and Reddit, and they focus on the Cryptocurrency IndeX (CRIX). Their result suggests that empirical analysis on text sentiment should be paid attention to, such as domain-specific lexicons. Shapiro et al. (2020) demonstrates text sentiment analysis tools on economic sentiment derived from economic and financial newspaper articles. They test dictionary based tools and natural languages process tools. Their results illustrate the gains from combining existing lexicons and from accounting for negation. Barbaglia, Consoli, and Manzan (2023) extract sentiment from news articles to study economy state using Dow Jones Data, News and Analytics (DNA) database.

More and more datasets have been used in textual analysis research and researchers are providing new findings and evidence. Li (2019) conduct their study in a different angle that they investigate the predictability of Chinese investor sentiment (CIS) for returns and volatilities of 12 Asia-pacific stock markets. The sentiment data comes from the China Investor's Sentiment Index Research Database in the China Stock Market & Accounting Research (CSMAR) Database. The sentiment measure in the database is mainly based on Baker and Wurgler (2006). They find a significant contagious effect from CIS to volatilities of Australia, Hong Kong, and India stock indexes, while very weak evidence of contagion from CIS to returns is found. Ben-David, Franzoni, Kim, and Moussawi (2021) study specialized ETFs that hold stocks with salient characteristics including high past performance, media exposure, and sentiment. They document that stocks that are included in specialized ETFs experience, after launch, a steep drop in their media sentiment and earnings surprises relative to the pre-launch period. Henderson, Pearson, and Wang (2020) construct a new sentiment measure for individual stock using Structured Equity Product (SEP) issuances. Their SEP sentiment measure predicts negative abnormal returns on the SEPs' reference stocks. Boudoukh, Feldman, Kogan, and Richardson (2013) identify the relevance of news by type and tone. This is a better manner to pick up information that has impact on related firms. Their findings supports the relationship between stock price and information. Boudoukh, Feldman, Kogan, and Richardson (2019) apply textual analysis tools to identify news related to firms and isolate the portion of return variance solely due to the arrival of these events. They use Visual Information Extraction Platform (VIP) and Ravenpack to select events- and firms-related news and study the contributions of news arrivals on variances in trading hours and overnight hours. Their results support the idea that stock prices are closely linked to identified relevant news and Information accounts for more overnight idiosyncratic volatility than trading hour volatility.

Different firms experience different news coverages that certain firms capture more attentions. Researchers are interested in studying links among firms so that it is possible to obtain predictability. Schwenkler and Zheng (2019) shows that news contains information on distressed firm links and the news-implied relationship correlates with financial uncertainty and credit risk contagion. The links generate a channel, through which they obtain the predictability of returns and downgrades. Tao et al. (2020) study the news-based links from the perspective of news co-coverage and find that investors are not seizing the news-implied information fast, leading to positive cross-firm return predictability

3. Data and Methodologies

Our news text data comes from Dow Jones Data, News and Analytics (DNA) database, which contains daily news articles text data from leading publishers. Each record in the Dow Jones DNA is tagged by a list of firm codes, indicating firms that has been mentioned in the news article. The ranking of firms in the list implies the relativity, in which the first firm in the list is the most relative firm of the news. We select the most relevant firm and label the news with it. Each article in our sample has one and only one firm label. Our sample consist of articles published via The Wall Street Journal (J), The New York Times (NYTF), Reuters News (LBA), The Washington Post (WP), The Dallas Morning News (DAL), Chicago Sun-Times (CHI), and The San Francisco Chronicle (SFC) from January 1, 2005 to June 30, 2021. The source code of each news source will be used consistently in this paper. To ensure that each article obtained in our sample maintains enough information, we keep only articles with more than four sentences. We obtain stocks data from CRSP. Using the looking-up table from Dow Jones DNA with firm codes and tickers, we merge the news data with CRSP data. Earnings announcements data is obtained from I/B/E/S data. The final sample contains 223,117 firm-day news observations on 2,945 firms.

3.1. VADER Sentiment Analysis

VADER (Valence Aware Dictionary for Sentiment Reasoning) is a natural language model that maps lexical features to emotion intensities and displays the results as sentiment scores. Similar to other methods used in text sentiment analysis, VADER relies on a dictionary which assigns an intensity score to each word, and by summing up the score of each word in the text, we obtain a sentiment score of the text. Unlike other methods such as Harvard General Inquirer (GI) Dictionary in which score of each word is assigned by researchers, VADER is trained by Twitter posts and this feature enables VADER to capture the sentiments among investors. In stead of analyzing the sentiment of an article directly, we start with sentence-level sentiments. We split each article into sentences and VADER generates sentence-level sentiment scores. Then an average score, which is weighted by the number of words in each sentences, is obtained to represent the sentiment of a news article. Specifically,

$$Sentiment_{j,t} = \frac{\sum_{i=1}^{N_{j,t}} m_i * SenSentiment_{j,i}}{N_{j,t}} \tag{1}$$

.

where $Sentiment_{j,t}$ is the weighted average sentiment score of article j on day t, $N_{j,t}$ is the number of sentence in the article j on day t, m_i is the number of words in sentence i, $SenSentiment_{j,i}$ is the VADER score of sentence i in article j. The sentiment score results range from -1 (negative) to +1 (positive).

3.2. Summary Statistics

The distribution of sentiment scores is shown in Figure 1. It approximately follows normal distribution with mean of 0.0891. The 25% and 75% thresholds are 0.0125 and 0.1702. Sentiments below 0.0125 and sentiments above 0.1702 are regarded as extreme sentiments. We present summary statistics of sentiment analysis in Table 1. Panel A reports the summary statistics by news sources. The length of articles varies a lot from 5 sentences to 1,727 sentences an articles. Articles from The New York Times (NYTF) and The Washington Post (WP) are generally longer than articles from other publishers. The lowest sentiment is

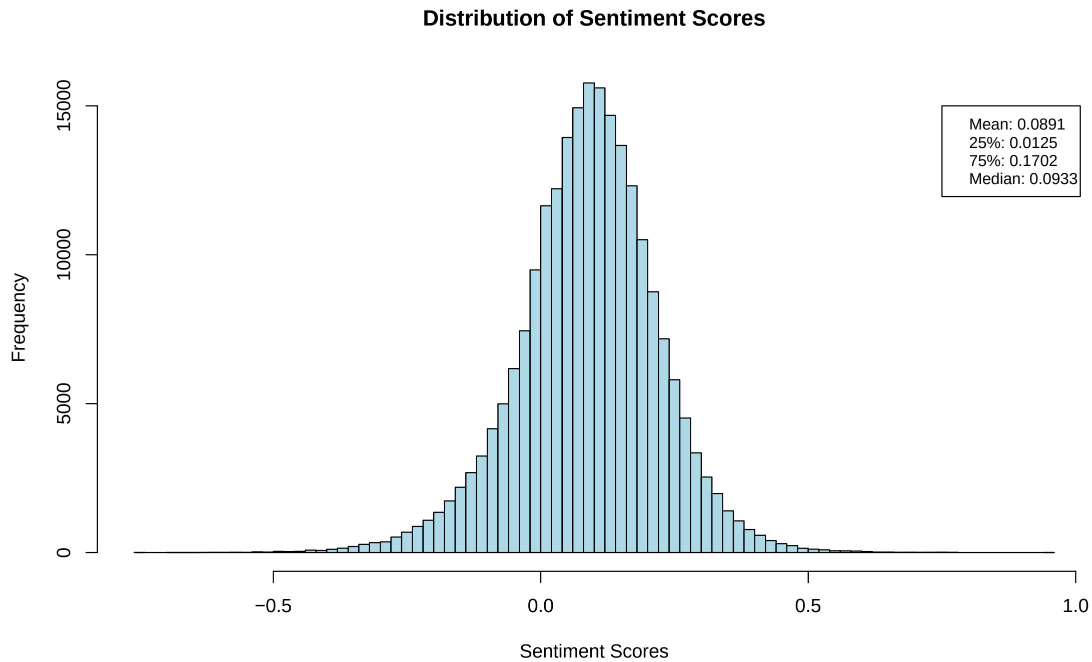

Fig. 1. Sentiment scores distribution.

-0.748 and the highest one is 0.949. All mean and median sentiments from our sample sources are slightly positive at around 0.1. We are interested in sentiment variation because we are reading articles sentence by sentence. We establish our beliefs every time when we finish one sentence and after finish reading the whole article, we have an updated understanding of the established belief. In some scenarios, an article can start with a negative sentiment and end with a positive sentiment, making the whole article a neutral one. In other scenarios, an articles can be composed with all neutral sentiment sentences, making the whole article a neutral one as well. But investors have different feelings when they read the 'neutral' articles and this leads to different trading behaviors. Due to the fact of this, the text structure matters. The statistics of sentiment score standard deviations offers evidence that articles have about 0.3 standard deviations on average and the max standard deviations can go up to around 0.8. Given the mean sentiment score is around 0.1, the standard deviations indicate that sentences in an article can vary from negative to positive. Panel B reports the summary statistics by year. There is a time-series trend in sentiment scores that sentiment scores are lower in the financial crisis and Covid-19 periods. We can observe the trend in Figure 2. However, the standard deviation does not have such trend that values are stable in the sample period. The key variables are sentiment scores and sentiment variations (measured by sentiment standard deviations) are visualized in Figure 2 and Figure 3.

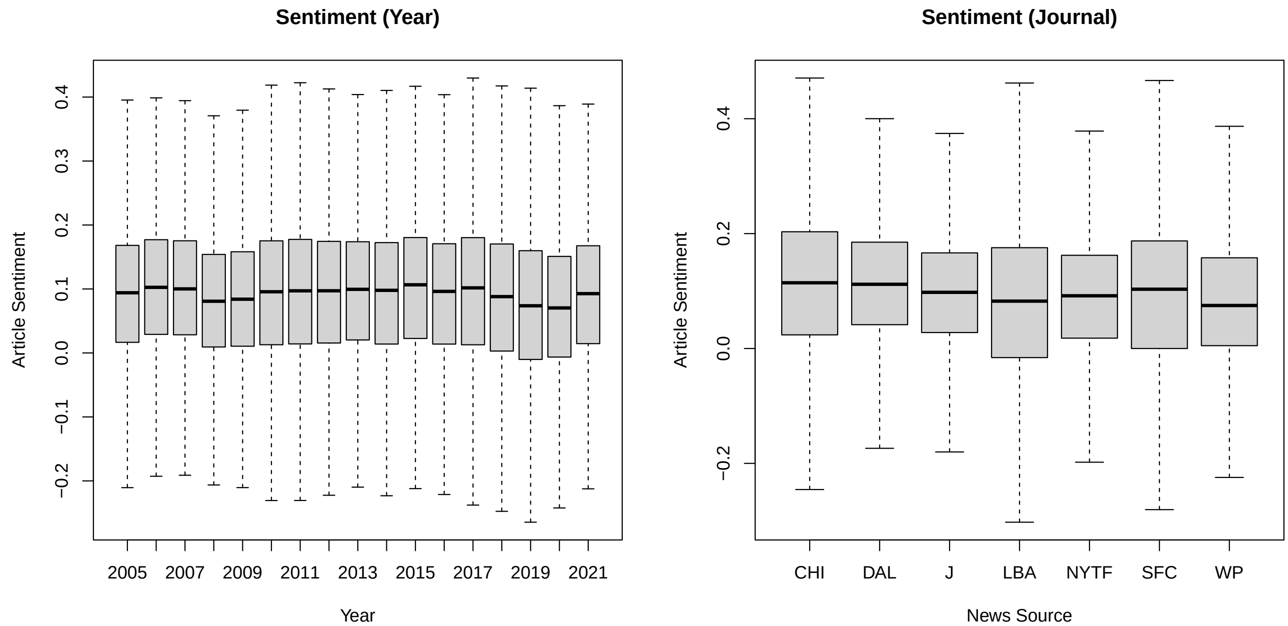

Fig. 2. Sentiment scores by year and news source.

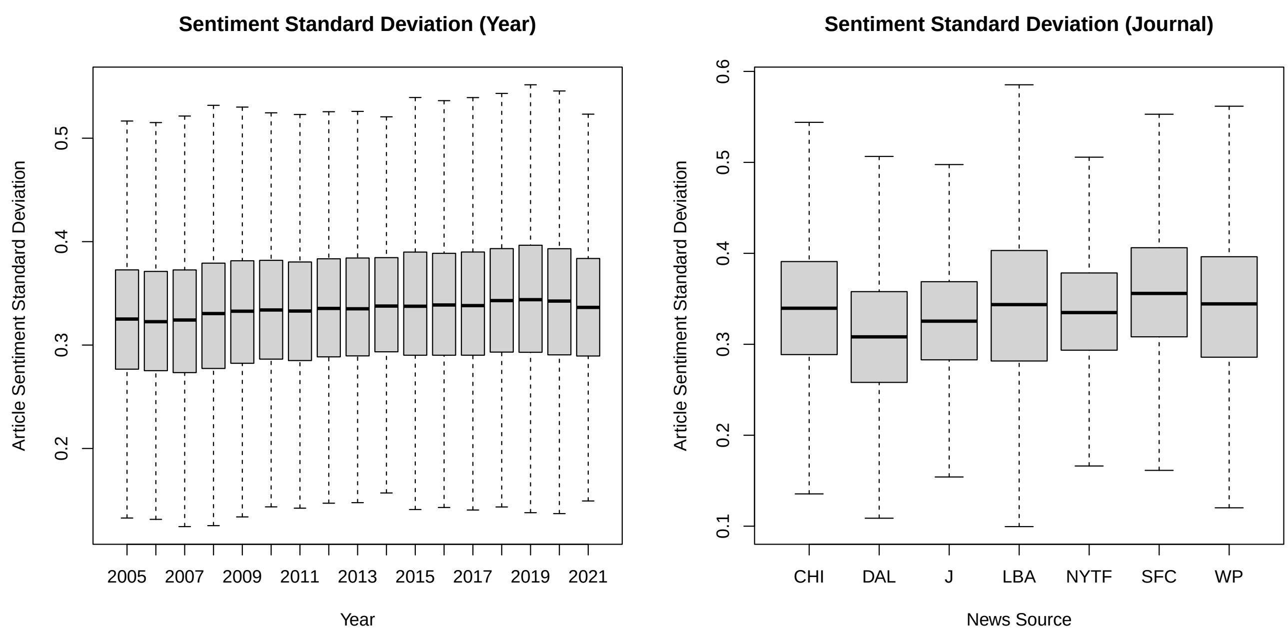

Fig. 3. Sentiment variations by year and news source.

Table 1: Summary Statistics

This table reports summary statistics of sentiment analysis. The minimum, mean, median, and maximum values are reported for each variable. *Sentence* and *Words* are the number of sentences and words in one article. *Sentiment* is article-level sentiment score calculated by weighted average sentence-level sentiment scores. *Std* is the standard deviation of sentence-level sentiment scores in one article, reflecting the sentiment variation.

Panel A: Summary Statistics by Source

	ALL	CHI	DAL	J	LBA	NYTF	SFC	WP
Observations	223117	9993	10789	79843	51870	36759	10981	22882
Sentence (min)	5	5	5	5	5	5	5	5
Sentence (mean)	31.765	23.486	29.082	30.799	16.868	41.882	29.597	58.571
Sentence (median)	23	18	22	26	14	35	25	29
Sentence (max)	1727	603	428	918	312	1393	632	1727
Words (min)	85	214	85	148	148	152	111	115
Words (mean)	3341	2582	2797	3339	2352	4410	3585	4352
Words (median)	2732	2102	2276	2874	1958	3883	3232	3571
Words (max)	108053	23053	35278	86643	47264	108053	40381	49419
Sentiment (min)	-0.748	-0.595	-0.61	-0.602	-0.748	-0.576	-0.554	-0.663
Sentiment (mean)	0.089	0.109	0.111	0.095	0.076	0.089	0.087	0.079
Sentiment (median)	0.093	0.114	0.112	0.098	0.082	0.092	0.103	0.075
Sentiment (max)	0.949	0.647	0.728	0.749	0.949	0.641	0.772	0.774
Std (min)	0	0	0	0	0	0	0	0
Std (mean)	0.332	0.339	0.305	0.325	0.340	0.335	0.357	0.328
Std (median)	0.334	0.34	0.308	0.325	0.344	0.335	0.356	0.344
Std (max)	0.822	0.78	0.689	0.705	0.822	0.697	0.762	0.714

Panel B: Summary Statistics by Year

	All	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021
Observations	223117	18190	16715	16339	17296	15625	14641	13933	13512	11774	11996	12944	12436	11145	11530	11060	10593	3388
Sentence (min)	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5
Sentence (mean)	31.765	33.350	34.520	32.023	31.692	31.571	30.261	29.108	30.305	31.201	30.944	26.649	26.244	27.184	31.678	42.831	35.494	45.188
Sentence (median)	23	24	23	22	22	23	24	23	25	26	25	21	21	21	22	23	22	30.5
Sentence (max)	1727	1727	1510	1364	889	983	721	875	831	753	804	785	1393	641	1119	1656	844	737
Words (min)	85	99	115	157	132	152	153	85	159	263	188	242	169	111	148	205	195	234
Words (mean)	3341	3380	3440	3206	3154	3164	3217	3192	3450	3647	3553	3129	3027	3145	3392	3741	3590	4363
Words (median)	2732	2664.5	2664	2568	2567.5	2650	2775	2753	2978	3151	3048	2590	2526.5	2546	2687	2830.5	2790	3656.5
Words (max)	108053	58904	86643	76122	35535	61986	74275	62437	49101	35375	44973	45340	108053	49070	63658	52184	45225	59218
Sentiment (min)	-0.748	-0.653	-0.596	-0.602	-0.613	-0.663	-0.576	-0.748	-0.689	-0.618	-0.538	-0.625	-0.595	-0.634	-0.537	-0.498	-0.533	-0.417
Sentiment (mean)	0.089	0.090	0.102	0.100	0.081	0.084	0.093	0.093	0.092	0.094	0.090	0.098	0.088	0.092	0.081	0.069	0.068	0.086
Sentiment (median)	0.093	0.094	0.103	0.100	0.081	0.084	0.096	0.097	0.097	0.099	0.098	0.107	0.096	0.102	0.088	0.074	0.070	0.093
Sentiment (max)	0.949	0.701	0.749	0.750	0.672	0.713	0.627	0.693	0.753	0.774	0.765	0.772	0.716	0.778	0.755	0.949	0.755	0.507
Std (min)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Std (mean)	0.332	0.322	0.321	0.321	0.324	0.328	0.332	0.331	0.335	0.336	0.339	0.340	0.340	0.340	0.342	0.340	0.338	0.332
Std (median)	0.334	0.325	0.323	0.324	0.330	0.333	0.334	0.333	0.335	0.335	0.338	0.338	0.339	0.338	0.343	0.344	0.343	0.336
Std (max)	0.822	0.711	0.741	0.780	0.771	0.716	0.703	0.748	0.822	0.736	0.786	0.722	0.742	0.743	0.683	0.755	0.672	0.608

4. News, Sentiment and Trading

4.1. News, Sentiment and Volumes

The impact of news sentiment has been broadly studied. We take the structure of text into consideration and study its impact. First, we analyze the impact on trading volumes. We consider the following regression,

$$\begin{aligned} Turnover_{i,t} = &\ \beta_1 Sentiment_{i,t} + \beta_2 Market\ Sentiment_t + \beta_3 Firm\text{–}Day\ News_{i,t} \\ &+ \beta_4 Day\ News_t + \beta_5 Size_{i,t} + \beta_6 News\ Variation_{i,t} + \beta_7 Earnings_{i,t} + \epsilon_{i,t}, \end{aligned} \tag{2}$$

where $Turnover_{i,t}$ is the log value of firm i trading turnover in day t, $Sentiment_{i,t}$ is the average sentiment of all news mentioned the firm i in day t, $Market\ Sentiment_t$ is the average sentiment of all news in day t, $Firm\text{–}Day\ News_{i,t}$ is the number of news on firm i in day t, $Day\ News_t$ is the total number of news in day t, $Size_i$ is the log value of firm i market capitalization, $News\ Variation_{i,t}$ is the average standard deviation of sentence-level sentiment in all news on firm i in day t, and $Earnings_{i,t}$ indicate whether there is a earnings announcement of firm i in day t.

Table 2 reports the results. Consistent with existing literature, our result shows that sentiment has positive and significant impact on turnover. Higher sentiments motivate trading in the market. Firm-level sentiment is the source of impact, while market-level sentiment does not have significant impact. News variation also has positive and significant impact on turnover. The magnitude is even stronger than sentiment itself which support our hypothesis that news variation motivates divergent expectation leading to more trades. Firm size has negative impact that larger firms are more stable. The number of news, which indicates the news exposure, has positive impact that when there are more articles on the firm, there are more trading on the firm. Earnings announcement motivates trading as well.

Then we consider what kind of firms are more likely to be affected. We consider the following regression,

$$\begin{aligned} Turnover_{i,t} = &\ \beta_1 Sentiment_{i,t} + \beta_2 Market\ Sentiment_t + \beta_3 Firm\text{–}Day\ News_{i,t} \\ &+ \beta_4 Day\ News_t + \beta_5 Size_{i,t} + \beta_6 News\ Variation_{i,t} \\ &+ \beta_7 Sentiment_{i,t} \times Size_{i,t} + \beta_8 Firm\text{–}Day\ News_{i,t} \times Size_{i,t} \\ &+ \beta_9 News\ Variation \times Size_{i,t} + \beta_{10} Earnings_{i,t} + \epsilon_{i,t}, \end{aligned} \tag{3}$$

where variables are the same as the variables in equation 2

Table 3 reports the results. Here, we focus on the impact on different firms. Results

Table 2: Sentiment and Turnover

	Dependent variable:		
	Turnover		
	(1)	(2)	(3)
Sentiment	0.177***	0.366***	0.180***
	(0.058)	(0.109)	(0.058)
Market Sentiment	−1.098	−0.047	0.059
	(0.775)	(0.335)	(0.270)
Firm-Day News	0.071***	0.099***	0.072***
	(0.009)	(0.012)	(0.009)
Day News	0.002	0.004***	0.004***
	(0.002)	(0.001)	(0.0004)
Size	−0.242**	−0.193***	−0.232**
	(0.091)	(0.033)	(0.084)
News Variation	0.027	0.813*	0.367***
	(0.220)	(0.400)	(0.122)
Earnings	0.553***	0.622***	0.591***
	(0.038)	(0.050)	(0.029)
Observations	160,983	160,983	160,983
FE	Firm	Time	Firm & Time
R^2	0.558	0.190	0.601
Adjusted R^2	0.552	0.190	0.594
Residual Std. Error	0.708 (df = 158588)	0.951 (df = 160959)	0.673 (df = 158572)
Note:			*p<0.1; **p<0.05; ***p<0.01

Table 3: Sentiment, Size and Turnover

	Dependent variable:		
	Turnover		
	(1)	(2)	(3)
Sentiment	0.171^{***}	0.330^{***}	0.182^{***}
	(0.019)	(0.023)	(0.018)
Market Sentiment	-1.031^{***}	0.037	0.120^{*}
	(0.071)	(0.099)	(0.071)
Firm-Day News	0.098^{***}	0.126^{***}	0.099^{***}
	(0.001)	(0.002)	(0.001)
Day News	0.002^{***}	0.004^{***}	0.004^{***}
	(0.0001)	(0.0002)	(0.0002)
Size	-0.149^{***}	-0.009	-0.147^{***}
	(0.008)	(0.010)	(0.008)
News Variation	0.285^{***}	1.512^{***}	0.562^{***}
	(0.058)	(0.073)	(0.057)
Earnings	0.548^{***}	0.617^{***}	0.588^{***}
	(0.016)	(0.022)	(0.016)
Sentiment $\times$Size	0.006	0.045^{***}	-0.001
	(0.009)	(0.011)	(0.008)
Fim-Day News $\times$ Size	-0.013^{***}	-0.014^{***}	-0.013^{***}
	(0.0005)	(0.001)	(0.0005)
News Variation $\times$ Size	-0.174^{***}	-0.501^{***}	-0.138^{***}
	(0.023)	(0.029)	(0.022)
Observations	160,983	160,983	160,983
FE	Firm	Time	Firm & Time
R^2	0.561	0.195	0.603
Adjusted R^2	0.554	0.195	0.597
Residual Std. Error	0.706 (df = 158585)	0.949 (df = 160956)	0.671 (df = 158569)

Note: *p<0.1; **p<0.05; ***p<0.01

on Table 2 show that more news exposure motivates trading. However, large firms are less affected by news exposure. When the firm size increases, the impact of firm news is decreasing. This indicates that small firms are more affected by news. The intuition is that large firms are always covered by media, the impact of an increase in news exposure is much weaker. But for small firms, they are often lack of attention and media coverage, so when there is an increasing news coverage, they capture the investors' attention. News variation shows similar pattern that higher variation motivates trading but the impact is weaker for large firms.

Next, we analyze the impact whether positive and negative environments work in the same way. We separate sentiment into sentiment below the average and sentiment above the average and consider the following regression,

$$
\begin{aligned}
Turnover_{i,t} = {} & \beta_1^{-} Sentiment_{i,t} \times I(Sentiment_{i,t} < \overline{Sentiment}) \\
& + \beta_1^{+} Sentiment_{i,t} \times I(Sentiment_{i,t} > \overline{Sentiment}) \\
& + \beta_2 Market\ Sentiment_t + \beta_3 Firm\text{–}Day\ News_{i,t} \\
& + \beta_4 Day\text{–}News_t + \beta_5 Size_{i,t} + \beta_6 News\ Variation_{i,t} \\
& + \beta_7^{-} Sentiment_{i,t} \times I(Sentiment_{i,t} < \overline{Sentiment}) \times Size_{i,t} \\
& + \beta_7^{+} Sentiment_{i,t} \times I(Sentiment_{i,t} > \overline{Sentiment}) \times Size_{i,t} \\
& + \beta_8 News\_Firm_{i,t} \times Size_{i,t} + \beta_9 News\_Variation \times Size_{i,t} \\
& + \beta_{10} Earnings_{i,t} + \epsilon_{i,t},
\end{aligned} \tag{4}
$$

where $\overline{Sentiment}$ is the mean sentiment in the sample period and other variables are the same as the variables in equation 3.

Table 4 reports the results. Both *Sentiment*$(+)$ and *Sentiment*$(-)$ have positive and significant coefficients, indicating that along with the increase in sentiment, there are more trading volumes. However, the coefficient of *Sentiment*$(-)$ is much larger than the coefficient of *Sentiment*$(+)$. The difference comes from the direction. To be more specific, *Sentiment*$(-)$ includes sentiments from negative to neutral and when sentiment is more closed to neutral, trading is more active. As for *Sentiment*$(+)$ which includes sentiment from neutral to positive, when sentiment increases, it moves away from the neutral part. Although higher sentiment is triggering trades, the impact is weaker compared to sentiment that moves towards neutral area.

Table 4: Directional Sentiment, Size and Turnover

	Dependent variable:		
	Turnover		
	(1)	(2)	(3)
Firm-Day News	0.097***	0.126***	0.098***
	(0.001)	(0.002)	(0.001)
Day News	0.002***	0.004***	0.004***
	(0.0001)	(0.0002)	(0.0002)
Size	−0.150***	−0.008	−0.149***
	(0.008)	(0.010)	(0.008)
News Variation	0.297***	1.529***	0.572***
	(0.058)	(0.073)	(0.057)
Earnings	0.548***	0.617***	0.587***
	(0.016)	(0.022)	(0.016)
Sentiment (-)	0.561***	0.444***	0.441***
	(0.051)	(0.064)	(0.048)
Sentiment (+)	0.029	0.287***	0.088***
	(0.025)	(0.032)	(0.024)
Market Sentiment (-)	1.527	6.749***	2.956*
	(1.800)	(2.421)	(1.719)
Market Sentiment (+)	−1.043***	−0.004	0.103
	(0.072)	(0.101)	(0.072)
Firm-Day News × Size	−0.013***	−0.014***	−0.013***
	(0.0005)	(0.001)	(0.0005)
News Variation × Size	−0.176***	−0.501***	−0.139***
	(0.023)	(0.029)	(0.022)
Sentiment (-) × Size	−0.038	0.064**	−0.034
	(0.024)	(0.030)	(0.023)
Sentiment (+) × Size	0.020*	0.037**	0.009
	(0.012)	(0.015)	(0.011)
Observations	160,983	160,983	160,983
FE	Firm	Time	Firm & Time
R^2	0.561	0.195	0.603
Adjusted R^2	0.554	0.195	0.597
Residual Std. Error	0.706 (df = 158582)	0.948 (df = 160953)	0.671 (df = 158566)
Note:			*p<0.1; **p<0.05; ***p<0.01

4.2. News, Sentiment and Volatility

We study the impact of sentiment variation on stock volatility. First, we consider the following regression,

$$\begin{aligned} Volatility_{i,t} = &\beta_1 Sentiment_{i,t} + \beta_2 Market\,Sentiment_t + \beta_3 Firm\text{–}Day\,News_{i,t} \\ &+ \beta_4 Day\,News_t + \beta_5 Size_{i,t} + \beta_6 News\,Variation_{i,t} + \beta_7 Earnings_{i,t} + \epsilon_{i,t}, \end{aligned} \tag{5}$$

where $Volatility_{i,j}$ is the stock price volatility of firm i in day j and it is measured by log(High/Low), $Sentiment_{i,t}$ is the average sentiment of all news mentioned the firm i in day t, $Market\,Sentiment_t$ is the average sentiment of all news in day t, $Firm\text{–}Day\,News_{i,t}$ is the number of news on firm i in day t, $Day\,News_t$ is the total number of news in day t, $Size_i$ is the log value of firm i market capitalization, $News\,Variation_{i,t}$ is the average standard deviation of sentence-level sentiment in all news on firm i in day t, and $Earnings_{i,t}$ indicate whether there is a earnings announcement of firm i in day t.

Table 5 reports the results. Volatility is not significantly affected by firm idiosyncratic news, while it is affected by market-wide sentiment. A higher market sentiment helps stabilize the market resulting in a low volatility. More news exposures increases volatility, no matter whether these articles are all about the target firm or about other firms in the market. The strongest impact comes from news variation that when articles deliver more variation, more volatility can be observed in the market.

Then we take consideration of firm size. We consider the following regression,

$$\begin{aligned} Volatility_{i,t} = &\beta_1 Sentiment_{i,t} + \beta_2 Market\,Sentiment_t + \beta_3 Firm\text{–}Day\,News_{i,t} \\ &+ \beta_4 Day\,News_t + \beta_5 Size_{i,t} + \beta_6 News\,Variation_{i,t} \\ &+ \beta_7 Sentiment_{i,t} \times Size_{i,t} + \beta_8 Firm\text{–}Day\,News_{i,t} \times Size_{i,t} \\ &+ \beta_9 News\,Variation \times Size_{i,t} + \beta_{10} Earnings_{i,t} + \epsilon_{i,t}, \end{aligned} \tag{6}$$

where variables are the same as the variables in regression 5

Table 6 reports the results. Generally, high sentiment helps stabilize market, but volatility is increasing when large firms are mentioned more by positive sentiment news. Meanwhile, large firms are less affected by the level of news coverage that when there is abundant news, large firms are less volatile than small firms.

Next, we analyze the directional sentiment that we separate sentiment into positive sen-

Table 5: Sentiment and Volatility

	Dependent variable:		
	Volatility		
	(1)	(2)	(3)
Sentiment	0.001	0.001	0.001
	(0.001)	(0.001)	(0.001)
Market Sentiment	-0.130^{***}	-0.045^{***}	-0.035^{***}
	(0.004)	(0.005)	(0.004)
Firm-Day News	0.004^{***}	0.004^{***}	0.003^{***}
	(0.0001)	(0.0001)	(0.0001)
Day News	0.00003^{***}	0.0001^{***}	0.0001^{***}
	(0.00001)	(0.00001)	(0.00001)
Size	-0.016^{***}	-0.009^{***}	-0.017^{***}
	(0.0001)	(0.0001)	(0.0001)
News Variation	0.002	0.050^{***}	0.030^{***}
	(0.003)	(0.003)	(0.003)
Earnings	0.015^{***}	0.013^{***}	0.013^{***}
	(0.001)	(0.001)	(0.001)
Observations	160,983	160,983	160,983
FE	Firm	Time	Firm & Time
R^2	0.299	0.208	0.349
Adjusted R^2	0.288	0.207	0.339
Residual Std. Error	0.041 (df = 158588)	0.043 (df = 160959)	0.040 (df = 158572)
Note:			*p<0.1; **p<0.05; ***p<0.01

Table 6: Sentiment, Size and Volatility

	Dependent variable:		
	Volatility		
	(1)	(2)	(3)
Sentiment	-0.005^{***}	-0.002^{**}	-0.004^{***}
	(0.001)	(0.001)	(0.001)
Market Sentiment	-0.122^{***}	-0.034^{***}	-0.028^{***}
	(0.004)	(0.004)	(0.004)
Firm-Day News	0.008^{***}	0.009^{***}	0.007^{***}
	(0.0001)	(0.0001)	(0.0001)
Day News	0.00004^{***}	0.0001^{***}	0.0001^{***}
	(0.00001)	(0.00001)	(0.00001)
Size	-0.009^{***}	-0.0004	-0.010^{***}
	(0.0005)	(0.0004)	(0.0004)
News Variation	0.011^{***}	0.073^{***}	0.037^{***}
	(0.003)	(0.003)	(0.003)
Earnings	0.014^{***}	0.013^{***}	0.012^{***}
	(0.001)	(0.001)	(0.001)
Sentiment × Size	0.005^{***}	0.005^{***}	0.004^{***}
	(0.0005)	(0.0005)	(0.0005)
Firm-Day News × Size	-0.002^{***}	-0.002^{***}	-0.002^{***}
	(0.00003)	(0.00003)	(0.00003)
News Variation × Size	-0.007^{***}	-0.016^{***}	-0.006^{***}
	(0.001)	(0.001)	(0.001)
Observations	160,983	160,983	160,983
FE	Firm	Time	Firm & Time
R^2	0.321	0.240	0.370
Adjusted R^2	0.311	0.240	0.360
Residual Std. Error	0.040 (df = 158585)	0.042 (df = 160956)	0.039 (df = 158569)
Note:			*p<0.1; **p<0.05; ***p<0.01

timent and negative and consider the following regression,

$$
\begin{aligned}
Volatility_{i,t} = & \beta_1^- Sentiment_{i,t} \times I(Sentiment_{i,t} < \overline{Sentiment}) \\
& + \beta_1^+ Sentiment_{i,t} \times I(Sentiment_{i,t} > \overline{Sentiment}) \\
& + \beta_2 Market\ Sentiment_t + \beta_3 Firm\text{–}Day\ News_{i,t} \\
& + \beta_4 Day\ News_t + \beta_5 Size_{i,t} + \beta_6 News\ Variation_{i,t} \\
& + \beta_7^- Sentiment_{i,t} \times I(Sentiment_{i,t} < \overline{Sentiment}) \times Size_{i,t} \\
& + \beta_7^+ Sentiment_{i,t} \times I(Sentiment_{i,t} > \overline{Sentiment}) \times Size_{i,t} \\
& + \beta_8 Firm\text{–}Day\ News_{i,t} \times Size_{i,t} + \beta_9 News\ Variation \times Size_{i,t} \\
& + \beta_{10} Earnings_{i,t} + \epsilon_{i,t},
\end{aligned}
\tag{7}
$$

where $Sentiment$ is the mean sentiment in the sample period and other variables are the same as the variables in regression 6.

Table 7 reports the results. The impact from sentiment shows similar pattern as we study its impact on turnovers that the coefficient is large at 0.441 when it is moving from negative to neutral, and the coefficient drops to 0.088 when it starts at neutral and moves up to positive. The number of news on firms and news variations both have positive and significant impact on volatility. However, firms of different sizes experience differently that large firms are less likely to be affected, while small firms are more affected by news.

4.3. News, Sentiment and Returns

After studying the impacts on turnover and volatility, we would like to analyze the impacts on return. We consider the following regression,

$$
\begin{aligned}
Return_{i,t} = & \beta_1 Sentiment_{i,t} + \beta_2 Market\ Sentiment_t + \beta_3 Firm\text{–}Day\ News_{i,t} \\
& + \beta_4 Day\ News_t + \beta_5 Size_{i,t} + \beta_6 News\ Variation_{i,t} + \beta_7 Earnings_{i,t} + \epsilon_{i,t},
\end{aligned}
\tag{8}
$$

where $Return_{i,t}$ is the log return of firm i trading turnover in day t, $Sentiment_{i,t}$ is the average sentiment of all news mentioned the firm i in day t, $Market\ Sentiment_t$ is the average sentiment of all news in day t, $Firm\text{–}Day\ News_{i,t}$ is the number of news on firm i in day t, $Day\ News_t$ is the total number of news in day t, $Size_i$ is the log value of firm i market capitalization, $News\ Variation_{i,t}$ is the average standard deviation of sentence-level sentiment in all news on firm i in day t, and $Earnings_{i,t}$ indicate whether there is a earnings announcement of firm i in day t.

The results are shown in Table 8. Sentiment has positive and significant impact on return.

Table 7: Directional Sentiment, Size and Volatility

	Dependent variable:		
	Volatility		
	(1)	(2)	(3)
Firm-Day News	0.008***	0.009***	0.007***
	(0.0001)	(0.0001)	(0.0001)
Day News	0.00004***	0.0001***	0.0001***
	(0.00001)	(0.00001)	(0.00001)
Size	−0.009***	−0.001	−0.010***
	(0.0005)	(0.0004)	(0.0004)
News Variation	0.013***	0.075***	0.039***
	(0.003)	(0.003)	(0.003)
Earnings	0.014***	0.013***	0.012***
	(0.001)	(0.001)	(0.001)
Sentiment (-)	0.018***	0.016***	0.011***
	(0.003)	(0.003)	(0.003)
Sentiment (+)	−0.013***	−0.009***	−0.009***
	(0.001)	(0.001)	(0.001)
Market Sentiment (-)	0.452***	0.711***	0.617***
	(0.103)	(0.108)	(0.100)
Market Sentiment (+)	−0.125***	−0.039***	−0.032***
	(0.004)	(0.004)	(0.004)
Firm-Day News × Size	−0.002***	−0.002***	−0.002***
	(0.00003)	(0.00003)	(0.00003)
News Variation × Size	−0.007***	−0.016***	−0.006***
	(0.001)	(0.001)	(0.001)
Sentiment (-) × Size	−0.003**	−0.003**	−0.002*
	(0.001)	(0.001)	(0.001)
Sentiment (+) × Size	0.008***	0.008***	0.006***
	(0.001)	(0.001)	(0.001)
Observations	160,983	160,983	160,983
FE	Firm	Time	Firm & Time
R^2	0.321	0.241	0.370
Adjusted R^2	0.311	0.240	0.360
Residual Std. Error	0.040 (df = 158582)	0.042 (df = 160953)	0.039 (df = 158566)
Note:			*p<0.1; **p<0.05; ***p<0.01

Table 8: Sentiment and Return

	Dependent variable:		
	Return		
	(1)	(2)	(3)
Sentiment	0.009^{***}	0.009^{***}	0.009^{***}
	(0.002)	(0.002)	(0.002)
Market Sentiment	0.021^{*}	0.018^{**}	0.013
	(0.010)	(0.008)	(0.009)
Firm-Day News	-0.001^{***}	-0.001^{***}	-0.001^{***}
	(0.0002)	(0.0002)	(0.0002)
Day News	0.00004	0.00000	0.00000
	(0.00004)	(0.00003)	(0.00003)
Size	0.004^{***}	0.001^{***}	0.004^{***}
	(0.001)	(0.0003)	(0.001)
News Variation	0.012	0.006	0.011
	(0.008)	(0.010)	(0.009)
Earnings	−0.002	0.0001	−0.0003
	(0.002)	(0.002)	(0.002)
Observations	160,811	160,811	160,811
FE	Firm	Time	Firm & Time
R^2	0.086	0.006	0.087
Adjusted R^2	0.072	0.005	0.073
Residual Std. Error	0.048 (df = 158423)	0.049 (df = 160787)	0.047 (df = 158407)
Note:			*p<0.1; **p<0.05; ***p<0.01

Good news usually releases positive signals to the market which results in a higher return. There is no evidence that text structure matters. High variation in text can generate more divergent expectations which motivates trading, but the impact on return is not dependent on structure. The results also show that the number of news on firms has negative impact, indicating that more news in a day lowers the return.

Since size plays a role in affecting turnover and volatility, we are interested in whether size influences returns. Then we consider the following regression,

$$\begin{aligned} Return_{i,t} = {} & \beta_1 Sentiment_{i,t} + \beta_2 Market\,Sentiment_t + \beta_3 Firm\text{–}Day\,News_{i,t} \\ & + \beta_4 Day\,News_t + \beta_5 Size_{i,t} + \beta_6 News\,Variation_{i,t} \\ & + \beta_7 Sentiment_{i,t} \times Size_{i,t} + \beta_8 Firm\text{–}Day\,News_{i,t} \times Size_{i,t} \\ & + \beta_9 News\,Variation \times Size_{i,t} + \beta_{10} Earnings_{i,t} + \epsilon_{i,t}, \end{aligned} \tag{9}$$

where variables are the same as the variables in regression 8

Table 9 shows the results. The results are consistent with previous findings that sentiment has positive and significant impact on return while news variation has no impact. The number of news a day on firm negatively affects return, and our results indicate that the impact varies for different firms. Large firms are not experiencing the negative impact from the increasing number of news, and the negative impact works more on small firms. At the same time, large firms are less likely to be affected by sentiment. Market-wide sentiment also positively affects return.

Next, we analyze the directional sentiment that we separate sentiment into positive sentiment and negative and consider the following regression,

$$\begin{aligned} Return_{i,t} = {} & \beta_1^- Sentiment_{i,t} \times I(Sentiment_{i,t} < \overline{Sentiment}) \\ & + \beta_1^+ Sentiment_{i,t} \times I(Sentiment_{i,t} > \overline{Sentiment}) \\ & + \beta_2 Market\,Sentiment_t + \beta_3 Firm\text{–}Day\,News_{i,t} \\ & + \beta_4 Day\,News_t + \beta_5 Size_{i,t} + \beta_6 News\,Variation_{i,t} \\ & + \beta_7^- Sentiment_{i,t} \times I(Sentiment_{i,t} < \overline{Sentiment}) \times Size_{i,t} \\ & + \beta_7^+ Sentiment_{i,t} \times I(Sentiment_{i,t} > \overline{Sentiment}) \times Size_{i,t} \\ & + \beta_8 News\_Firm_{i,t} \times Size_{i,t} + \beta_9 News\_Variation \times Size_{i,t} \\ & + \beta_{10} Earnings_{i,t} + \epsilon_{i,t}, \end{aligned} \tag{10}$$

where $\overline{Sentiment}$ is the mean sentiment in the sample period and other variables are the same as the variables in regression 9.

Table 9: Sentiment, size and Return

	Dependent variable:		
	Return		
	(1)	(2)	(3)
Sentiment	0.013***	0.014***	0.013***
	(0.001)	(0.001)	(0.001)
Market Sentiment	0.018***	0.015***	0.011**
	(0.005)	(0.005)	(0.005)
Firm-Day News	−0.002***	−0.003***	−0.002***
	(0.0001)	(0.0001)	(0.0001)
Day News	0.00004***	0.00000	0.00001
	(0.00001)	(0.00001)	(0.00001)
Size	0.002***	−0.0004	0.003***
	(0.001)	(0.001)	(0.001)
News Variation	0.014***	0.002	0.013***
	(0.004)	(0.004)	(0.004)
Earnings	−0.001	0.0004	−0.0002
	(0.001)	(0.001)	(0.001)
Sentiment × Size	−0.004***	−0.005***	−0.004***
	(0.001)	(0.001)	(0.001)
Firm-Day News × Size	0.001***	0.001***	0.001***
	(0.00003)	(0.00003)	(0.00003)
News Variation × Size	−0.001	0.002	−0.001
	(0.002)	(0.002)	(0.002)
Observations	160,811	160,811	160,811
FE	Firm	Time	Firm & Time
R^2	0.089	0.010	0.090
Adjusted R^2	0.075	0.010	0.077
Residual Std. Error	0.047 (df = 158420)	0.049 (df = 160784)	0.047 (df = 158404)
Note:			*p<0.1; **p<0.05; ***p<0.01

Table 10: Directional sentiment, size and Return

	Dependent variable:		
	Return		
	(1)	(2)	(3)
Firm-Day News	-0.003^{***}	-0.003^{***}	-0.002^{***}
	(0.0001)	(0.0001)	(0.0001)
Day News	0.00004^{***}	0.00000	0.00001
	(0.00001)	(0.00001)	(0.00001)
Size	0.002^{***}	-0.0004	0.003^{***}
	(0.001)	(0.001)	(0.001)
News Variation	0.013^{***}	0.001	0.012^{***}
	(0.004)	(0.004)	(0.004)
Earnings	-0.001	0.0005	-0.0001
	(0.001)	(0.001)	(0.001)
Sentiment (-)	0.009^{***}	0.010^{***}	0.009^{***}
	(0.003)	(0.003)	(0.003)
Sentiment (+)	0.014^{***}	0.015^{***}	0.014^{***}
	(0.002)	(0.002)	(0.002)
Market Sentiment (-)	-0.766^{***}	-0.765^{***}	-0.768^{***}
	(0.121)	(0.125)	(0.121)
Market Sentiment (+)	0.024^{***}	0.020^{***}	0.016^{***}
	(0.005)	(0.005)	(0.005)
Firm-Day News × Size	0.001^{***}	0.001^{***}	0.001^{***}
	(0.00003)	(0.00003)	(0.00003)
News Variation × Size	-0.001	0.002	-0.001
	(0.002)	(0.002)	(0.002)
Sentiment (-) × Size	-0.005^{***}	-0.005^{***}	-0.005^{***}
	(0.002)	(0.002)	(0.002)
Sentiment (+) × Size	-0.003^{***}	-0.005^{***}	-0.003^{***}
	(0.001)	(0.001)	(0.001)
Observations	160,811	160,811	160,811
FE	Firm	Time	Firm & Time
R^2	0.089	0.011	0.091
Adjusted R^2	0.076	0.010	0.077
Residual Std. Error	0.047 (df = 158417)	0.049 (df = 160781)	0.047 (df = 158401)
Note:			*p<0.1; **p<0.05; ***p<0.01

The results are shown in Table 10. We can observe similar results in Table 9 that sentiment and number of daily news on firm have positive and negative impacts on return respectively. Market sentiment shows different impacts from different directions. The coefficient of $Market\,Sentiment(-)$ is -0.768, suggesting when the market-wide sentiment goes from negative to neutral, the return is decreasing. The coefficient of $Market\,Sentiment(+)$ is 0.016 that when sentiment moves more positive, the return is higher. Given the results here, the relationship between return and market sentiment is not linear that the neutral part, which motivates more divergent expectations, lowers the return.

4.4. The Impact of Neutral News

From Table 4 we observe $Sentiment(+)$ and $Sentiment(-)$ have different impacts on turnover, and $Sentiment(+)$ has much lower coefficient suggesting that when sentiment is above the average, the impact is much lower. To further analyze the different impacts from sentiment, we consider the following regression,

$$\begin{aligned} Turnover_{i,t} = {} & \beta_1 Sentiment(H/L)_{i,t} + \beta_2 Market\,Sentiment_t + \beta_3 Firm\text{–}Day\,News_{i,t} \\ & + \beta_4 Day\,News_t + \beta_5 Size_{i,t} + \beta_6 News\,Variation_{i,t} \\ & + \beta_7 Earnings_{i,t} + \epsilon_{i,t}, \end{aligned} \tag{11}$$

where $Return_{i,t}$ is the log return of firm i trading turnover in day t, $Market\_Sentiment_t$ is the average sentiment of all news in day t, $News\_Firm_{i,t}$ is the number of news on firm i in day t, $News\_Market_t$ is the total number of news in day t, $Size_i$ is the log value of firm i market capitalization, $News\_Variation_{i,t}$ is the average standard deviation of sentence-level sentiment in all news on firm i in day t, and $Earnings_{i,t}$ indicate whether there is a earnings announcement of firm i in day t. For $Sentiment(H/L)_{i,t}$, we separate the whole sample into two groups that $Sentiment(H)_{i,t}$ includes only observations of which sentiment score is greater than the mean score of our sample, while $Sentiment(L)_{i,t}$ includes the lower part.

Table 11: Positive/Negative sentiment and Turnover

	Dependent variable:					
	Turnover					
	(1)	(2)	(3)	(4)	(5)	(6)
Sentiment (H/L)	−0.129***	−0.045	−0.089***	0.387***	0.481***	0.326***
	(0.036)	(0.047)	(0.035)	(0.032)	(0.042)	(0.030)
Market Sentiment	−0.761***	−0.074	0.091	−1.633***	−0.339**	−0.217**
	(0.100)	(0.142)	(0.101)	(0.100)	(0.142)	(0.100)
Firm-Day News	0.073***	0.117***	0.079***	0.068***	0.087***	0.067***
	(0.002)	(0.002)	(0.002)	(0.001)	(0.001)	(0.001)
Day News	0.003***	0.004***	0.005***	0.002***	0.004***	0.004***
	(0.0002)	(0.0003)	(0.0002)	(0.0002)	(0.0003)	(0.0002)
Size	−0.227***	−0.196***	−0.223***	−0.246***	−0.195***	−0.226***
	(0.003)	(0.002)	(0.004)	(0.003)	(0.002)	(0.004)
News Variation	−0.035	0.052	0.272***	0.020	1.402***	0.420***
	(0.066)	(0.092)	(0.066)	(0.062)	(0.081)	(0.061)
Earnings	0.503***	0.601***	0.550***	0.611***	0.649***	0.640***
	(0.022)	(0.029)	(0.021)	(0.025)	(0.034)	(0.024)
Observations	82,575	82,575	82,575	78,408	78,408	78,408
FE	Firm	Time	Firm & Time	Firm	Time	Firm & Time
R^2	0.564	0.173	0.601	0.583	0.214	0.631
Adjusted R^2	0.553	0.173	0.591	0.573	0.214	0.622
Residual Std. Error	0.704 (df = 80522)	0.958 (df = 82551)	0.674 (df = 80506)	0.693 (df = 76525)	0.940 (df = 78384)	0.652 (df = 76509)

Note: $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

Table 11 reports the results. In column (1) - (3), we consider sentiments above the average. In our definition, sentiment around the mean value is considered as neutral sentiment. The coefficient of $Sentiment(H)$ is -0.089, suggesting when the sentiment moves away from neutral sentiment, turnover decreases. In column (4) - (6), we use sentiment below the average. The positive coefficient (0.326) indicates that when sentiment moves from negative to neutral sentiment, turnover increases. According to the results, neutral sentiments have stronger impact on turnover. Neutral sentiment is not as clear as positive or negative sentiments and investors would explain the signals from their own perspectives, resulting in more trading activities.

To further test the impact of neutral sentiment, we consider the following regression,

$$\begin{aligned} Turnover_{i,t} = \beta_1 Sentiment(Tail/Neutral)_{i,t} + \beta_2 Market\_Sentiment_t + \beta_3 News\_Firm_{i,t} \\ + \beta_4 News\_Market_t + \beta_5 Size_{i,t} + \beta_6 News\_Variation_{i,t} + \epsilon_{i,t}, \end{aligned} \tag{12}$$

where $Return_{i,t}$ is the log return of firm i trading turnover in day t, $Market\_Sentiment_t$ is the average sentiment of all news in day t, $News\_Firm_{i,t}$ is the number of news on firm i in day t, $News\_Market_t$ is the total number of news in day t, $Size_i$ is the log value of firm i market capitalization, and $News\_Variation_{i,t}$ is the average standard deviation of sentence-level sentiment in all news on firm i in day t. $Sentiment(Tail)_{i,t}$ includes sentiments scores which are in the bottom and the top 25% of the sample sentiment scores, while $Sentiment(Tail)_{i,t}$ are sentiment scores which are in the middle 50%.

We select 25% and 75% quartile values of the sample sentiment scores as the thresholds and separate our sample into two groups. The neutral sentiment group contains sentiment values that are greater than the bottom 25% sentiment scores and smaller than the top 25% sentiment scores, while the tail sentiment group includes sentiment values which are smaller than the bottom 25% and greater than the top 25%.

Table 12 shows the results. Previous results suggest that sentiment has positive significant impact on turnover. We can find positive and significant coefficients of both neutral sentiment and tail sentiment. However, the coefficient of neutral sentiment is 0.255 which is much higher than the coefficient of tail sentiment (0.173), indicating that sentiment in the neutral group has greater impact rather than sentiment on two tails. This offers evidence to support our hypothesis that neutral sentiments are not clear as positive or negative sentiments, and this creates more space for divergent expectations and interpretations of information so that there are more trading activities in the market.

Table 12: Tail Sentiments and Neutral Sentiments

	Dependent variable:		
	Turnover		
	(1)	(2)	(3)
Market Sentiment	-1.100^{***}	-0.057	0.056
	(0.071)	(0.100)	(0.071)
Firm-Day News	0.071^{***}	0.099^{***}	0.072^{***}
	(0.001)	(0.001)	(0.001)
Day News	0.002^{***}	0.004^{***}	0.004^{***}
	(0.0001)	(0.0002)	(0.0002)
Size	-0.242^{***}	-0.193^{***}	-0.232^{***}
	(0.002)	(0.001)	(0.003)
News Variation	0.026	0.811^{***}	0.366^{***}
	(0.046)	(0.061)	(0.045)
Earnings	0.553^{***}	0.623^{***}	0.591^{***}
	(0.017)	(0.022)	(0.016)
Sentiment (Neutral)	0.218^{***}	0.562^{***}	0.255^{***}
	(0.036)	(0.047)	(0.034)
Sentiment (Tail)	0.174^{***}	0.347^{***}	0.173^{***}
	(0.016)	(0.021)	(0.016)
Observations	160,983	160,983	160,983
FE	Firm	Time	Firm & Time
R^2	0.558	0.190	0.601
Adjusted R^2	0.552	0.190	0.595
Residual Std. Error	0.708 (df = 158587)	0.951 (df = 160958)	0.673 (df = 158571)
Note:			*p<0.1; **p<0.05; ***p<0.01

5. Conclusion

In this paper, we study the impact of news on stock market from two perspectives: sentiment and sentiment variation. Starting from VADER sentiment analysis tool, we conduct sentence-level sentiment analysis which further allows us to measure the news implied sentiment variation within articles. Our results support the findings in existing literature that positive sentiment contributes to more trading. Furthermore, we find that the impact from sentiment is stronger for small firms. We also find that sentiment variation has greater impact than sentiment itself, which results in higher trading volumes. Our analysis shows that news exposure increases trading volumes but decrease returns. Existing literature often focuses on large firms which generally have more media exposure, but small firms are those who experience stronger influence from sentiment, sentiment variation and news exposure. We provide evidence that neutral news has stronger impact than positive or negative news since neutral sentiment creates more space for divergent expectations and different interpretations, encouraging more trading in the market.

References

Baker, M., Wurgler, J., 2006. Investor sentiment and the cross-section of stock returns. The journal of Finance 61, 1645–1680.

Baker, M., Wurgler, J., 2007. Investor sentiment in the stock market. Journal of economic perspectives 21, 129–152.

Baker, M., Wurgler, J., Yuan, Y., 2012. Global, local, and contagious investor sentiment. Journal of financial economics 104, 272–287.

Baker, S. R., Bloom, N., Davis, S. J., Sammon, M. C., 2021. What triggers stock market jumps? .

Barbaglia, L., Consoli, S., Manzan, S., 2023. Forecasting with economic news. Journal of Business & Economic Statistics 41, 708–719.

Ben-David, I., Franzoni, F., Kim, B., Moussawi, R., 2021. Competition for attention in the etf space. Tech. rep., National Bureau of Economic Research.

Black, F., 1986. Noise. The journal of finance 41, 528–543.

Boudoukh, J., Feldman, R., Kogan, S., Richardson, M., 2013. Which news moves stock prices? a textual analysis .

Boudoukh, J., Feldman, R., Kogan, S., Richardson, M., 2019. Information, trading, and volatility: Evidence from firm-specific news. The Review of Financial Studies 32, 992–1033.

Chen, C. Y.-H., Després, R., Guo, L., Renault, T., 2019. What makes cryptocurrencies special? investor sentiment and return predictability during the bubble .

Das, S. R., Chen, M. Y., 2007. Yahoo! for amazon: Sentiment extraction from small talk on the web. Management science 53, 1375–1388.

De Long, J. B., Shleifer, A., Summers, L. H., Waldmann, R. J., 1990. Noise trader risk in financial markets. Journal of political Economy 98, 703–738.

Hart, R. P., 2000. Diction 5.0. Austin, TX: Digitex .

Henderson, B. J., Pearson, N. D., Wang, L., 2020. Retail derivatives and sentiment: A sentiment measure constructed from issuances of retail structured equity products. Available at SSRN 3642883 .

Huang, D., Jiang, F., Tu, J., Zhou, G., 2015. Investor sentiment aligned: A powerful predictor of stock returns. The Review of Financial Studies 28, 791–837.

Ke, Z. T., Kelly, B. T., Xiu, D., 2019. Predicting returns with text data .

Kyle, A. S., 1985. Continuous auctions and insider trading. Econometrica: Journal of the Econometric Society pp. 1315–1335.

Li, X., 2019. Does chinese investor sentiment predict asia-pacific stock markets? evidence from a nonparametric causality-in-quantiles test. Finance Research Letters p. 101395.

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

Manela, A., Moreira, A., 2017. News implied volatility and disaster concerns. Journal of Financial Economics 123, 137–162.

Merton, R. C., 1980. On estimating the expected return on the market: An exploratory investigation. Journal of financial economics 8, 323–361.

Saurabh, S., Dey, K., 2020. Unraveling the relationship between social moods and the stock market: Evidence from the united kingdom. Journal of Behavioral and Experimental Finance 26, 100300.

Schwenkler, G., Zheng, H., 2019. The network of firms implied by the news. Boston University Questrom School of Business Research Paper .

Shapiro, A. H., Sudhof, M., Wilson, D. J., 2020. Measuring news sentiment. Journal of Econometrics .

Stambaugh, R. F., Yu, J., Yuan, Y., 2012. The short of it: Investor sentiment and anomalies. Journal of Financial Economics 104, 288–302.

Stambaugh, R. F., Yu, J., Yuan, Y., 2014. The long of it: Odds that investor sentiment spuriously predicts anomaly returns. Journal of Financial Economics 114, 613–619.

Stone, P. J., Dunphy, D. C., Smith, M. S., 1966. The general inquirer: A computer approach to content analysis. .

Tao, R., Yim, A., Han, T., 2020. News-based links and cross-firm return predictability. Available at SSRN 3687222 .

Tetlock, P. C., 2007. Giving content to investor sentiment: The role of media in the stock market. The Journal of finance 62, 1139–1168.

Yu, J., Yuan, Y., 2011. Investor sentiment and the mean–variance relation. Journal of Financial Economics 100, 367–381.

Zhu, L., Wu, H., Wells, M. T., 2023. News-based sparse machine learning models for adaptive asset pricing. Data Science in Science 2, 2187895.