When Sentiment Is News:
Topic-Adaptive Syntax Approach (TASA)

Nazanin Babolmorad[a]

Nadia Massoud

This version: November 2021

First Version: September 2019

Abstract

We present a novel approach for the *sentiment analysis* that combines *topic modelling* with the analysis of the deep structure of text from financial new headlines. The Topic-Adaptive Syntax Approach (TASA) provides two major contributions. First, TASA assesses the sentiment of text using tone-syntax patterns rather than a set of words. Second, the *topic-adaptive sentiment lexicon* constructed by TASA increases the accuracy of sentiment analysis across fine-grained topics in finance. We integrate different machine learning algorithms with our approach to examine whether and to what extent machine learning improves sentiment models in finance. In our empirical analyses, the sentiment of news assigned by TASA better explains market trading activities than existing sentiment models, ranging from bag of word models to fully automatic models. More importantly, our results reveal that fully automatic models underperform compared to models that combine topic-adaptive sentiment lexicon and human knowledge. The results from TASA resolves existing anomalies in the empirical evidence.

Keywords: Sentiment analysis, topic modelling, topic-adaptive sentiment lexicons, financial media news, machine learning algorithms, semantic-syntactic format.

JEL Classifications: G1, G12, G4, G41

Link to TASA tone assignment dataset: https://www.whensentimentisnews.com

---

[a] Corresponding author. Nadia Massoud is a professor of finance at Melbourne Business School, University of Melbourne, Australia, e-mail: N.Massoud@mbs.edu. Nazanin Babolmorad is a PhD candidate at Melbourne Business School, University of Melbourne, Australia, e-mail: n.babolmorad@student.unimelb.edu.au. We acknowledge financial support from Melbourne Business School internal research grants. We would like to thank Adam Shapiro, Ali Akyol, Peter Bossaerts, James Brugler, Oleg Chuprinin, Diego Garcia, Mark Kamstra, Patrick J. Kelly, Srikumar Krishnamoorthy, Pekka Malo, Ronald Masulis, Debarshi Nandy, Mario Schabus, and Robert Wood for their valuable feedback. We also thank our research assistants at the University of Melbourne and Amirkabir University, specially linguist expert, Rachel Soderstrom. This paper benefited from helpful comments from workshop participants at Schulich school of Business (York University), the Melbourne Business School, University of Melbourne, University of New South wales, invited Keynote speeches at the 3rd Sydney Banking and Financial Stability Conference, university of Sydney and the 9th Behavioral Finance and Capital Markets Conference, La Trobe University.

1. Introduction

Investment decision making is influenced by a number of information sources including the financial news published by major media outlet and on online. Studies have shown that sentiments expressed in financial news influence stock prices (e.g. Garcia, 2013; Agarwal, Chen, and Zhang, 2016) and, therefore, are relevant for both investment decision making and companies' risk perception. As the volume and speed of financial news reporting has increased, analysis of the content and tone has exceeded unaided human processing capabilities and machines are now deployed to augment human analysis of text, particularly the tone[2] expressed in reports. *Sentiment Analysis* has emerged as a subfield of *Natural Language Processing* (NLP) in which different combinations of NLP *tools* are used to extract the sentiments. These include polarity dictionaries, Machine Learning (ML) algorithms, Part of Speech (PoS) Taggers, etc. Regardless of the tools used, extracting an accurate and comprehensive representation of text to machines has proven to be a challenge in sentiment analysis particularly in the trade-off between the degree of *complexity* and *dimension reduction*. Degree of complexity is reflected in the complex grammatical structures and the rich interactions among words, while dimension reduction refers to the set of restrictions used to reduce the dimensionality of text to a manageable level. Dimension reduction enhances processing efficiency of the algorithms but often comes at the cost of inaccuracy in categorizing text. To convert the text to a machine format, researchers propose two main solutions.

The simplest solution is the Bag of Words (BoW) format in which text is presented to machines as a set of words or noun phrases regardless of word order or position. Many researchers use the combination of BoW format with either polarity dictionaries (e.g. Tetlock, 2007; Garcia, 2013; Agarwal et al., 2016; Shapiro, Sudhof, and Wilson, 2020) or supervised ML (e.g. Schumaker, Zhang, Huang, and Chen, 2012; Manela and Moreira, 2017; Garcia, Hu, and Rohrer, 2020) to study the sentiment of financial news articles. A concern with this format is that it ignores the deep structure of text which may lead to misclassification of the tone.[3] Additionally, using the BoW format does not allow for the

[2] Tone, sentiment, and polarity are used interchangeably in this paper.
[3] For example, Malo, Sinha, Korhonen, Wallenius, and Takala (2014), Salas-Zárate, Valencia-García, Ruiz-Martínez, and Colomo-Palacios (2017), and Shapiro et al. 2020.

possibility of different tone assignments for multiple firms mentioned in the same text.[4]

A more advanced solution is the Semantic and Syntactic format in which grammar, word order and word position are included as features of the text converted to machine readable format. The field of sentiment analysis in computer science has made tremendous progress in this form of text conversion. However, researchers in finance and accounting have not incorporated this treatment of text complexity regardless of the applied tools when analysing financial texts (e.g. Malo et al., 2014; Meyer, Bikdash, and Dai 2017; Krishnamoorthy, 2018; Shapiro et al., 2020).,

A new emerging technique in computer science combines topic modelling and sentiment analysis (e.g. Deng, Jing, Yu, Sun, and Ng, 2019; Garcia-Pablosa, Cuadrosa, and Rigaub, 2018; Ali et al., 2019). The output from this technique is *Topic-Adaptive Sentiment Lexicons*, which has a promising prospect in sentiment analysis of financial news articles, as it adjusts the tone assignment for words depending on the context in which they are mentioned. That is, it captures the granularity of topics.[5]

These advances in computer science research and the limitation of BoW formats have motivated us to develop a new sentiment analysis approach that combines the semantic and syntactic format and the topic-adaptive sentiment lexicon for the first time in the analysis of financial and business news. We integrate our approach with machine learning algorithms and techniques to examine whether and to what extent machine learning improve sentiment analysis in finance. We examine how abnormal returns can be used for decreasing or eliminating human involvement in building financial sentiment models by developing a wide range of sentiment models from a crafted rule sets model (e.g., Meyer et al, 2017 and Krishnamoorthy, 2018) to fully an automatic model (e.g., Bybee, Kelly, Manela, and Xiu, 2020 and Garcia et al. 2020). The intuition behind automatic sentiment models is that machines can learn sentiment analysis from the market's trading reactions to financial news, which is assessed using abnormal returns and volume.

In this paper we focus on news headlines because they highlight the key features of news article.

---

[4] This is called mixed-tone noise in NLP field of where a text conveys different tones for different name entities (e.g. firms) tagged within the text.

[5] For example, 'sell' conveys negative sentiment in financial market related news, while it has positive polarity in text about selling products in business related news; 'close' conveys neutral in in financial market related news about closing of the market, while it has negative sentiment in text about closing manufacturing plants in business related news.

Headlines are articulated carefully to attract readers' attention and to communicate the main message of the story. The news headline is described by news providers (e.g. The New Yorker, 2014) as the most important part of a news article because it affects the way people read and remember an article; and sets the tone for what follows. Therefore, news headlines summarize both sentiment and content into few phrases, which is a useful approach for representing text to the machine.

The schematic in Figure I presents the conceptual structure of our approach for classifying the tone of business and financial media news at the firm level. The schematic starts from the input of news headlines and shows the steps of the processing system leading to the output. On the right side of the schematic is the sentiment score (s) at the firm level for different topics (business versus finance). As can be seen in Figure I, the structure of our approach consists of seven components. The different components serve important functions, and each is designed to work in tandem with other components to assign a sentiment to news headlines at the firm level. Some of these components are based on a crafted rule set.

Components one and **four** employ the *NLTK toolkit*.[6] **Component two**, *Company/Sector dictionary* is designed for *name-entities recognition* within the financial news text. **Component three**, Topic Lexicon, is constructed by topic modelling including *Cluster Analysis* and Unsupervised ML (i.e. *Latent Dirichlet Allocation* /LDA) for higher-level sentiment classification. This Lexicon classifies phrases into one of three topics including *business*, *finance*, and *mixed* using 3,027 features for *topics recognition*. Finance News is defined as news that directly relates to fundamental or technical analysis of stocks and financial markets. Finance news covers current or forecasted future trends in stock prices (e.g. news about earning reports, dividend, stock prices, etc). Business News describes business details, such as firm operations and management, that go beyond the terminology of trading in capital markets (e.g. launching a new product, marge-and-acquisition, macro-economic news, etc). Such judgements probably are more the province of professional investors than retail investors.

Component five, Topic-Adaptive Sentiment Lexicon, determines the topic-adapted tones and

[6] The Natural Language Toolkit, or more commonly NLTK, is a suite of libraries and programs for symbolic and statistical natural language processing for English written in the Python programming language (Wikipedia).

semantic orientations[7]. This component is unique. For the first time in the literature we combine topic adapted tones and syntax roles for verbs, adjective, adverbs, nouns, noun phrases, and negations to extract *tone-syntax patterns* within the text. For example, the tone-syntax pattern of “Apple May Soon Lose the Title of "Most Valuable Public Company" to This $2 Trillion IPO”, from MotleyFool, 18 Mar. 2017, is [Verb-Negative, NounPhrase-Positive], where ‘lose’ is considered as *negative verb* and “most valuable public company” is a positive noun phrase (i.e. the other *Parts of Speech* are neutral and do not change the tone in this example).[8] This dictionary includes the 2,562 most impactful words and phrases in the financial news headlines.

Component six, the *Tone-Syntax Pattern Lexicon* determines the tone of each tone-syntax pattern (e.g. the tone of [Verb-Negative, NounPhrase-Positive] pattern is negative). Having a comprehensive *topic-adaptive sentiment lexicon* enables extraction of the most impactful 5,432 tone-syntax patterns in financial and business news headlines to link each of them to a specific tone. This lexicon is unique because, instead of linking the tone of text to set of words and phrases, which is a common approach in the literature, it links the tone to syntax patterns. We name our approach the *Topic-Adaptive Syntax Approach (TASA)* as it uses tone-syntax patterns to measure the sentiment of text within the fine-grained topics at the firm level. **Component seven**, Main Core or *TASA*, is a rules-based core, which connects all six components together and leads the system to assign a unique topic-adapted tone to each firm mentioned in a news headline.

TASA is trained on pure text including approximately 111,000 unique news headlines, tagged for S&P 500 firms from January 2014 to Dec 2018 using both ML algorithms and manual labelling. TASA’s design raises three important questions. First, what is the contribution of each component of TASA to the accuracy of its output? For example, does topic-adaptive sentiment matter more to TASA’s output than the semantic and syntactic representation of text. Second, does the human involvement in the different components of the TASA model (e.g. tagging of tone in step 6) increase or decrease bias

[7] Semantic orientation means that a word or a phrase does not convey a special polarity to readers by itself, but the final polarity is based on the other parts of the sentence.

[8] Another example of Tone Syntax Patterns: “Tesla Inc (TSLA) Stock Is Flying Too Close to the Sun” from InvestorPlace.com, Jun. 2020. The tone syntax pattern of this headline is [Noun-Positive-Up, Verb-Positive]. Where, ‘stocks’ stands for ‘Noun-Positive-Up’, as a semantic orientation for stock prices, and ‘flying’ is a ‘Verb-positive’ in finance news, the other parts of speech are neutral. This tone syntax pattern conveys a positive tone.

in the final output compared to alternative approaches, e.g. a fully automated approach? Third, how does TASA perform in explaining trading activities (i.e., AR and AV) compared with alternative approaches. To address these questions, we integrate different machine learning algorithms and techniques by a stepwise replacement of components in TASA to establish whether and to what extent machine learning can improve sentiment analysis in finance. We propose one automatic machine learning model and two hybrid models. The automatic model relies on machine learning techniques to learn sentiment analysis from abnormal returns without any human involvement. The two hybrid models combine a subset of the TASA's components with automatic machine learning algorithms. The hybrid models tune their performance through learning from the reactions of the market to financial news in terms of abnormal returns.

To test the accuracy of the final output of TASA, i.e. topic-adapted tone assignment at the firm level, we conduct two sets of tests using an out-of-sample dataset[9]: "verification" and "validation". The verification tests use classification metrics, a data science method, to compare the accuracy of TASA topic-adapted tone assignment versus actual tone assignment (e.g. Malo et al., 2014; Krishnamoorthy, 2018) as reflected by the market abnormal returns. For a comparison, we also compare the tone assignments of the two Hybrid models, the Automatic Model, and two common sentiment models versus with true tone assignment. The first common model incorporates the BoW representation model and the polarity dictionary of Loughran and McDonald (L&M), which we refer to as L&M BoW. This model simply counts the number of positive and negative words to measure sentiment of a news headline. L&M BoW is commonly used to analyse media content in extant finance studies (e.g., Engelberg, et al., 2012; Ferguson, et al., 2015, Garcia, 2013, Liu and McConnell, 2013). The second common model is the VADER Sentiment model (Valence Aware Dictionary and Sentiment Reasoner). The VADER Sentiment is a lexicon and rule-based sentiment analysis tool. It is publicly available in the NLTK package and can be applied directly to unlabelled textual data.

The results show that the accuracy of tone assignment by TASA is relatively high, and it

[9] The output from the TASA analyses of news headlines published by 15 news providers for firms in the S&P 500 from 2014 to 2018 is posted for academic researcher on this website: https://www.whensentimentisnews.com/

dominates the accuracy of all other sentiment models across all evaluation metrics. The improvements in the accuracy of sentiment output by TASA compared with BoW L&M is 31%, VADER Sentiment is 24%, Automatic Model is 11%, Hybrid Model 1 is 8%, and Hybrid Model 2 is %5.

Our results also demonstrate that the TASA's Topic-Adaptive Sentiment Lexicon is a key component in improving the performance of the sentiment models. Additionally, combining the Topic-Adaptive Sentiment Lexicon with word order and syntax and a machine learning algorithm (i.e. Hybrid model 2) significantly improves the performance of the model but it does not outperform the manual labelling model proposed by TASA.

To conduct the validation tests, we examine investor reactions to the publication of news articles. If tone of news headlines matters to investors, our measure of negative (positive or neutral) news headline should, on average, generate a negative (positive or neutral) market response immediately following publication of the news headlines. To validate this hypothesis, we consider the impacts of firm level tone assignments on the risk adjusted abnormal returns (AR) and abnormal trading volume (AV) using the Fama-French three-factor (1993) and the Carhart four-factor (1997) models. In baseline tests of market reactions to tone, we use the Close-to-Close window of daily news headlines published between the market close on one day and the market close on the next day and the stock returns and volume for the same window. Using the TASA, our analyses reveal associations between the tone of news and market responses that are consistent with expectations but for one set of findings, vary from previously reported results. TASA positive tone assignment is positively associated with daily abnormal returns and trading volumes, especially for finance news. In contrast, many prior studies report a non-significant or, counter-intuitively, negative relationship between positive news and stock returns, which we attribute to measurement error in previous approaches (e.g. Tetlock, 2007; Engelberg et al., 2012). As in previous research, our measure of negative headlines is negatively associated with daily AR, positively associated with daily trading, and has stronger associations for finance news than business news. The results are robust for intraday and close-to-open windows used in robustness tests (i.e. The close-to-open window captures the impact of accumulated news overnight (lead) on trading activities (lag) at the beginning of the trading session).

We conduct further validation tests for the close-to-close window by comparing TASA with Hybrid model 2, Automatic Model, L&M BoW and VADER Sentiment model. A sentiment analysis method is assumed to be superior if it is less likely to misclassify words and sentences while considering other exogenous variable, i.e. the tone assignments are more strongly associated, economically and statistically, with patterns observed in stock market returns post publishing news headlines. We call this hypothesis the Superiority Hypothesis. To examine the Superiority Hypothesis, we compare the power of the sentiment models in interpreting contemporaneous AR in both separate and joint cross-sectional regressions. We control for year fixed effects, firm fixed effects and annual market capitalization fixed effects using quartiles. The results show that the coefficients on both negative and positive tones for the TASA are statistically and economically larger than for other sentiment models. Consistent with verification test results, models that do not include semantic, syntax, word order and no machine learning to tune the model underperform all other models. In addition, the Hybrid models outperform the Automatic model, showing that integration of machine learning and human knowledge improves sentiment models.

The remainder of the paper is organized as follows. Section 2 highlights our contribution to the literature. Section 3 discusses our methodology for development of TASA. Section 4 discusses the machine learning models in detail. Section 5 introduces the verification tests and results. Section 6 reports validation results and robustness tests. Finally, Section 7 presents our conclusions.

2. Contributions to the Literature

Sentiment analysis of financial text typically employs a combination of the BoW format and polarity dictionaries (e.g. Harvard psychosocial dictionary and Loughran and McDonald (L&M) dictionary) and simply counts the number of positive and negative words in text to assign the tone to the text (e.g. Tetlock, 2007; Loughran and McDonald, 2011; Garcia, 2013; Jegadeesh and Wu, 2013; Chen, Demers, and Lev, 2018). This approach is simple but imprecise, as the measure of tone is based solely on the words defined in the polarity dictionaries without any consideration of the semantic content of the words (Malo et al., 2014; Meyer et al., 2017; Manela and Moreira, 2017; Krishnamoorthy, 2018; Garcia et al., 2020). TASA assesses the sentiment of text using tone-syntax patterns rather than the simple set of words used in BoW format. TASA also provides a topic-adaptive sentiment lexicon for the first time in the literature instead of using polarity dictionaries. As a result, TASA significantly improves the accuracy of sentiment assignment at to textual phrases and to firms.

TASA can also resolve the existing limitations in applying application of supervised ML algorithms to sentiment analysis of media news. Das and Chen (2007), Schumaker et al. (2012), and Manela and Moreira (2017), Bybee, Kelly, Manela, and Xiu (2020), and Garcia et al. (2020), examine the tone of financial news by applying the combination of the BoW format and supervised ML algorithms. In these studies, the algorithms are trained using 'lagged returns' or 'the implied volatility' for a specific *time-window* as an exogenous variable. This produces the training data set, which includes a set of financial news linked to the labels of tones. This approach provides significant improvements in sentiment analysis of financial text, however, as well recognised in the literature, the approach is limited by the use of the BoW format for the reasons mentioned above. The accuracy of the supervised ML approach can be enhanced by replacing the BoW format with the tone-syntax patterns proposed by TASA. Specifically, the TASA and Hybrid model 2 outperform both Hybrid model 1 and Automatic model. All models use machine learning and there is only one significant difference between these models and that is the representation format of text. The TASA and Hybrid model 2 use the tone-syntax patterns, while both Hybrid model 1 and Automatic model apply the BoW format.[10]

Recently, as a response to the limitation of the integration of BoW model and ML, researchers (e.g. Malo et al. 2014; Meyer et al., 2017; Chan and Chong, 2017; Krishnamoorthy, 2018) proposed supervised ML algorithms trained using manually tagged data. This approach goes beyond the polarity of words to include grammar, word order and semantic orientation using a collection of phrase banks for financial news articles. These studies offer promise for improving the precision of estimation of the tone of financial news, but only when the algorithm is trained on large training datasets. If datasets are

[10] Another limitation of this supervised ML algorithm is that it seems no consensus is established around the appropriate target time-window. For example, Research considers the target time window from 20 minutes after news publishment to close-to-close stock market time within the time of releasing of news as an optimal choice (e.g. Schumaker et al., 2012; Manela and Moreira, 2017). Because the act of correlating financial news articles to a stock price movement is a challenging task, even around major corporate events, e.g. earning announcements and M&A (Gidofalvi, 2001; Garcia, Hu, and Rohrer, 2020). This Tagging all the financial news articles in the target time-window with the same tone decrease the accuracy of the approach by misclassify the tone of text in the training data. Because ML algorithms cannot learn more than what is provided in the training data.

not sufficiently large, the machine learning algorithm leads to overfitting.[11] The maximum labelled dataset used in studies reported to date (e.g. Malo et al., 2014; Meyer et al., 2017) contain around 5,000 phrases randomly selected from news articles. While this labelling is a large manual task, the resulting sample is small compared to datasets used to train algorithms in other areas of research, including movies and products reviews, and discussion forums (Meyer et al., 2017). To address this issue in the literature, we develop two sentiment models that integrate ML and large training datasets using the semantic and syntactic representation i.e., the TASA and Hybrid model 2. The TASA and Hybrid model 2 are trained based on large training sets including about 111,000 and 178,000 unique news headlines, respectively. The difference between these two models is the training sets. The TASA training set is built on more manually tagging than automated process, while the Hybrid model 2's training set is almost based on abnormal returns. The TASA accuracy is significantly higher that the Hybrid model 2 as shown in the verification and validation tests.

3. Methodology

Confidence in findings from sentiment analyses of financial news has been limited due to measurement errors in the assignment of tone (e.g. Kearney and Liu 2014; Malo et al., 2014; Garcia et al., 2020). We address this limitation of past methods by developing a comprehensive approach to assess the tone within the deep structure of text. Our strategy is to let the data empirically determine the impactful words, phrases, and semantic-syntactic structures in the text. Our approach contrasts to that of existing sentiment analyses that have been built using the BoW format coupled with polarity dictionaries or ML algorithms. It is designed to process the tone within syntax patterns at the firm level. This results in high precision sentiment analyses of financial news, even when multiple firms are mentioned in the same text. For the first time in the literature, we provide an approach that combines the two sub-fields of NLP, topic modelling and sentiment analysis, to construct a *topic-adaptive sentiment lexicon* to determine the tone of text across fine-grained topics. We call our novel approach the Topic-Adaptive Syntax Approach (TASA). This section describes how we build the different

[11] Building a large training dataset is time consuming and costly.

components of TASA and the role of each component in the ecosystem.

The schematic in Figure I presents the conceptual model of our approach for classifying the tone of business and financial media news at the firm level. The schematic starts from the input of news headlines to the processing system and finishes with the assigned tone to a firm. On the right side of the schematic is the sentiment score (s) at the firm level and its topic (business versus finance). As can be seen in Figure I, the structure of our approach consists of seven components: 1) Headline Splitter, 2) Company/Sector Dictionary, 3) Topic Lexicon, 4) Part of Speech Tagger, 5) Topic-Adaptive Sentiment Lexicon, 6) Tone-Syntax Pattern Lexicon, and 7) TASA, the main Core which completes the classification of headline tone at the firm level. Each component works in tandem with other components to assign a sentiment to news headline at the firm level. The remainder of this section describes the purpose of each component, how it is designed and built, and how it works by using the example in Panel B of Figure I. We first describe the textual dataset, which is applied to train and test our approach.

3.1 Dataset

To build and verify our approach, we collect 236,790 news headlines from 15 leading news providers for S&P 500 firms from January 2014 to Dec 2018.[12] The news providers include Seeking Alpha, Zacks, WSJ, Market Realist, Motley Fool, Yahoo Finance, Reuters, Bloomberg, InvestorPlace, Investor's Business Daily, GuruFocus, 247WallSt, Barron's, Fox Business, and Benzinga. The 236,790 news headlines are randomly spilt into three subsamples. The first subsample, the "in-sample" contains 110,000 (47%) news headlines which we use to train our approach. The second subsample, "the verification sample" contains 1,520 (0.5%) news headlines and is used for testing the accuracy of TASA compared to sentiment analysis by human (i.e. verification phase). The third subsample, the "out-of-sample" contains 124,637 (52.5%) news headlines which is used to examine the performance of TASA in explaining abnormal market trading activities (i.e. AR and AV) compared to commonly used

[12] We consider all firms, which were or are in S&P 500 index constituents from 2000 to 2018.

approaches in the literature (i.e. validation phase).

3.2 Headline Splitter (Component one)

The main objective of this study is to capture the tone of text at the firm level, even when a text conveys different tones for multiple firms (i.e. mixed-tone text). This requires assignment of tone at a more granular level than the BoW format. For example, the information in this headline: “Oakbrook Investments Llc Buys Apple Inc, Starbucks Corp, Boeing Co, Sells General Electric Co, ...”, from GuruFocus.com on 18th of January 2018, conveys to investors a positive tone about ‘Apple Inc, Starbucks Corp, and Boeing Co’ based on the verb ‘*buy*’, and a negative tone about ‘General Electric Co’ based on the verb ‘*sell*’. Our tone assignment contrast with that in previous studies which assigns the tone at the BoW format and applies it equally to all firms mentioned in text.

To capture the sentiment of text at this granular firm level, the Headline Splitter breaks the text into phrases based on punctuation marks using the NLTK Sentence Tokenizer. The use of punctuation marks in news headlines is extensive. The headline above, from GuruFocus.com on 18th of January 2018 demonstrates how the use of a comma (,) as a phrase splitter plays an important role by separating the two verbs in the sentence, ‘*buys*’ and ‘*sells*’, so that each can convey a different tone about the firms they refer to. Inspecting our in-sample news headlines, we propose the following splitters: comma ', ', semicolon ';', dot '. ', question mark '?', dash '-', 'despite', 'while', etc. For example, using the Headline Splitter rules to pre-process the Guru Focus’ headline, identifies the comma and the dot as the punctuations and generates 5 phrases.

In Panel B, Figure I, we show an example of how our conceptual model analyses the headline “Alcoa To Close Two Spanish Aluminium Plants, Cut Jobs.” from Reuters, 18 Oct. 2018. Using the punctuation rules, the main punctuations in this headline are the comma and the dot, and the headline is separated in two phrases by the comma. Our approach utilizes the punctuations to improve the accuracy of the tone assignment at the phrase level rather than the BoW format.

3.3 Company-Sector Dictionary (Component two)

This dictionary is applied for *name-entity recognition* in text of where firms are considered as name entities. We consider different variations of company names including full name, ticker, abbreviations, and nick names of firms traded on three exchanges NASDAQ, NYSE, and AMEX in the Company-Sector dictionary. The company sector is taken from the NASDAQ. Based on our full sample of 236,790 news headlines collected for S&P 500 firms from 2014 to 2018, we find financial news for 221 firms belonging to 13 sectors. Table I shows the summary statistics of our list of firms based on sector and exchange. The example in Panel B of Figure I shows that the Company/Sector Dictionary assigns Alcoa to the first phrase, which is in the *basic industries sector*. The dictionary does not assign a company/sector label to the second phrase as it does not contain any relevant keywords.

3.4 Topic Lexicon (Component three)

In this paper, we apply topic modelling to construct a topic-adaptive sentiment approach for the tone classification across fine-grained topics in finance including finance and business. It is widely recognized that sentiment varies even in different topics within a domain. To address this issue, computer science studies apply *topic modelling* (e.g. Latent Dirichlet allocation, LDA, Latent Semantic Indexing, LSI, and Principle Component Analysis, PCA) to build topic-adaptive sentiment lexicons and topic-adaptive sentiment classifiers (e.g. Deng, Jing, Yu, Sun, and Ng, 2019; Garcia-Pablosa, Cuadrosa, and Rigaub, 2018; Ali et al., 2019).

However, finance studies ignore the variability of the sentiment polarities in different topics in finance. For example, 'sold', a variant of 'sell', conveys a negative tone in the news headline "Dr Pepper Snapple: 7 Different Insiders have Sold Shares During the Past 30 Days", from Seeking Alpha, 5 Mar. 2015. Because investors in the capital market are more likely to read this 'finance news' as negative given the belief that those insiders might be better informed about the firm's future cashflow. While 'sell' has a positive tone in the news headline "Boeing May Sell Chinook Helicopters to Brazil", from Investors.com, 16 Oct. 2014. Investors may perceive this 'business news' as positive because it predicts or confirms future revenue for Boeing. As you can see, the word 'sell' or variants communicate different tones in the two examples of financial news headlines. That is a negative tone for the stock market news headline in Seeking Alpha, and a positive tone for the business news headline in Investors.com. There are many words that can have different tones based on the context of financial news including 'sell', ‘buy’, ‘close’, ‘open’, ‘start’, ‘win’, ‘defeat’, ‘expensive’, etc. To the best of our knowledge, there is no sentiment analysis approach in the finance literature that is capable to differ sentiment across fined-grained topics in finance.

To address this issue, we construct a Topic Lexicon to classify financial news into fine-grained topics and we then use classified text to build a topic-adaptive sentiment lexicon. which includes the most impactful features across topics in financial news. The Topic Lexicon helps TASA to recognize the fine-grained topics in financial news to implement the tone assignment considering the semantic within the context. To build this lexicon, we employ the *Cluster Analysis* and an Unsupervised ML algorithm, i.e. Latent Dirichlet Allocation (LDA). LDA is a widely used unsupervised modelling technique for identifying topics that best describe a set of documents (e.g. Bao and Datta, 2014; Huang, Lehavy, Zang, and Zheng, 2018). LDA considers an assigned number of topics and allocates words and phrases to topics based on similarity but does not label the topics. We run the LDA with a fixed set of 50 topics. We also construct clusters across firms (i.e. 221 clusters) and sectors (i.e. 13 clusters). We then extract the most impactful words and phrases (i.e. features) in terms of their appearance across clusters and topics using a weighting factor. The weighting factor is based on Term Frequency Inverse Document Frequency (TFIDF)[13] in which cluster/topic is considered as document. A feature receives higher weighting if it is a high frequency word/phrase in a special cluster/topic and is not repeated at high frequency in the other clusters/topics. The 20% most highly weighted features in each cluster/topic (i.e. we have 284 clusters/topics in total) are selected. Example of features extracted for topic recognition using these models include crude oil prices, iPhone, FDA, aircraft market, animal health, artificial intelligence, cable providers, chicken, consumer healthcare, defence contracts, etc.

Subsequently, our annotators evaluate the extracted features across the 282 clusters and combine them within two broad meta-clusters (i.e. finance and business). Our final Topic Lexicon includes 3,027 features comprising 1,744 business features and 1,283 finance features. The 282 clusters could be combined into more than the two meta topics. For illustration purposes, we consider just the

[13] TFIDF is a numerical statistic that reflects how important a word is to a document in a collection of textual documents.

two most common meta topics of business and finance for this paper. For example, business features can capture possible products at the sector level for the 13 sectors. Business features also capture business activities (e.g., mergers & acquisitions (M&A), changing of board members, contracts, supply and demand, etc.) and business ramifications (e.g., natural disasters, air crashes, trading wars, legal problems, etc.). Finance features flag news headline that directly discuss either fundamental factors (e.g., earnings, profit, dividend, etc.) or technical factors (e.g., stock prices, market volatilities, bear market, beat earnings, etc.).

TASA is trained to apply the Topic Lexicon for classifying each financial news headline into one of four categories: i.) Finance only features (e.g., *"Why Caterpillar, Goldman Sachs, and American Express Led the Dow to Record Highs Last Week",* The Motley Fool, 7 Jun. 2014), ii.) Business only features (e.g. *"There's a Boom in Airline Travel, and Jet Makers Are Making the Most of It",* from The Wall Street Journal, 12 Feb. 2018), iii.) Mixed, those with business and finance features (e.g., *"FDA Oks Merck's Keytruda for Two New Uses in Bladder Cancer; Shares Ahead 1% After Hours- Seeking Alpha",* 18 May 2017), and iv.) Unlabelled, which are those news headlines that have neither feature. This extensive Topic Lexicon is unique in terms of size and application in the finance literature of sentiment analysis.

For the example shown in Panel B of Figure I, the news headline is categorized as a business topic. This is because in the first phrase of news headline news "Plants" and in the second phrase "Jobs" both belong to the business topic.

3.5 NLTK POS Tagger (Component four)

We use NLTK POS tagger to determine parts of speech such as nouns, adjectives, verbs, etc. In Panel B of Figure I, we show how NLTK breaks down and tags words and phrases for an example headline. For the first phrase "Alcoa To Close Two Spanish Aluminium Plants": the *verb* is '*close*', the *nouns* include '*Alcoa*', '*aluminium*', and '*plants*'. '*Spanish*' is an *adjective* and there is a *noun-phrase*: '*Spanish aluminium plants*'. There are no other phrases or negations. For the second phrase *'Cut Jobs'*: the verb is *'cut'*, the noun is '*jobs'*. There is no adjectives, noun-phrases, other phrases or negations. NLTK POS tagger sometimes label POS incorrectly. For example in this headline *"Buying Oil on The*

Dip? Stick to Buying Quality Oil Companies That Will Profit from A Rise in Oil Prices", from Seeking Alpha, 31 Dec. 2014, the word *'stick'* is a verb while NTLK tags *'stick'* as a noun. Our Topic-Adaptive Sentiment Lexicon fixes such labelling errors, as shown in the next step.

3.6 Topic-Adaptive Sentiment Lexicon (Component five)

Many sentiment dictionaries have been proposed. These can be classified as domain-independent (e.g. Harvard psychosocial dictionary) or domain-specific (e.g. Loughran and McDonald (L&M) dictionary; Garcia et al., 2020). These dictionaries are constructed either manually, e.g. L&M dictionary, or using ML based on statistical co-occurrence information between the market trading activities and sentiment labels (e.g. Garcia et al., 2020). Existing dictionaries in the finance literature have not identified variation in the meaning of the words within different contexts in a domain. Recent studies in computer science address this issue by providing topic-adaptive sentiment lexicons, which consider the sentiment of words across different fine-grained topics.

In this component, we leverage the topic-adaptive sentiment lexicon approach to consider the sentiment of words and phrases within context using both syntactic and semantic structures across different topics. The output of this lexicon is the tone-syntax patterns. To construct this lexicon, we apply our Topic Lexicon for in-sample data and classify the headlines into the finance and business topics. We then use NLTK POS Tagger to determine Parts of Speech (POS) within text. Two lists are developed for each topic, one using the unigram approach (i.e. it includes four lists, verbs, nouns, adjectives, and adverbs) and the other using phrases (i.e. it includes noun phrases and other phrases). To finalize the Topic-Adaptive Sentiment lexicon, we let the data empirically determine the most impactful words and phrases in headlines at the syntax-topic level.

Unigram list of Topic-Adaptive Sentiment lexicon

In the unigram list, we consider four groups of words from the NLTK POS Tagger label: verbs, nouns, adjectives, and adverbs. The dictionary also includes a list of common negation words (e.g. *not, no, nobody, never, doesn't, don't*, etc). Figure II shows the classification process for the four groups in the unigram list, based on the tone and semantic orientations in finance and business. In Panel A, verbs are classified into two sub-groups: fixed-sense and directional. The fixed-sense verbs always convey a fixed tone such as positive, negative or neutral. For example, *'acquire'* and *'expand'* convey a positive tone, *'fall'* and *'hurt'* convey a negative tone, while *'announce'* and 'compare' convey a neutral tone. Verbs in the directional sense group work as modifiers and do not convey sentiment but show an increase or decrease (i.e. *'Up'* and *'down'*, respectively) in the size, amount, or degree of relevant nouns and noun phrases. For example, synonyms for the verb up include boost, climb, increase, jump, pick up, soar, raises, etc, while synonyms for the verb down include cut, drop, slide, lowered, etc.

In Panel B of Figure II, the *nouns* are classified into two sub-groups, fixed sense and biased sense (i.e. semantic orientations). The fixed-sense noun categories are very similar to the fixed-sense verbs in that they convey a tone and are not related to other words or phrases in a sentence. For example, *'growth'* and *'developments'* convey a positive tone, *'Bubble'* and *'collapse'* convey a negative tone, and *'product'* and *'contract'* convey a neutral tone. Biased sense nouns display polarities based on the orientation of a modifier or a verb. For example, *'employment'* and *'price'* convey a positive tone if they are followed by a modifier indicating an increase (e.g. up, boost), while tax and risk convey a positive tone if they are followed by a modifier indicating a decrease (e.g. drop, down). Similarly, in Panels C and D of figure II, we show how we classify the adjectives and adverbs in our unigram dictionary based on the tone.

Phrases list of Topic-Adaptive Sentiment lexicon

Noun phrases play an important role in finance and business headlines such that considering words only may lead to misclassification of headline sentiments. For example, in the headline *"How Low Oil Prices Impacted Texas Business Activity in January",* from Market Realist, 26th January 2016, '*low oil prices*' is a noun phrase[14] in which none of words convey a fixed sense while the semantic structure of the phrase conveys a negative sentiment. The classification of noun-phrases is similar to that of nouns. We report the discussion of noun phrases sentiment categories, noun phrases grammatical structure and other phrases in Internet Appendix IA.A.

Identifying the most impactful words and phrases

To build the different components of the Topic-Adaptive Sentiment lexicon explained above

[14] 'low oil prices' is a noun phrase which consists of 'low' as an adjective, 'oil' as a singular noun, and 'prices' as a plural noun.

we let the data empirically identify the most impactful words and phrases in finance and business news headlines. To achieve this, we follow a six-step procedure. (1) We extract the unigrams and phrases in the in-sample dataset and create a list of 7 different groups considering POS tags (i.e. verbs, nouns, adjectives, adverbs, noun phrases and other phrases). (2) We compute the proportion (reported as a percentage) for the word count (or phrase count) divided by the total number of words in the groups identified in step 1. (3) The words (or phrases) are then listed in descending order of proportions. (4) The cumulative proportion from top to the bottom of the ranked list is the calculated. (5) A target level of 75% is set and only words (or phrases) that fall above the target level (i.e., in the top 3 quartiles) of the accumulated proportion are retained. (6) Our annotator team then tag each of the retained words (phrases) with the sentiment of that word (phrase) for the new headlines across the 18 different polarity groups, as shown in Table II. Each filtered word (phrase) is tagged by at least three annotators before the final sentiment label is chosen, based on a majority vote. As an additional verification step, a random sample of 600 words and phrases from our Topic-Adaptive Sentiment lexicon are validated by a linguistic expert from the University of Melbourne, who endorsed the manual tagging for all 600 words and phrases.

Columns 2 and 3 in Table II contain a summary of our Topic-Adaptive Sentiment Lexicon across the 18 categories. In column 2, there are 635 (25%) words and phrases with a negative tone, 1,222 (48%) words and phrases with a positive tone, and 479 (19%) words in semantic orientation that do not convey a tone without a modifier. There are 226 (9%) directional (i.e. modifier) words. In total, our Topic-Adaptive Sentiment lexicon consists of 2,562 words and phrases. This list does not include the neutral words.

Our approach is not likely to be subject to the endogeneity problem explained in the survey paper of L&M (2020), i.e., if a small set of negative words is identified, managers will avoid them in the future. This is because our Topic-Adaptive sentiment Lexicon includes a large number of negative and positive words. Additionally, in general, reporters (different from insiders to the firm) do not have an incentive to conceal the sentiment of their news report to the contrary they seek to highlight it.

We investigate how the L&M (2011) dictionary classifies the 2,562 words in the Topic-Adaptive Sentiment Lexicon. The results are presented in in the last three columns of Table II. For the

L&M dictionary, 176 (7%) words are classified as negative, 79 (3%) words are classified as positive, 1,009 (40%) words are classified as neutral, and 1298 (51%) are classified missing. The intersection of our Topic-Adaptive Sentiment lexicon with L&M is very low, particularly for the positive word list. The L&M dictionary only classified 6% of our positive word list as positive, 4% as neutral and 90% as missing. In unreported results[15], we obtain similar results when we analyse how the Harvard psychosocial dictionary categorizes the 2,562 words in our Topic-Adaptive Sentiment lexicon.

In the example shown in Panel B of Figure I, we illustrate how our Topic-Adaptive Sentiment Lexicon assigns polarity to words/phrases at the topic-syntax level and generates a Tone-Syntax Pattern for each phrase in the news headline *"Alcoa To Close…"*. The output from the Topic-Adaptive Sentiment Lexicon for each phrase is a novel format, that we call it *tone-syntax pattern*. It is a sequence[16] of POS tags (syntax) linked to either a tone or semantic orientation within a phrase. The tone-syntax pattern keeps word order in the phrase. We use square brackets [.] to refer to a tone-syntax pattern. In the first phrase *"Alcoa To Close Two Spanish Aluminium Plants"*: the verb is '*close*' and its polarity is *verb-negative*. In this phrase all the nouns (i.e. '*Alcoa*', '*aluminium*', and '*plants*') and the adjective (i.e. '*Spanish*') have no polarity and accordingly they are not labelled. The Tone-Syntax Pattern of the first phrase is [verb-negative]. In the second phrase the verb '*Cut*' is classified as *verb-down*, and '*job*' is classified as *noun-pos-up*'. Consequently, the tone-syntax pattern of the second phrase is [verb-down, noun-pos-up].

3.7 Tone-Syntax Patterns Lexicon (Component Six)

Having a comprehensive *topic-adaptive sentiment lexicon*, facilitates extracting all tone-syntax patterns in the in-sample dataset. Subsequently, to construct our Tone-Syntax Patterns Lexicon, each extracted tone-syntax pattern is linked to a unique tone (i.e. positive, negative, neutral).

Specifically, we extract the most impactful 5,432 tone-syntax patterns, for phrases in the in-sample dataset. Our annotator team then review each tone-syntax pattern and assign it a tone (i.e.

[15] Available from corresponding author.
[16] In mathematics, a sequence is an enumerated collection of objects in a particular order, in which repetitions are allowed.

positive, neutral, and negative) by considering the original news. As a result, we have 2,489 patterns with a positive tone, 1,895 with a negative tone and 1,048 with a neutral tone. In Table III we provide some examples of patterns from the Tone-Syntax Patterns Lexicon, plus their frequency and tone. For example, the [verb-negative] phrase, the first phrase of the example in Panel B of Figure I, is repeated in 5,691 phrases in the in-sample data and is considered as negative tone. Another of the examples shows a negation pattern, ['Negation', 'Verb-Positive'], which is repeated in 10,431 phrases and is assigned to negative tone.

3.8 Main Core, TASA (Component Seven)

TASA is more than just being ML algorithms or polarity dictionaries. It is an automated system designed for sentiment analysis of financial news at the firm level. Component Seven, Main Core, integrates the outputs for the other six components, using a unified set of rules to assign a unique tone to each firm mentioned in a news headline at the topic level (see Figure I). These rules are developed to capture the deep structure of news headlines. The Main Core makes TASA capable to sentiment analysis of text, even when text is a mixed tone for multiple firms. TASA uses the following 4-step algorithm to assign the polarity of a phrase to its adjacent phrases and determine the final sentiment for each firm in a news headline:

I. Allocating the sentiment polarity to adjacent phrases from left phrase to right.

II. If there is at least one negative (positive) phrase on a news headline and there is not any positive (negative) phrase, then the overall sentiment is negative (positive) and is assigned to all phrases.

III. If there are both positive and negative phrases in a headline, TASA keeps both those sentiment labels and does not let positive and negative labels neutralize each other. In such cases, the headline has a mixed phrase label. If there are neutral phrases in a mixed headline, TASA assigns the nearest preceding neighbour's non-neutral label to the neutral phrase by giving a higher priority to the nearest left-hand side neighbour.

IV. If all phrases in a headline are either neutral or non-neutral, TASA stops. Otherwise TASA repeats steps I to III.

To show how this algorithm works to assign a tone at the firm level, we consider three scenarios.

Scenario 1, a headline mentions only one company. Scenario 2, a headline mentions more than one company. Scenario 3, the headline does not directly mention any company, but the relevant companies are tagged on the article by the news provider.

In Scenario 1, there are three cases depending on the number of different tones assigned to each phrase. In Case 1, the headline contains only one phrase and TASA assigns the tone of that phrase as the final tone for the firm. In Case 2, the headline contains more than one phrase and the tones of those phrases do not conflict with each other. That is, all phrases have either a mix of positive and neutral or negative and neutral tones, or tones are either all positive or all negative. TASA assigns the dominant tone as the final tone for the firm. In Case 3, the headline contains more than one phrase, and the tones of those phrases contain a mixture of positive and negative tones. In this case, TASA assigns two final tones to the firm, one positive and one negative.

In the example headline shown in Panel B of Figure I, we identify two phrases where the first phrase includes a company name, Alcoa, and the second phrase does not have any company labels. Both phrases are labelled negative by the Tone-Syntax Pattern lexicon. This fits Case 2 above. The headline mentions one firm and contains more than one phrase, but the tones of those phrases do not conflict with each other. Accordingly, TASA assignes a negative tone to Alcoa for this headline. The explanation of scenarios 2 and 3 are presented in Internet Appendix IA.B.

4. Machine Learning Sentiment Models

TASA is a rule-based sentiment analysis model that integrates textual analysis techniques and manually crafted rule sets. There are two main questions regarding the TASA model. First, what is the contribution of each component of TASA in the model performance? For example, how much do Topic-Adaptive Sentiment Lexicon, component 5 in TASA, and Tone-Syntax Patterns Lexicon, component 6 in TASA, contribute to the accuracy of the final output of TASA? Second, does the replacement of human inputs in TASA with automated processes improve the model performance? Specifically, does replacement of all or a subset of the human inputs in TASA with automatic machine learning algorithms improve model performance? In this section, we propose two hybrid models (e.g. Malo et al. 2014; Meyer et al., 2017; Chan and Chong, 2017; Krishnamoorthy, 2018) and one automatic machine learning model (e.g., Bybee et al., 2020; Garcia et al., 2020) to address these two questions.[17] The first Hybrid model incorporates the Support Vector Machine (SVM), BoW representation model and the Topic-Adaptive Sentiment lexicon from TASA. This model includes component 5 from TASA (i.e. semantic and topic-adaptive sentiment) but excludes component 6 of TASA (i.e. Tone-Syntax Lexicon Pattern), as it considers the BoW model. The second Hybrid model is also based on SVM and component 5 from TASA and excludes the component 6 i.e., the Tone-Syntax Lexicon Pattern. But in comparison to Hybrid model 1, it considers the semantic and syntactic representation using word order. Specifically, our second Hybrid model learns the sentiment of syntactic and word order from the market through abnormal returns while TASA applies a rule-based syntactic dictionary. The third Model is an automatic machine learning model that incorporates the SVM and BoW representations. This model ignores semantic, syntactic, and word orders. The model is fully automated and learns sentiment patterns from the market through abnormal returns. All three models have less human intervention than TASA through removal of the dictionary-based components of TASA and the use of abnormal returns to train the SVM instead of manual tagging. We first explain the general training and prediction processes used in the SVM. Each model is then explained in detail.

4.1 The Training and Prediction Processes in the SVM

In the training process, a machine learning algorithm learns how to associate a text input to the related sentiment label (i.e., positive, negative, or neutral) based on the data used for training i.e., the training set. A feature extractor is required to convert an input text into a vector of numbers as machines cannot read textual data. Each input text in the training set is assigned to a predefined sentiment label. Pairs of vectors and labels are fed into the algorithm to build a sentiment model. The feature extractor and training set each have key functions in improving the performance of sentiment analysis models.

In the prediction process, the feature extractor is applied to transform unobserved text inputs, the test set, into numeric vectors. These vectors are then fed into the model to predict the sentiment labels.

[17] Automatic models rely on machine learning techniques to learn from training data. Hybrid models integrate both crafted rule sets and automatic machine learning algorithms.

Our goal is to build a model that generalizes to any new data. Thus, both test and training sets must be large enough to yield statistically meaningful results. Our entire dataset consists of 236,790 news headlines excluding the post period sample. Manual tagging of all news headlines would be costly and time consuming. Even if achievable, manual tagging raises concerns about bias and fitness. Instead, we develop three algorithms which apply abnormal returns to define the true sentiment label for each pair of firm-headline in our dataset.[18] In other words, we let the market judge the sentiment of news articles and show its verdicts in terms of stock price movements. We then randomly split the entire dataset for each iteration into the training and test sets pairs, with the split percentage 75% (i.e. for training) and 25% (test), respectively.[19] The training and test sets include 178,184 and 58,606 news headlines, respectively. We incorporate each model for 10 iterations and consider the average of the results of all iterations as the model performance. The three algorithms for creating true sentiment labels are as follows.

The first algorithm uses Abnormal Returns on day 0 for firm j ($AR_{0,j}$), the day of publication of news headlines, to assign the true sentiment label to each pair of firm-headline. Panel A of Figure IV shows the histogram built by $AR_{0,j}$. The distribution of abnormal returns is very close to the normal distribution. Zero AR is in bin 22 in Panel A. We assume bin 22 (zero AR) and the two surrounding bins (AR -0.348% in bin 21 and AR +0.348% in bin 23) convey neutral sentiment as there is no significant market prices response to the firm-headline pair. This represents 40.9% of the entire data. The firm-headline pairs linked to $AR_{0,j}$ higher than 0.348% are tagged as positive (i.e. 32% of the entire data) and the rest of the pairs are labelled as negative (i.e. 27.1% of the entire data). The second algorithm applies abnormal returns on the preceding day of publishing news headlines for firm j ($AR_{-1,j}$), to assign the true sentiment label to each pair of firm-headline. Panel B of Figure IV shows the histogram built by $AR_{-1,j}$. similar to Panel A, in Panel B, we assign neutral sentiment to the zero AR bin and the two surrounding bins, i.e., abnormal returns in range [-0.2% to 0.2%]. As a result, 36.6% of this sample is labelled as neutral, 31.1% positive, and 32.3% as negative. The third Algorithm uses

[18] Headlines with multiple firms are repeated in the dataset, each time they tagged for a specific firm.

[19] We also repeat the experiments for the split percentage 70%-30%. We observed the same results in terms of ranking the models. However, the accuracy of all models fall compared with the split percentage 75%-25%.

Cumulative Abnormal returns on firm j on trading days -1 to +1 from the day of publication, day 0, CAR_j(-1,+1). Panel C of Figure IV shows the histogram built by CAR_j(-1,+1). The third algorithm also considers CAR in the range [-0.348%, 0.348%] as neutral. As a result, the algorithm tags 37.7% of the entire data as neutral, 29.3% as positive, and 33% as negative. Finally, we apply the majority votes algorithm to determine the final true sentiment label for each pair of firm-headline using the tone assignment from the three algorithms. The distribution of tone for headline-firm pairs using the majority vote is 38.2% neutral, 30% positive, and 31.8% negative. This mechanism of determining true sentiment labels allows us to avoid imbalanced classification problems. Imbalanced classifications cause a challenge for the prediction process as most of the machine learning algorithms used for classification are built based on the assumption of an equal number of samples for each class.

4.2 Hybrid Model 1 (Topic-Adaptive Sentiment Lexicon & BoW & SVM)

Hybrid model 1 includes the TASA's Topic-Adaptive Sentiment Lexicon (component 5), BoW representation model, and the SVM. This model uses a feature extractor that integrates both the TASA's Topic-Adaptive Sentiment Lexicon and BoW model to transform a text into a vector of numbers. First, an input text is converted into a Tone-Syntax Pattern using the TASA's Topic-Adaptive Sentiment Lexicon. The pattern is then transformed into a BoW representation model, which is readable by the SVM.

The BoW model uses a vocabulary of all labels defined in the TASA's Topic-Adaptive Sentiment Lexicon, that is the 18 tokens presented in Table II. For example, "None-Positive", "Verb-Negative", "Noun-Positive-Up", etc. The vocabulary is used as a reference by the feature extractor to make a fixed-length vector of 18 for any pattern. The vector includes one position to score each token listed in the vocabulary. A token is scored in terms of the number of times it appears within the pattern. For example, the headline "3 Reasons to Avoid Apple stock, Motley Fool, 2018-April-10" is converted to "Verb-Negative Noun-Positive-Up" using the TASA polarity dictionary.[20] The TASA pattern is then

[20] - Verb-Negative stands for "avoid" and Noun-Positive-Up is the TASA token for "stock".

transformed into the feature vector [0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0], in which positions 6 and 16 in the feature vector stand for the ‘Noun-Positive-Up’ and ‘Verb-Negative’ tokens in the vocabulary, respectively. The scores for both tokens are equal to 1, as each occurs once in the pattern. This process takes account of the semantic and topic-adaptive sentiments via the TASA’s Topic-Adaptive Sentiment Lexicon. But it ignores word and syntactic order, as it applies the BoW model. The feature vectors are unique as the vector length is just 18. This small fixed-length representation of the vocabulary is remarkably efficient in comparison to the thousands of words used in common BoW models.

Finally, pairs of vectors and sentiment labels are fed into the SVM to build a sentiment model. In the training process, the SVM learns the relationship between the feature vectors and sentiment labels. In the prediction process, the SVM predicts the sentiment label for each feature vector in the test set. This model reduces human inputs compared with TASA because all manually rule-based components of the TASA, except the Topic-Adaptive Sentiment Lexicon, are replaced with the SVM, which is trained on a training set built by abnormal returns. The SVM is fitted best with Radial Basis Function (RBF) kernel. For a personal computer with 16 GB RAM and Intel® Core i7 processor, training time is ~1,231 seconds and prediction time is ~550 seconds.

4.3 Hybrid Model 2 (Topic-Adaptive Sentiment Lexicon & Word Order & SVM)

Hybrid model 2 integrates the Topic-Adaptive Sentiment Lexicon, Company Dictionary, and SVM. The feature extractor of this model includes a vocabulary of the 18 TASA tokens presented in Table II and a new token, “CompanyName”. Each token in the vocabulary is linked to a unique number. First, the feature extractor replaces all companies mentioned in an input text with the same token, “CompanyName”, using the Company Dictionary. The feature extractor then reads the input text for the second time and replaces all tokens with the TASA tokens using the Topic-Adaptive Sentiment Lexicon. The tokens that are not listed in the Company Dictionary and Topic-Adaptive Lexicon are removed from the input text by the feature extractor. The feature extractor then reads the text for the last time, token by token, from left to right, and replaces each token with the relevant unique number using the vocabulary. Thus, the text is converted to a vector of numbers considering word order. We let the data calculate the fixed size of feature vectors. We apply the feature extractor to convert all news headline in the entire dataset into feature vectors. We then determine the maximum size, which is 11. Figure V presents the distribution of size of feature vectors using the entire news headlines dataset. Since the SVM works with a fixed length for feature vectors, we set the length of the feature vector to 11.

For example, the headline “3 Reasons to Avoid Apple stock, Motley Fool, 2018-April-10” is converted to “Verb-Negative CompanyName Noun-Positive-Up” using the Company Dictionary and Topic-Adaptive Sentiment Lexicon. The pattern includes 3 tokens and is then initially transformed into [7,1,2], in which 7, 1, and 2 are unique numbers for Verb-Negative, CompanyName, Noun-Positive-Up, respectively. Since the size of the feature vector must be 11, the feature extractor extends to the feature vector [7, 1, 2, 0, 0, 0, 0, 0, 0, 0, 0].

Finally, Pairs of vectors and sentiment labels are fed into the SVM to build a sentiment model. In the training process, the SVM learns the relationship between feature vectors and sentiment labels. In the prediction process, the SVM predicts the sentiment label for each feature vector in the test set. This model takes into consideration word order, syntactics and semantics. This model differs from the TASA in that it replaces the TASA Polarity Pattern Dictionary and TASA Core with the SVM. Hybrid Model 2 eliminates the human inputs by letting the machine learning algorithm learn how to label the patterns using the market abnormal returns. The SVM is fitted best with an RBF kernel. For a personal computer with 16 GB RAM and Intel® Core i7 processor, training time is ~2,321 seconds and prediction time is ~943 seconds.

4.4 Automatic Model (SVM and TF-IDF)

The third model is an automatic machine learning model. The SVM and BoW representation model are used to build a sentiment model devoid of any human inputs. The feature extractor applies a TF-IDF Vectorizer to transform an input text to a vector of numbers.[21] First, the model makes a vocabulary

[21] - TF-IDF is an abbreviation for Term Frequency Inverse Document Frequency. This is very common algorithm to convert an input text into a meaningful representation of numbers. The number representation is

of most influential words using the training set and TF-IDF algorithm. The feature extractor uses the vocabulary and converts an input text into a fixed-length vector of numbers. The size of feature vectors equals the size of the vocabulary. A vector includes one position for each word listed in the vocabulary. This model often generates sparse vectors, as the vocabulary includes thousands of words. The sentiment labels are created using the abnormal returns as explained in the Training and Prediction Processes section. Finally, pairs of vectors and sentiment labels are fed into the SVM to build a sentiment model. In the training process, the SVM learns the relationship between the feature vectors and sentiment labels, which are created by the market. In the prediction process, the SVM predicts the sentiment label for each feature vector in the test set. This model ignores topic-adaptive sentiment, semantics, syntactics, and word order. However, it fully removes the human input into building the model. The SVM is fitted best with a linear kernel. For a personal computer with 16 GB RAM and Intel® Core i7 processor, training time is ~5,655 seconds and prediction time is ~1,353 seconds.

5. Verification of TASA performance

TASA is trained to detect the tone of news headline phrases in an optimal manner through the combination of ML algorithms and human knowledge. Optimality depends on the proportion of correct classifications of news headlines into tone categories (positive, neutral, or negative) by TASA. To verify this claim, we use a common method from data science i.e., classification metrics to assess the accuracy of tone-adapted assignment at the headline-firm level using TASA versus true tone assignment (e.g. Malo et al., 2014; Dridi and Atzeni, 2018; Krishnamoorthy, 2018). We let true tone assignment is defined by the market abnormal returns instead of manually tagging as explained in section 4.1. For a comparison, we also consider the tone assignments of the Hybrid models 1 and 2, Automatic Model, as well as those of two common sentiment analysis models versus true sentiment labels. The first common model incorporates the BoW representation model and the polarity dictionary of L&M, which we refer to as L&M BoW. L&M BoW simply counts for number of positive and negative words to measure

used to fit machine learning algorithms for prediction. we use "TfidfVectorizer" of "SciKit-learn" which is publicly accessible, and usable in various contexts.

sentiment of a news headline. L&M BoW is commonly used to analyse media content in extant finance studies (e.g., Engelberg, et al., 2012; Ferguson, et al., 2015, Garcia, 2013, Liu and McConnell, 2013). The second common model is the VADER Sentiment model (Valence Aware Dictionary and sEntiment Reasoner). This model includes a lexicon and rule-based sentiment analysis tool. It is available in the NLTK package and can be applied directly to unlabelled textual data.

Classification relevance tests are conducted to estimate the accuracy of TASA and other models versus true sentiment labels using four common evaluation metrics including "precision", "recall", "F1-score", and "accuracy". Precision assesses the exactness of a classifier where a lower (higher) precision means more false positives (less false positives). This is measured as: $Precision = \frac{True\ Positive}{Actual\ Result}\ or\ \frac{True\ Positive}{True\ Positive+False\ Positive}$. Recall measures the sensitivity of a classifier, where higher (lower) recall shows less false negatives (more false negatives), which is measured as: $Recall = \frac{True\ Positive}{Predicted\ Result}\ or\ \frac{True\ Positive}{True\ Positive+False\ Negative}$. Finally, F1-score is the harmonic mean of the precision and recall, where an F1-score reaches its best value at 1 (perfect precision and recall). This is measured as: $F1-score = 2 * \frac{Precision*Recall}{Precision+Recall}$. Accuracy is a metric that encapsulates the performance of a classification model as the number of correct predictions divided by the total number of predictions.

The results are presented in Panels A to F of Table IV. Panel A reports the evaluation metrics for the TASA. Panels B, C and D show the evaluation metrics for Hybrid Model 1, Hybrid Model 2, and the Automatic Model, respectively. The evaluation metrics for L&M BoW are presented in Panel E. Panel F shows the evaluation metrics for the VADER Sentiment model. The results in Panel A indicate that the predictive accuracy of tone assignment by TASA is relatively high and it dominates the predictive accuracy of the other sentiment models across all evaluation metrics. In our analysis, we focus on the accuracy and F1-score. The accuracy summarizes the performance of a sentiment model, and the F1-score is the harmonic mean of the recall and precision. As can be seen from Panels C and D, the accuracy of the Hybrid model 1 is 0.63, which is higher than the accuracy of 0.61 for the Automatic model. Specifically, the F1-score for positive (neutral) tone increased from 0.56 (0.62) for the Automatic model to 0.64 (0.69) for the Hybrid model 1. There is only one significant difference between these two models, that is the vocabularies used. The Hybrid model 1 applies the TASA's

Topic-Adaptive Sentiment Lexicon compared with the automated vocabulary built by TF-IDF in the Automatic model. This indicates the importance of Topic-Adaptive Sentiment and human knowledge for sentiment models. As can be seen from Panels C and B, the accuracy for Hybrid model 2 is 0.65, higher than the 0.63 for Hybrid model 1. Specifically, the F1-score for negative (neutral) tone increases from 0.55 (0.69) for the Hybrid model 1 to 0.60 (0.71) for the Hybrid model 2. Both models apply the TASA's Topic-Adaptive Sentiment Lexicon, indicating that the rise in the accuracy is due to considering word order and syntactics in Hybrid model 2 compared with the BoW representation in Hybrid model 1. As can be seen from Panels B and A, the accuracy for TASA is 0.68, higher than the 0.65 for Hybrid model 2. The F1-score for positive (negative) tone increases from 0.61 (0.60) for Hybrid model 2 to 0.66 (0.61) for TASA. Additionally for neutral tone, the F1-score for TASA is 0.83, higher than the 0.71 for Hybrid model 2, demonstrating the function of crafted rule sets in improving sentiment models. As anticipated, the results of tone assignments by the L&M BoW and VADER Sentiment are less satisfactory than for other sentiment models across all classification metrics. Because these two models do not take account of semantics, syntactics and word order there is no training process to tune the sentiment models.

In summary, the verification tests provide evidence of the high accuracy of topic-adapted tone assignments to news headlines at the firm level using TASA. The TASA Topic-Adaptive Sentiment Lexicon is a key function in the TASA performance. Considering word order and syntactics improves sentiment models in which a machine learning algorithm might be a good option for syntax learning. Hybrid models that incorporate both human knowledge and machine learning algorithms are more effective than automatic models.

4. Validation tests using exogenous variables

Sentiment of finance news has been theorized to have a significant impact on market valuation and stock returns of the firms reported on (Tetlock, 2007; Garcia, 2013; Engelberg, et al., 2012). If the classification of the sentiment of headlines using TASA is reliable then our measure of the tone (negative, positive or neutral) for news headlines should, on average, relate to a negative, positive or neutral market response immediately following publication of the news article. In the first validation test, we examine this hypothesis. In the second test, we compare the predictive variability of TASA with the other sentiment models. A sentiment analysis approach is assumed to be superior if the tone assignments more strongly associated with stock market returns post publishing news headlines. In the following two subsections we explain the variable constructions to conduct these tests.

5.1 Variable Constructions

Sentiment measures are constructed to capture the daily media sentiment across "observations" for different combinations of tone, topics, and intensity of press coverage. An observation is a unique pair of firm and time in a trading day. For the Tone-Index group, the negative and positive indexes are defined per observation as:

$$Negative\ (Positive)\ index_{t,j} = \frac{the\ number\ of\ negative\ (positive)\ news\ headlines\ _{t,j}}{total\ number\ headlines\ _{t,j}} \qquad (1)$$

for firm *j* in trading day *t*. In the baseline analysis we consider the Close-to-Close window which focuses on the news headlines, published after the market closes on one day up to the market close on the next day.[22] In robustness tests to take into consideration the time of news publication and the trading hours at the stock exchanges we consider two additional windows, close-to-open and open-to-close windows. The close-to-open window, accounts for the news headlines published after the market closes on one day and before the market opens the next day. The open-to-close window focuses on the news published intraday, i.e. the core trading session from the open to close on a day.

The second group of variables is constructed using the different combinations of topics (business, finance, mixed, and without label) and tone indexes (positive, negative, neutral). For example:

$$\boldsymbol{Business} - Negative(Positive)\ index_{t,j} = \frac{number\ of\ \boldsymbol{business\ news}\ headlines\ with\ negative(positive)\ tone\ _{t,j}}{total\ number\ of\ news\ headlines\ _{t,j}} \qquad (2)$$

[22] The core trading session for the firms in our study is from 9:30am to 4:00pm Eastern Standard time zone (GMT-4).

As can be seen, the indexes are weighted by the intensity of news for firm *i* during time *t*. We call this group of variables the *Topic-Tone-Index*. We also construct a *Media-Optimism index*, which is the difference between the positive index and negative index scores. A score greater than zero indicates a level of media optimism and a score less than zero indicates media pessimism at the firm level.

Fang and Peress (2009) study the cross-sectional relationship between the intensity of media coverage and expected stock returns. They find that high media intensity can alleviate informational frictions and affect security pricing even if it does not supply genuine news. In the third group of variables for the sentiment index, we seek to supplement this previous measure of media intensity by accounting for intensity in both the number of headlines and the number of news providers. Specifically, we define '*negative* (*positive) index-press'* as follow:

$$Negative\ (Positive) - Press\ index\ _{t,j} = \ Negative\ (Positive)\ index\ _{t,j}\ * Press\ _{t,j} \quad (3)$$

Where $Press_{t,j}$ is the number of news providers that published negative (positive) headlines about firm *i* during time *t*. We call this group of variables the Tone-Index-Press. One of the innovations in introducing this measure is to assign higher sentiment score when more news providers share similar sentiments about a firm during time *t*.

For example, on April 20, 2018, Seeking Alpha published three different negative news articles about Apple Inc (NASDAQ: APPL), MotleyFool published three articles (two negative and one neutral), 247WallSt.com published two articles (one negative and one neutral), and InvestorPlace.com published two articles (one positive and one neutral). On this day, there are ten news articles published about APPL, six negative, three neutral and one positive. For the Tone-Index group, the score for APPL based on the *'Negative index'* is 0.6 (6 negative headlines divided by 10 total headlines), *'Positive index'* is 0.1, and '*Neutral index'* is 0.3. For the Tone-Press index group of variables that account for news providers, the score for APPL based on the *'Negative-Press index'* score is 1.8 (*Negative index* of 0.6 multiplied by three for the news providers Seeking Alpha, MotleyFool and 247WallSt.com), and *'Positive-Press index'* score is 0.1 (*'Positive index'* of 0.1 multiplied by one for InvestorPlace.com news provider). Internet Appendix IA.C provides a summary table for variable definitions.

5.2 Validation Tests Using Out-Of-Sample Dataset

We allocate 52.5% (124,637) of the total 236,790 news headlines to the out-of-sample dataset. In this sample there are 170,738 news headlines at the firm level. The summary statistics for these classifications are presented in Panel A of Table V. TASA assigns 13.80%, 39.11%, 47.09% of the out-of-sample headlines at the firm level to negative, neutral, and positive tones, respectively. For this sample 27%, 28.54%, 13.33%, and 31.05% news headlines are classified as business, finance, mixed, and without label at the topic level, respectively.

For our baseline analysis, the aggregation of the 124,637 headlines during the Close-to-Close window resulted in 66,977 firm-day observations. These are used to calculate the daily index score across observations for the different combinations of media sentiment, topics, and news coverage. The summary statistics of the analyses are presented in Panel B of Table V. There is considerable variability in the scores of the different indexes. For example, the mean and median, respectively, 0.47 and 0.50 are for the *Positive index*, 0.13 and 0.00 are for the *Negative index*, 0.13 and 0.00 are for *Finance-Positive index*, and 0.04 and 0.00 are *Finance*-N*egative index*.

The out-of-sample validation tests are based on expectations for the contemporaneous daily market reaction to the tone of the published news headlines. The stock market reaction is measured by risk adjusted Abnormal Returns (*AR)* and Abnormal Trading volume (*AV)* across observations using the Fama-French three-factor (1993) and the Carhart four-factor (1997) models to adjust the trading strategy returns for the value weighted returns on the market, size, book-to-market, and momentum factors. The model is estimated for individual firm-day observations over the period from the 252nd to the 31st trading days prior publishing the news headlines. Observations with fewer than 60 trading days available over the estimation window are excluded. The return is computed as the $close - to - close\ return_{t,j} = \frac{close\ price_{t,j}}{close\ price_{t-1,j}} - 1$, where *t* is time in trading day and *j* is the firm. We refer to the event *t=0* as the existence of any news headlines at firm-day level categorised across tones and topics using the Topic-Adaptive Syntax Approach (TASA).

Panels C and D report summary statistics for $AR_{t,j}$ and Cumulative AR, $CAR(t_1, t_2)$, and $AV_{t,j}$, where *t* equals *-5, -4,..-1, 0, +1, ..,+5*. (t_1, t_2) refers to windows for different combinations of *t*, where

t_1 is the beginning window, and t_2 is the end of the window.[23] As shown in Panel C, the mean and median of $AR_{0,j}$ are 0.008% and -0.01%, respectively.

As a preliminary test, we consider the average daily AR around the event day in four panels for the 221 firms, from 2014 to 2018. Where trading days are reported around day 0 at -10, -4, ...-1, 0, +1, ..., +10. The *Finance Negative* and *Business Negative* variables shown in Panels A and B are dummy variables equal to one when *Finance Negative index* and *Business Negative index* are greater than zero. We construct two similar dummy variables for *Finance Positive* and *Business Positive* graphs shown in Panels C and D. The reported significance of daily AR is computed using the Adjusted Standardized Cross-Section Test (AStdCSect Z) to account for cross-sectional and serial correlation (Kolari and Pynnönen, 2010). Figure III shows a significant association between publishing news headlines and AR on day 0 and that this effect is short lived as the AR rebounds almost fully after a few days. Additionally, the graphs show an asymmetry in this association as the AR is larger for negative headlines than for positive ones and it is larger for the finance headlines than for business headlines. Specifically, the daily AR on event day 0 for the negative finance and business headlines are -0.32% and -0.25%, while those for positive finance and business headlines are 0.15% and 0.06%, respectively. We next report formal tests using tone aggregation at firm-day level utilizing the tone-topic indexes defined in our variable construction.

Regression results for the Close-to-Close Window

To examine the links between the tone of news headlines and contemporaneous abnormal returns, we estimate the following model using cross-sectional Ordinary Least Squares (OLS) regression analysis:

$$AR_{t,j} = \beta' \, sentiment\ index_{\,t,j} + \emptyset' L(5-t)\left(sentiment\ index_{\,t,j}\right) + \gamma' \ L(5-t)\left(AR_{t,j}\right) + F_j + \varepsilon_{j,t} \quad (4)$$

Where ($AR_{t,j}$) is abnormal returns for firm j on day t; *Sentiment index*$_{t,j}$ is a vector of different sentiment indexes as defined above; and $L(5\text{-}t)$ are lag operators for time (5 days -t) for the vector of variables including $AR_{t,j}$, and *Sentiment index*$_{t,}$. F_j is a vector of fixed effects including firm, industry using 4-

[23] All values for CAR and CAV are winsorized at the 1st and the 99th percentiles.

digit SIC code, time, annual market capitalization quartile,[24] and $\varepsilon_{j,t}$ is the robust error term clustered by industry using 4-digit SIC code. The sample sizes drop from 66,977 observations to 66,877 when we run the event study to compute AR and AV and to 63,428 observation when we include market capitalization.

The results are reported in Table VI, Panels A to E. In each panel, we report four models based on different combinations of the ($AR_{j,t}$), year fixed effects, firm fixed effects, industry fixed effects using 4-digit SIC code, and market capitalization fixed effects using quartiles. Panel A shows the *Positive index* and *Negative index*. In all four models, the results show that the coefficient for the *Positive index* is positive and those for the *Negative index* are negative and that both are economically and statistically significant predictors of contemporaneous abnormal returns (AR_0). For example, the coefficient estimates on the *Positive index* suggest that a one standard deviation increase in the *index* is associated with an increase in AR_0 of 3.99 basis points. This amounts to 499% of the mean of AR_0 in our sample.[25] After controlling for fixed effects, the autocorrelation between AR_0 and the lagged AR is very weak. Only AR_{-1} has a robust statistically significant coefficient in models 3 and 4 and the size of the coefficient is relatively small.

These results are robust when using alternative measures for the tone indexes, as shown in Panel B. The coefficients and adjusted R-squares for the *Negative-Press index* and *Positive-Press index* (i.e. when the tone indexes are weighted by the number of news providers) are larger than pure tone indexes. It shows that media has a stronger impact on the market when different news providers publish the same tone for a firm on the same day. In addition, The *Negative index* and *Negative-Press index* coefficients are larger than the *Positive index* and *Positive-Press index*. Veronesi (1999) and Epstein and Schneider (2008) demonstrate theoretically how the response to positive and negative news information can be asymmetric. In general, these arguments are consistent with the patterns displayed for AR in Figure III and differ from the existing evidence for the impacts of positive sentiment from textual analysis of

[24] We obtain annal market capitalization for 221 firms in our study from COMPUSTAT. We sort market capitalization into quartiles. We create a dummy for each quartile per year.

[25] Economic significance is computed as follows: % of AR_0 that can be explained by ton-index = (std dev of tone-index) * (coefficient of tone-index)/ (mean of AR_0). Accordingly, % of AR_0 that can be explained by Positive-Index= (0.0951 x 0.42) / 0.008 x100=499%.

media news. In existing textual analysis studies, the relationships between positive sentiment in news reporting and market responses have been mostly found to be non-significant (e.g. Tetlock, 2007; Engelberg et al., 2012) or, counter intuitively, negative (e.g. Antweiler and Frank, 2004).

Panel C of Table VI reports the link between AR and media sentiment using the tone and topic indexes. As can be seen, all news headline types have significant correlations with $AR_{0,j}$ for the positive and negative tones. Additionally, the finance topic for positive and negative indexes explain a larger proportion of AR than the business news and mixed business and finance news. Specifically, in Model 1, the coefficient for the *Finance-Negative index* is -0.3299 compared to -0.1657 for the *Business-Negative index*. Both coefficients are economically and statistically significant, and the difference between the two coefficients is statistically significant. Additionally, the coefficient for the *Finance-Positive index* is 0.1802 compared to 0.0380 for the *Business-Positive-Index*. Both are statistically significance, at the 1% and 10% levels, respectively. In Model 1, the coefficient estimates on the *Positive-Finance-Index* suggest that a one standard deviation increase is associated with an increase in AR_0 of 1.06 basis points, which amounts to 133% of mean of AR_0 in our sample. However, the coefficient on *Business-Positive index* is statistically weaker than *Finance-Positive-Index*. This may reflect the difficulty of judging the impact of business news on market activities compared to finance news, especially by retail investors.

In Panels D and E of Table VI we use another alternative measure of tone index, i.e. the Media-Optimism indexes. Consistent with the results for other indexes, there is positive relationship between the contemporaneous abnormal returns and the *Media-Optimism index*, *Business-Optimism index,* and *Finance-Optimism index*. This contrasts with previous studies on textual analysis of sentiment in media news, which have mainly discussed pessimism-indexes because of the lack of evidence for positive media news.

A well-known Wall Street proverb “it takes volume to make prices move” has motivated many studies to investigate the relationship between price changes and trading volume. The current evidence validates the positive relationship between change in price and trading volume (e.g. Morgan, 1976; Harris and Gurel, 1986). Our results from Table VI show that there is a significant association between the tone of the news and market returns. These results also suggest that there must be a positive relationship between the tone of the news headlines and the trading volume. In Table VII, we examine the links between *Abnormal volume (AV)* and sentiment dictionaries using Equation 4, in which AR is replaced with *AV*. We obtain abnormal trading volume from the event study using the Fama-French-Momentum Market Model and value-weighted volume index. Consistent with our earlier results we find a positive correlation between *AV* and tone-indexes (Panel A), and tone and topic indexes (Panel B). The only index that shows negative correlation with *AV* is *Business-Positive-Index*.

An extra test is conducted to examine whether there is a delay or a reversal in the reaction of stock market returns to the tone of the headlines. Table VIII reports the cross-sectional OLS estimates for Equation 5 shown below, where the dependent variable in model 1 is *AR* on day 0, in model 2 is *AR* on day 1, in Model 3 is *CAR (0,1)* window, in Model 4 is *AR* on day 2, and Model 5 is *CAR(2, 5)* window, all for firm j.

$$CAR_j\,(t_1,\ t_2) = \beta'\ sentiment\ index\ _{0,j}\ +\ \emptyset' L(5)\big(sentiment\ index\ _{0,j}\big) + \gamma'\ L(5)\big(AR_{0,j}\big) + F_j + \varepsilon\,_j(t_1,\ t_2) \quad (5)$$

The results show that there is no significant market delay to the impact of the tone of news headline during the following five trading days, however, there is a significant reversal in market reactions during days 2 to 5 following negative news headlines. Specifically, in Model 5, the coefficient estimates on the *Negative index* suggest that a one standard deviation increase is associated with an increase in *CAR (2,5)* of 4.07 basis points which amount to -127% of mean *CAR (2,5)* in our sample. This result reveals that the media's negative tone predicts immediate negative returns with gradual reversals within 5 trading days. Based on these results, investors could establish a profitable short-term trading strategy using the tone of new headlines.

In summary, the off-sample validation results show that the tone of news headlines generated from syntactic and semantic dictionary using the TASA approach performs well in explaining the changes in prices and trading volumes.

Regression results for the Close-to-Open and Open-to-Close Windows

To test the robustness of our results, we consider the possibility that news published outside the stock market trading session may have different impact on market activities in comparison to those published during the trading session. Many of the 15 news agencies included in this study publish their news outside the daily trading sessions times of the NY Stock Exchange. News headlines could be published in the afternoon after the trading session ends at 4:00pm EST and before the market opens at 9:30am EST. Impacts due to the tone of the accumulated news headlines published outside the training session might not be as strong or as consistent as those of new headlines published during the trading session. These effects would not be evident for analyses using the *Close-to-Close window* and might affect the precision of our previous results. The impact of the tone of the afternoon and early morning headlines on investors and market prices will be mostly incorporated into the market open-prices relative to previous day close-prices. However, the tone of the headlines published during the day can be captured by the close-prices relative to the same day open-prices.

Accordingly, for every day (24 hours) we consider two windows. In the first window, we examine the link between the tone of news headlines published outside the trading session on the open stock prices. In this session we compute the return from the previous day close price to same day open price (i.e. $open\ return_{\ t,j} = \left(\frac{open\ price_{\ t,j}}{close\ price_{\ t-1,j}}\right) - 1$). We call this session Close-to-Open window. In the second window, we examine the link between the tone of news headlines published during the trading session (i.e. intraday) on the closing stock prices. In this session we compute the return from the same day open price to close price (i.e. $intraday\ return_{\ t,j} = \left(\frac{close\ price_{\ t}}{open\ price_{\ t,j}}\right) - 1$). We call this session Open-to-Close window.

To estimate abnormal returns for Close-to-Open and Open-to-Close windows, we first utilize the market model $R_{t,,j} = \alpha_{t,\,j} + \beta_{t,\,j}\, R_{t,\,m}$ to estimate the daily rolling beta ($\beta_{t,\,j}$) and alpha ($\alpha_{t,\,j}$) for firm j at time t. The daily returns and the value weighted market returns for S&P 500 (R_m) are computed separately for the Close-to-Open and Open-to-Close window. The estimation window is 250 days that start 30 days before the date of publication (event day) and the minimum estimation window is 60 days. For every event we estimate the abnormal returns using the event study approach.

To test the relationship of the tone of news with the Close-to-Open Abnormal Returns ($Open\ AR_{t,j}$) and the Open-to-Close Abnormal Returns ($Intraday\ AR_{t,j}$), we re-estimate Equation 4 replacing AR with either $Open\ AR_{t,j}$ or $Intraday\ AR_{t,j}$. The results are presented in Tables IX and X. In general, the results confirm that the Close-to-Close window captures the market reaction to news headlines during the Close-to-Open and Open-to-Close windows. Both are consistent with and weaker than, but not significantly different from, the relationships between tone of news and market reactions for the Close-to-Close window. For example, the coefficient for the Negative index drops from -0.2829 in model 1 of Table VI (for Close-to-Close window) to -0.1231 in Model 1 of Table IX (for Close-to-Open Window), and to -0.1357 in Model 1 of Table X (Open-to-Close Window). Table XI shows that the sum of the coefficients -0.1231 and -0.1357 is not statically significant from -0.2829. In summary, we confirm the robustness of the results in Table VI using the Close-to-Close window by looking at the sensitivity of the results to the timing of publishing headlines using both Close-to-Open window and Intraday.

5.3 Validation Robustness Tests Using a Post Period Sample

We conduct robustness tests to investigate the sensitivity of our validation results to the news headlines published between January and December 2019, immediately following the period covered by our main sample, i.e. 2014 to 2018. For the post period sample, we collect the news headlines published by the same 15 news providers for 125 firms randomly selected from the 221 firms in the company/sector dictionary, resulting in a total of 20,116 news headlines. In Table XII, we estimate the cross-sectional regression for the post period sample using equation 5 above and the Close-to-Close window. Panel A reports the results for the tone indexes while Panel B reports results for the tone and the topic indexes. In each panel, we report five models where the dependent variable in model 1 is *AR* on day 0, in model 2 is *AR* on day 1, in Model 3 is *CAR (0,1)* window, in Model 4 is *AR* on day 2, and Model 5 is *CAR(2, 5)* window, all for firm j.

Although our TASA approach is not updated using the news headlines in the post period, the patterns of results in Panels A and B in Table XII remain the same. For example, Model 1 in Panel A shows that the coefficient for the *Positive index* is positive and for the *Negative index* it is negative. Both coefficients are economically and statistically significant predictors of contemporaneous abnormal returns (AR_0). Models 2 to 5 show that there are no significant market delays in the impact of the tone of news headline during the following five trading days. However, there is a significant reversal in the *CAR(2, 5)* window following negative news headlines.

5.4 Validation tests: Comparing TASA with the Other Sentiment Models

In the second validation test, we compare the performance of the content analysis using TASA with the Hybrid Model 2 (TASA's Topic-Adaptive Sentiment Lexicon & word order & SVM), Automatic Model (SVM and TFIDF), and L&M BoW. A content analysis method is assumed to be superior if it is less likely to misclassify words and sentences while considering other exogenous variable, i.e. the tone assignments are more strongly associated, economically and statistically, with patterns observed in stock market returns post publishing news headlines. We call this hypothesis the Superiority Hypothesis.

To examine the Superiority-Hypothesis, we first use the test set and construct the tone indexes for the Close-to-Close window across 19,635 observations, for all sentiment models. The test set is an unobserved news headlines dataset for the sentiment models and explained in Section 4.1. Second, we conduct a joint cross-sectional regression analyses to estimate Equation 4 across 8 models in which the vector $Sentiment\ Index_{j,t}$ includes tone indexes for all sentiment models (*Sentiment-Index TASA* $_{t,j}$, *Sentiment-Index Hybrid 2* $_{t,j}$, *Sentiment-Index Automatic* $_{t,j}$, *and Sentiment Index L&M BoW* $_{t,j}$).

The Pearson's correlations are presented in Panels A and B of Table XIII for all tone indexes and contemporaneous abnormal return. The results show that the TASA's tone assignment has the highest correlation with contemporaneous abnormal returns. It follows by the Hybrid Model 2, Automatic Model, and L&M BoW, respectively. The results of the regression analysis for the 8 models are presented in Panels A and B of Table XIV. The dependent variable in all models is *AR* on day 0 for firm *j*, *i.e.*, $AR_{0,j}$.

Panel A of Table XIV presents the results of for both Positive and Negative indexes. As can be seen, Models 5 to 8 run the regression separately for each sentiment model. The results show that the coefficients on both Negative and Positive indexes for TASA are statistically and economically larger than that for the other sentiment models. Specifically, the coefficient on *Negative index* for TASA is -0.069 compared to -0.047, -0.040, -0.037 for the Hybrid Model 2, Automatic Model, and L&M Bow, respectively. The coefficient on *Positive index* for TASA is 0.050 compared to 0.045, 0.039, 0.028 for the Hybrid Model 2, Automatic Model, and L&M Bow, respectively. In addition to, Models 1 to 4 present the results of a joint cross-sectional regression studies in which the explanatory power for AR $_{0,\ j}$ is compared across the sentiment models. Specifically, Model 4 indicates that the coefficients on both Positive and Negative indexes of TASA economically and significantly higher than those for the other sentiment models. For the Negative index, the TASA's coefficient is -0.056 compared with -0.024, -0.026, and -0.012 for the Hybrid Model 2, Automatic Model, and L&M BoW, respectively. For the Positive index, the TASA's coefficient is 0.039 compared with 0.026, 0.033 and 0.021 for the Hybrid Model 2, Automatic Model, and L&M BoW, respectively. Model 3 shows that the Hybrid Model 2 outperforms both Automatic Model and L&M BoW. In all models of Panel A, the differences between the Tone-Indexes coefficients for each pair of sentiment models are statistically significant, but that for the Hybrid and Automatic models in the regression Model 4.

Panel B of Table XIV reports the regression study's results for the 8 models in which the vector $Sentiment\ Index_{j,t}$ includes Optimism index $_{0,\ j}$ for all sentiment models. The results are consistent with Panel A. As can be seen from Panel B, the coefficient on Optimism index for TASA is positive and statistically and economically larger than that for the other sentiment models. Specifically, as shown in Models 5 to 8, the coefficient on Optimism index for TASA is 0.058 compared with 0.041, 0.032, and 0.027 for the Hybrid Model 2, Automatic Model, and L&M BoW.

In summary, the results reported in this section supports the following conclusions. First, for estimating the relationship between tone of news and stock returns, all sentiment models are superior to L&M BoW especially for negative and positive news. These results are consistent with Engelberg et al. (2012), Garcia (2013), and Garcia et al. (2020), who find asymmetric predictability in the sentiment of news content using L&M BoW and stock market activities. It indicates the importance of semantic, word order, syntactic, and training for upgrading the sentiment models in finance. Second, the TASA model significantly outperforms all other machine learning models i.e., the Hybrid and Automatic models. Third, the Hybrid Model is superior to the Automatic Model for explaining the market movements. Conclusions 2 and 3 confirm that adding human knowledge to sentiment models improves the model's explanatory power for the stock market in comparison to using fully automatic models. Forth, the results are consistent with the verification study in Section 5.

5. Conclusion

We propose a novel approach for *sentiment analysis* of finance and business news that combines *topic modelling* with analysis of the deep structure of text. TASA assesses the sentiment of text using tone-syntax patterns rather than the BoW format. A major contribution of TASA is the improved accuracy of tone assignment through construction of the *Topic-Adaptive Sentiment Lexicon*.

We consider a battery of tests to verify the accuracy of the topic-adapted tone assignment at the firm level by TASA. In our main tests we compare the performance of TASA with a variety of other models that include the basic BoW model, hybrid models and fully automated models. In the hybrid models we incrementally integrate components of TASA with supervised machine learning approaches, e.g. SVM. Our main objective is to identify the impacts of different components of TASA on overall performance of the model and to gauge whether fully automated textual analysis and machine learning models outperform the rule-based approach of TASA.

Accuracy in the categorization of the tone of the news headlines by TASA is relatively high and outperforms categorizations using all other models. In the validation tests, we show that the TASA topic-adapted tone assignment at the firm-day level is correlated with stock market returns and trading volumes and that it outperforms all other models. These results confirms that supervised machine learning algorithms can be enhanced when the *tone-syntax patterns* proposed by TASA replace the *Bag-of-Word format*. Additionally, it shows that adding human knowledge to sentiment models improves the model's explanatory power for the stock market in comparison to using fully automatic models.

The TASA approach is adaptable to address new investigations. Specifically, the *Topic Lexicon* can be modified from business and finance to include a wide set of topics (e.g. merger and acquisitions, earning announcements, etc). Additionally, the *Topic-adaptive Sentiment Lexicon*, can be expanded to include new words and phrases using ML algorithms (e.g. 'Covid-19' and 'Coronavirus' are noun and noun-phrases, respectively, and will be categorized as 'positive-down). The TASA approach can also be applied in analyses of short texts including news feed and social media written in formal English (e.g. tweets published by a firm).

References

Ali, F., Kwak, D., Khan, P., El-Sappagh, S., Ali, A., Ullah, S., & Kwak, K. S. (2019). Transportation sentiment analysis using word embedding and ontology-based topic modeling. *Knowledge-Based Systems*, 174, 27-42.

Agarwal, S., Chen, V. Y., & Zhang, W. (2016). The information value of credit rating action reports: A textual analysis. Management Science, 62(8), 2218-2240.

Antweiler, W., & Frank, M. Z. (2004). Is all that talk just noise? The information content of internet stock message boards. *The Journal of finance*, *59*(3), 1259-1294.

Bao, Y., & Datta, A. (2014). Simultaneously discovering and quantifying risk types from textual risk disclosures. Management Science, 60(6), 1371-1391.

Chan, S. W., & Chong, M. W. (2017). Sentiment analysis in financial texts. *Decision Support Systems*, *94*, 53-64.

Chen, J., Demers, E., & Lev, B. (2018). Oh what a beautiful morning! Diurnal influences on executives and analysts: Evidence from conference calls. *Management Science*, 64(12), 5899-5924.

Das, S. R., & Chen, M. Y. (2007). Yahoo! for Amazon: Sentiment extraction from small talk on the web. *Management science*, *53*(9), 1375-1388.

Deng, D., Jing, L., Yu, J., Sun, S., & Ng, M. K. (2019). Sentiment lexicon construction with hierarchical supervision topic model. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 27(4), 704-718.

Dridi, A., Atzeni, M., & Recupero, D. R. (2019). FineNews: fine-grained semantic sentiment analysis on financial microblogs and news. *International Journal of Machine Learning and Cybernetics*, *10*(8), 2199-2207.

Engelberg, J. E., Reed, A. V., & Ringgenberg, M. C. (2012). How are shorts informed?: Short sellers, news, and information processing. *Journal of Financial Economics*, *105*(2), 260-278.

Epstein, L. G., & Schneider, M. (2008). Ambiguity, information quality, and asset pricing. *The Journal of Finance*, *63*(1), 197-228.

Fang, L., & Peress, J. (2009). Media coverage and the cross-section of stock returns. *The Journal of Finance*, *64*(5), 2023-2052.

Ferguson, N. J., Philip, D., Lam, H., & Guo, J. M. (2015). Media content and stock returns: The predictive power of press. *Multinational Finance Journal*, *19*(1), 1-31.

Garcia, D. (2013). Sentiment during recessions. *The Journal of Finance*, *68*(3), 1267-1300.

Garcia, D., Hu, X., & Rohrer, M. (2020). The Colour of Finance Words. *Working Paper*, Available at SSRN 3630898.

García-Pablos, A., Cuadros, M., & Rigau, G. (2018). W2VLDA: almost unsupervised system for aspect based sentiment analysis. *Expert Systems with Applications*, 91, 127-137.

Harris, L., & Gurel, E. (1986). Price and volume effects associated with changes in the S&P 500 list: New evidence for the existence of price pressures. *the Journal of Finance*, *41*(4), 815-829.

Huang, A. H., Lehavy, R., Zang, A. Y., & Zheng, R. (2018). Analyst information discovery and interpretation roles: A topic modeling approach. Management Science, 64(6), 2833-2855.

Jegadeesh, N., & Wu, D. (2013). Word power: A new approach for content analysis. *Journal of financial economics*, *110*(3), 712-729.

Kearney, C., & Liu, S. (2014). Textual sentiment in finance: A survey of methods and models. *International Review of Financial Analysis*, *33*, 171-185.

Kolari, J. W., & Pynnönen, S. (2010). Event study testing with cross-sectional correlation of abnormal returns. *The Review of financial studies*, 23(11), 3996-4025.

Konnikova, M. (2014). How headlines change the way we think. *The New Yorker*, Retrieved May 26, 2015.

Krishnamoorthy, S. (2018). Sentiment analysis of financial news articles using performance indicators. *Knowledge and Information Systems*, *56*(2), 373-394.

Liu, B., & McConnell, J. J. (2013). The role of the media in corporate governance: Do the media influence managers' capital allocation decisions? *Journal of Financial Economics*, *110*(1), 1-17.

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

Loughran, T., & McDonald, B. (2020). Textual Analysis in Finance. *Working Paper,* Available at SSRN 3470272.

Malo, P., Sinha, A., Korhonen, P., Wallenius, J., & Takala, P. (2014). Good debt or bad debt: Detecting semantic orientations in economic texts. *Journal of the Association for Information Science and Technology*, *65*(4), 782-796.

Manela, A., & Moreira, A. (2017). News implied volatility and disaster concerns. *Journal of Financial Economics*, *123*(1), 137-162.

Meyer, B., Bikdash, M., & Dai, X. (2017, March). Fine-grained financial news sentiment analysis. In *Southeast Con 2017* (pp. 1-8). *IEEE*.

Morgan, I. G. (1976). Stock prices and heteroscedasticity. *The Journal of business*, *49*(4), 496-508.

Salas-Zárate, M. D. P., Valencia-García, R., Ruiz-Martínez, A., & Colomo-Palacios, R. (2017). Feature-based opinion mining in financial news: an ontology-driven approach. *Journal of Information Science*, *43*(4), 458-479.

Shapiro, A. H., Sudhof, M., & Wilson, D. (2020, March). Measuring news sentiment. *Federal Reserve Bank of San Francisco*.

Schumaker, R. P., Zhang, Y., Huang, C. N., & Chen, H. (2012). Evaluating sentiment in financial news articles. *Decision Support Systems*, 53(3), 458-464.

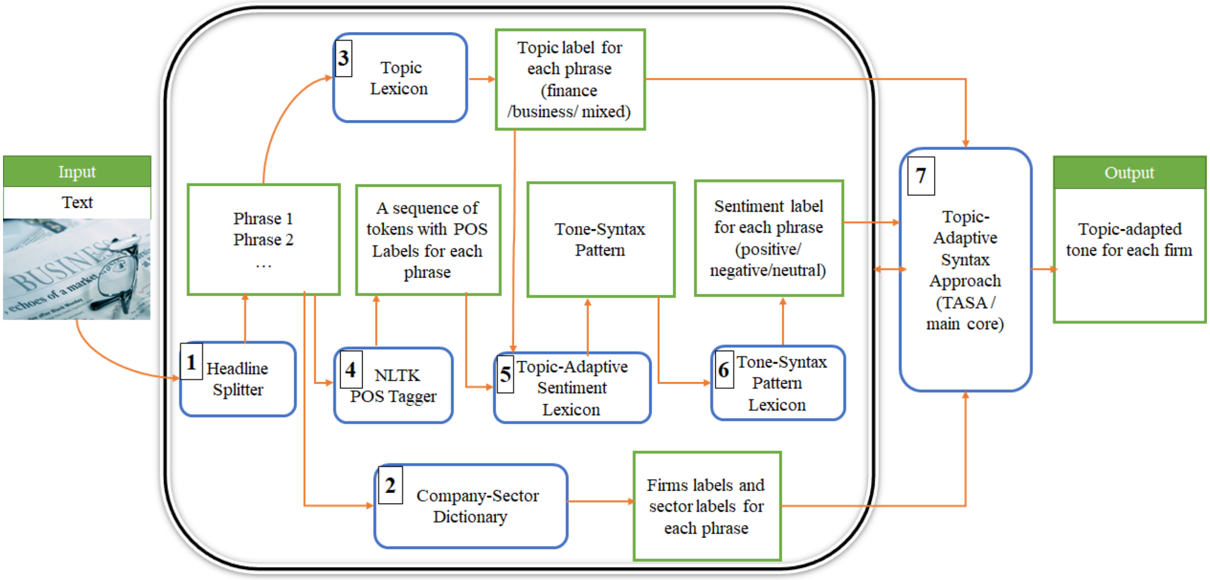
Tetlock, P. C. (2007). Giving content to investor sentiment: The role of media in the stock market. *The Journal of finance*, *62*(3), 1139-1168.

Veronesi, P. (1999). Stock market overreactions to bad news in good times: a rational expectations equilibrium model. *The Review of Financial Studies*, *12*(5), 975-1007.

Figure I: The Conceptual model of our approach for sentiment analysis of media news

The schematic in Figure I presents the conceptual structure of our approach for classifying the tone of news headlines at the firm level. The schematic starts from the input of business and financial media news and shows the processing system step leading to the output of categorizing news sentiment at the firm level. On the right side of the schematic is the sentiment score (s) at the firm level and its topic (business versus finance).Components are shown by the blue rectangles, workflow by orange arrows, and the outputs of each components are presented in green rectangles. This Figure is presented in two panels. Panel A shows a conceptual model of our approach and Panel B presents an example using the conceptual model.

Panel A: Conceptual model



Panel B: An example using the conceptual model

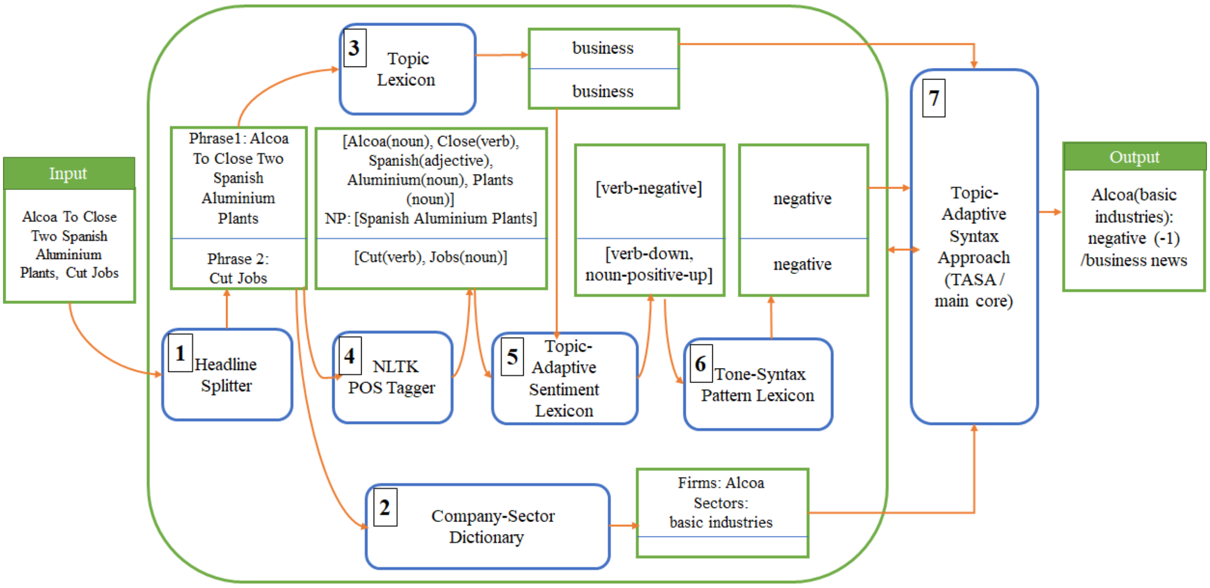
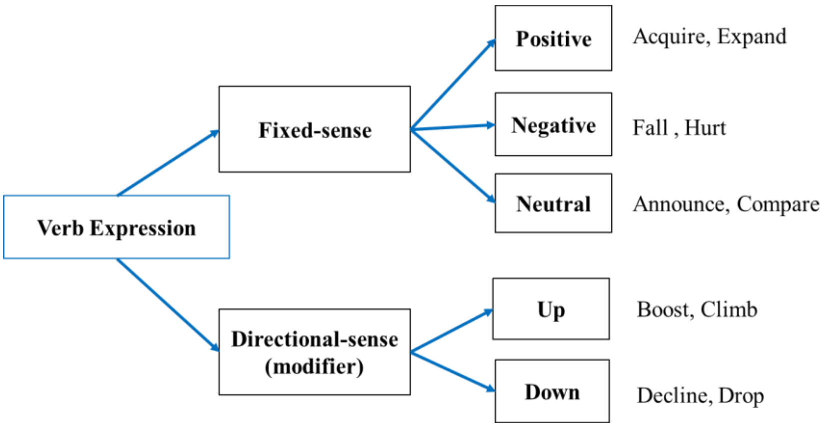


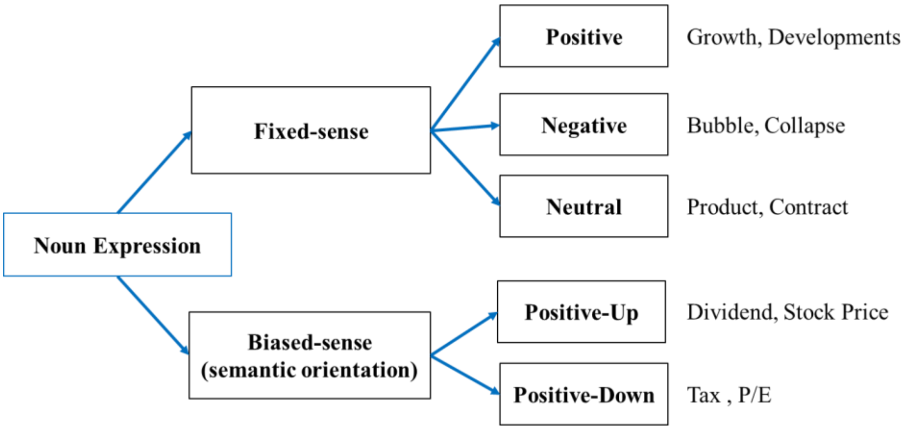
Figure II. The unigram list of topic-adaptive sentiment lexicon.

We show the classification for verbs in Panel A, for nouns in Panel B, for adjectives in Panel C, and for adverbs in Panel D.

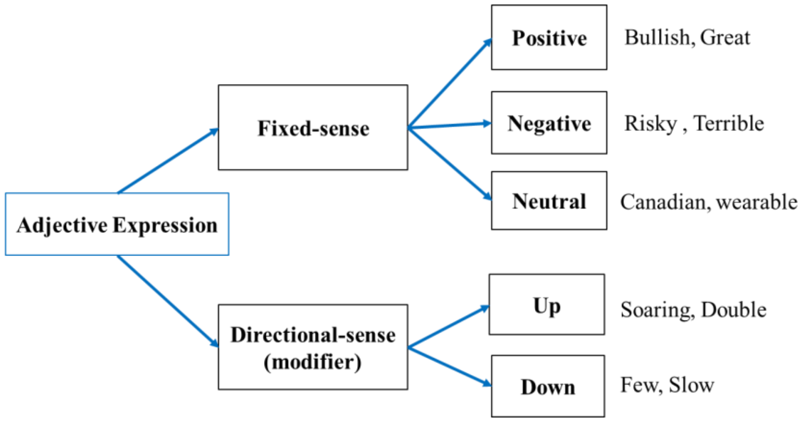
Panel A: The classifications of *verbs* at the tone level.



Panel B: The classifications of *nouns* at the tone level.



Panel C: The classifications of *adjectives* at the tone level.



Panel D: The classifications of *adverbs* at the tone level.

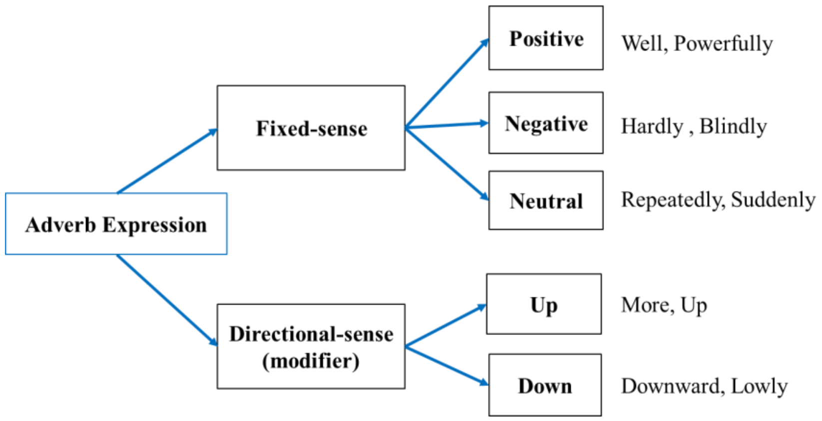


Figure III. Abnormal Returns (AR) around the events of releasing media news

We obtain AR from event study tests using the Fama-French-Momentum model and value weighted index. The *Finance Negative* and *Business Negative* variables in the Panels A and B are dummy variables equal to one when *Finance Negative index* and *Business Negative index* are greater than zero. We construct two similar dummy variables *Finance Positive* and *Business Positive* in Panels C and D. The four panels show the paragraph of the daily average AR around the event of publishing news headlines at the firm level for the 221 firms, from 2014 to 2018. The event day of the releasing of the news day is day 0 and the trading days are labelled around day 0 as -10, … ,-1 , 0, +1, ...,+10. $, *, **, and *** denote statistical significance at the 0.10, 0.05, 0.01 and 0.0, respectively. The significance of daily AR is computed using Adjusted Standardized Cross-Section Test (AStdCSect Z) to account for cross-sectional and serial correlation, Kolari and Pynnönen (2010).

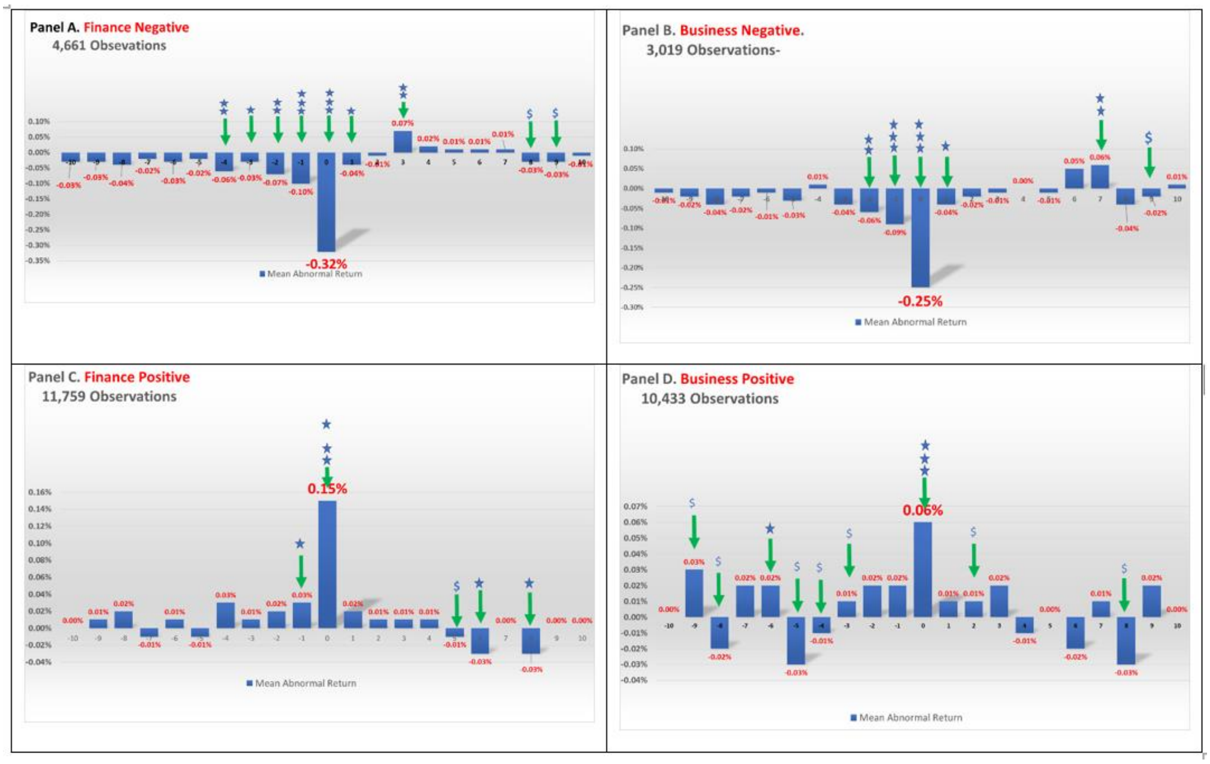


Figure IV. Distribution of Abnormal Returns (AR) for pairs of headline-firm in the entire news headline dataset.

We obtain AR from event study tests using the Fama-French-Momentum model and value weighted index. Panel A shows the distribution of abnormal returns on day 0 for firm j (AR0,j), the day of publishing news headlines, across pairs of headline-firm. The distribution is fitted into a histogram consists of 43 bins, in which the width of each bin equals to 0.232%. Panel B presents the histogram built by $AR_{-1,\,j}$ across pairs of headline-firm. Where, $AR_{-1,\,j}$ is AR on the preceding day of publishing news headlines for firm j. The histogram includes 43 bins, in which the width of each bin equals to 0.133%. Panel C shows the distribution of CAR_j(-1,+1) across the pairs of headline-firm. CAR_j(-1,+1) stands for Cumulative Abnormal returns on firm j on trading days -1 to +1 form the day of publishing headlines day 0. The distribution is fitted into a histogram including 43 bins, in which the width of each bin equals to 0.232%. The solid line curve shows the normal distribution in all panels.

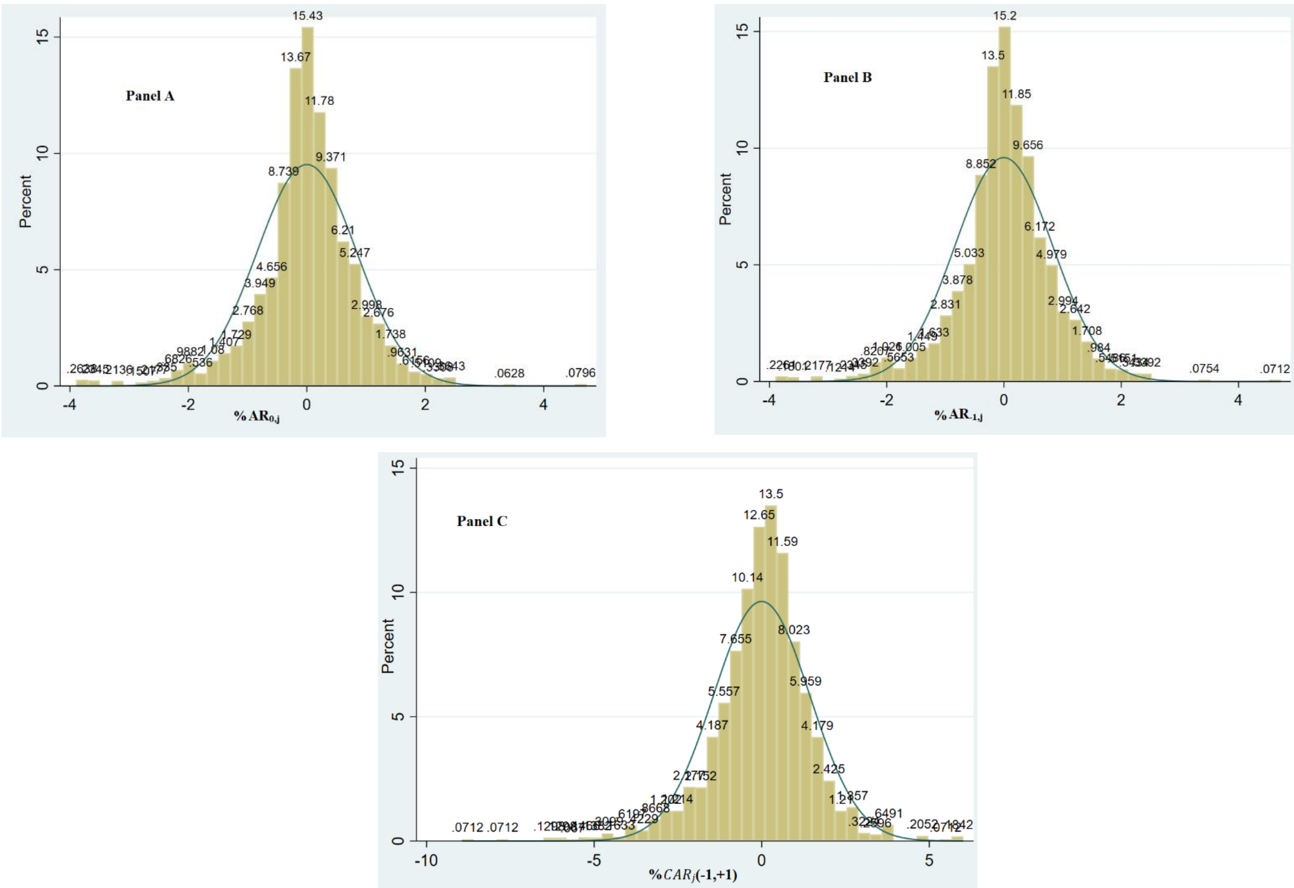


Figure V. Distribution of the size of the feature vectors across the entire news headline dataset

The feature extractor from Model 2 is applied to generate feature vectors across the entire news headline dataset.

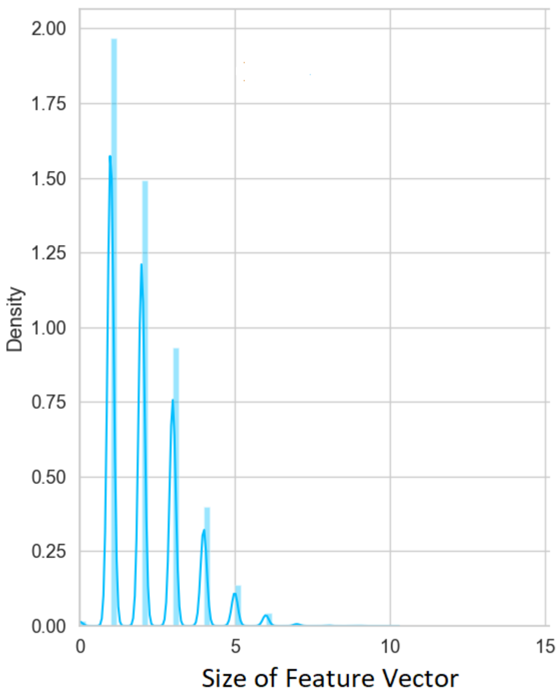


Table I. Summary statistics of company-sector dictionary.

Our company-sector dictionary is constructed from S&P 500 firms. Using the full sample of 322,000 news headlines from 15 major news provider, we found 221 out of S&P 500 firms were mentioned in those headlines. The 221firms belong to 13 sectors, and their shares are traded on three exchanges NYSE, NASDAQ and AMEX.

Sector	Stock exchanges			**Total**
	NASDAQ	NYSE	AMEX	
Basic Industries		8	1	**9**
Capital Goods	4	18	1	**23**
Consumer Durables		6		**6**
Consumer Non-Durables	7	8		**15**
Consumer Services	11	20	1	**32**
Energy	1	11		**12**
Finance	7	14		**21**
Health Care	14	22	2	**38**
Miscellaneous	5	5	1	**11**
Public Utilities	1	6		**7**
Technology	27	11	1	**39**
Transportation	2	2		**4**
n/a	3	1		**4**
Total	**82**	**132**	**7**	**221**

Table II: Summary of our Topic-Adaptive Sentiment Lexicon versus the L&M dictionary

The Topic-Adaptive Sentiment Lexicon contains 2,562 words and phrases including 635 (25%) words and phrases with a negative tone, and 1,222 (48%) words and phrases with a positive tone. 479 (19%) words and phrases fall in the semantic orientation category because they do not convey tone without a modifier. A further 226 (9%) words are directional (i.e. modifiers). In the last three columns, we compare our Topic-Adaptive Sentiment Lexicon with L&M's.

Categories		Total	Count	Intersect with L&M's Negative	Intersect with L&M's Positive	Intersect with L&M's Neutral
Negative	Noun-Negative	635	182	69	0	96
	Verb-Negative		240	71	0	147
	Adjective-Negative		24	10	0	11
	NounPhrase-Negative		170	0	0	0
	Phrase-Negative		19	0	0	0
Positive	Noun-Positive	1,222	230	0	27	178
	Verb-Positive		354	0	32	291
	Adjective-Positive		52	0	14	29
	NounPhrase-Positive		562	0	0	0
	Phrase-Positive		24	0	0	0
semantic orientation	Noun-Positive-Down	479	9	1	0	7
	Noun-Positive-Up		103	0	0	78
	NounPhrase-Positive-Down		11	0	0	0
	NounPhrase-Positive-Up		356	0	0	0
Direction	Verb-Up	226	129	3	6	107
	Verb-Dow		64	21	0	34
	Adjective-Up		25	0	0	24
	Adjective-Down		8	1	0	7
Grand Total		2,562	2,562	176	79	1,009

Table III: A Sample of tone-syntax patterns from the Tone-Syntax Pattern Lexicon

Using 111,000 in-sample news headlines, we extract 5,432 tone-syntax patterns. Subsequently, our annotators team review each pattern in each phrase and give it a tone by considering the original news. In this Table we give some examples from the 5,432 tone-syntax patterns and their assigned tone.

Pattern	Frequency in in-sample news headlines	Assigned Sentiment Label by Audit Team
[‘Noun-Positive’]	25,377	Positive (+1)
[‘Noun-Up’]	22,291	Neutral (0)
‘Verb-Positive’	14,488	Positive (+1)
[‘Negation’, ‘Verb-Positive’]	10,431	Negative (-1)
[‘Noun-Negative’]	10,204	Negative (-1)
[‘Verb-Negative’]	5,691	Negative (-1)
[‘Verb-Positive’, ‘Noun-Positive-Up’]	5,649	Positive (+1)
[‘Adjective-Positive’, ‘Noun-Positive’]	5,530	Positive (+1)
[‘NounPhrase-Positive-up’]	4,717	Neutral (0)
[‘Adjective-Positive’]	4,660	Positive (+1)
[‘Verb-Positive’, ‘Noun-Positive’]	4,237	Positive (+1)
[‘NounPhrase-Positive’]	3,619	Positive (+1)
[‘Verb-Up’, ‘NounPhrase-Positive-up’]	3,390	Positive (+1)
[‘NounPhrase-Positive’, ‘Adjective-Positive’]	2,618	Positive (+1)
[‘Verb-Neg’, ‘Noun-Up’]	2,473	Negative (-1)
['Verb-Up', 'Noun-Positive-Up', 'Noun-Positive']	104	Positive (+1)
['Verb-Negative', 'Noun-Negative', 'Noun-Positive-Up']	132	Negative (-1)

Table IV. Verification performance tests

This table presents classification relevance tests conducted to estimate the performance of the sentiment models versus true sentiment labels using the test set. The test set and true sentiment labels are explained in Section 4.1. Four common evaluation metrics are used including "precision", "recall", "F-measure", and "accuracy". The evaluation metrics are defined in Section 5. Panel A reports the evaluation metrics for the TASA. Panels B, C, D show the evaluation metrics for Hybrid Model 2, Hybrid Model 1, Automatic Model, respectively. The evaluation metrics for L&M BoW are presented in Panel E. Panel F shows the evaluation metrics for the VADER Sentiment model.

Panel A) TASA

	precision	recall	f1-score	accuracy	support
negative	0.56	0.65	0.61	-	14,452
neutral	0.98	0.68	0.83	-	23,854
positive	0.6	0.72	0.66	-	20,300
Total	-	-	-	0.68	58,606

Panel B) Hybrid Model 2 (TASA's Polarity Dict. & Word Order & SVM)

	precision	recall	f1-score	accuracy	support
negative	0.55	0.65	0.6	-	14,452
neutral	0.72	0.68	0.71	-	23,853
positive	0.6	0.61	0.61	-	20,300
Total	-	-	-	0.65	58,605

Panel C) Hybrid Model 1 (TASA's Polarity Dict. & BoW & SVM)

	precision	recall	f1-score	accuracy	support
negative	0.51	0.58	0.55	-	14,452
neutral	0.78	0.6	0.69	-	23,853
positive	0.58	0.7	0.64	-	20,300
total	-	-	-	0.63	58,605

Panel D) Automatic Model (SVM-TF-IDF)

	precision	recall	f1-score	accuracy	support
negative	0.53	0.63	0.59	-	14,452
neutral	0.57	0.66	0.62	-	23,853
positive	0.59	0.53	0.56	-	20,300
total	-	-	-	0.61	58,605

Panel E) L&M BoW

	precision	recall	f1-score	accuracy	support
negative	0.48	0.2	0.31	-	14,452
neutral	0.56	0.98	0.77	-	23,853
positive	0.6	0.19	0.39	-	20,300
total	-	-	-	0.52	58,605

Panel F) VADER Sentiment

	precision	recall	f1-score	accuracy	support
negative	0.41	0.32	0.35		14,452
neutral	0.6	0.73	0.66		23,853
positive	0.53	0.5	0.52		20,300
total	-	-	-	0.55	58,605

Table V. Summary statistics for the Validation sample

In this Table we present the summary statistics for the out-of-Sample 124,637 news headlines. Panel A presents the TASA tone assignment for 170,738 news headlines at the firm level. The aggregation of the news headlines during the Close-to-Close window resulted in 66,977 observations. Panel B presents summary statistics for the daily index score for the different combinations of media sentiment, topics, news coverage using Close-to-Close window. The definition of the tone indexes is provided in Internet Appendix IA.C. Panel C reports summary statistics for $AR_{t,j}$ and Cumulative AR, $CAR(t_1, t_2)$, while Panel D reports summary statistics for $AV_{t,j}$. Where *j* denotes firm, *t* equals *-5, -4,..-1, 0, +1, ..,+5*, and (t_1, t_2) refers to window of different combination of trading day *t*. Panel E reports the summary statistics of the market capitalization. AR and AV are estimated from Fama-French three-factor (1993) and the Carhart four-factor (1997) using the value weighted index. The model is estimated for individual firms over 252nd trading day to the 31st trading day prior publishing news headline. We refer to the event day as 0. Firm-day Observations with fewer than 60 trading days available over the estimation window are excluded. The return is computed as $Close-to-Close\ return_t = \frac{close\ price_t}{close\ price_{t-1}} - 1$ .The market variables are winsorized at the 1st and the 99th percentiles.

Panel A. Tone and topic classification using TASA approach

	Negative	Neutral	Positive	Total	%Total
Business	6,287	18,181	21,756	46,224	27.07%
Finance	7,129	16,758	24,850	48,737	28.54%
Mix	3,289	7,575	11,903	22,767	13.33%
Without label	6,854	24,260	21,896	53,010	31.05%
Total	23,559	66,774	80,405	170,738	100.00%
%Total	13.80%	39.11%	47.09%	100.00%	-

Panel B: Sentiment indexes

Index	Min	5%	25%	50%	75%	95%	Max	Mean	Std. Dev.
Negative index	0	0	0	0	0.10	1.00	1.00	0.13	0.28
Positive index	0	0	0	0.50	1.00	1.00	1.00	0.47	0.42
Business-Negative index	0	0	0	0	0	0.25	1.00	0.04	0.15
Business-Positive index	0	0	0	0	0	1.00	1.00	0.13	0.28
Finance-Negative index	0	0	0	0	0	0.25	1.00	0.04	0.16
Finance-Positive index	0	0	0	0	0	1.00	1.00	0.13	0.28
Mixed-Negative index	0	0	0	0	0	0.00	1.00	0.02	0.11
Mixed-Positive index	0	0	0	0	0	0.50	1.00	0.07	0.22
Media-Optimism index	-1.00	-1.00	0	0.33	1.00	1.00	1.00	0.34	0.59
Business-Optimism index	-1.00	-0.17	0	0	0	1.00	1.00	0.09	0.32
Finance-Optimism index	-1.00	-0.17	0	0	0	1.00	1.00	0.09	0.34
Mixed-Optimism index	-1.00	0	0	0	0	0.50	1.00	0.06	0.25
Negative-Press index	0	0	0	0	0.10	1.00	5.25	0.16	0.36
Positive-Press index	0	0	0	0.50	1.00	2.00	6.59	0.66	0.70

Panel C: Percentage Abnormal Returns (AR) and Cumulative Abnormal Returns (CAR)

AR_t / CAR (t_1, t_2)	Min	5%	25%	50%	75%	95%	Max	Mean	Std. Dev.
AR_{-5}	-4.79	-1.90	-0.58	-0.02	0.56	1.95	4.69	-0.006	1.220
AR_{-4}	-4.33	-1.94	-0.58	-0.02	0.56	1.99	4.78	-0.003	1.238
AR_{-3}	-4.95	-1.99	-0.59	-0.01	0.60	2.06	5.01	0.006	1.288
AR_{-2}	-4.95	-2.07	-0.60	-0.02	0.59	2.10	5.74	0.001	1.336
AR_{-1}	-5.76	-2.16	-0.61	-0.02	0.61	2.21	5.93	-0.001	1.443
AR_{0}	-6.07	-2.26	-0.62	-0.01	0.63	2.35	6.62	0.008	1.536
AR_{+1}	-4.98	-2.09	-0.61	-0.02	0.58	2.02	5.02	-0.019	1.307
AR_{+2}	-4.87	-2.04	-0.60	-0.02	0.59	2.06	5.10	-0.004	1.304
AR_{+3}	-4.91	-1.96	-0.59	-0.02	0.56	1.98	4.86	-0.008	1.250
AR_{+4}	-4.65	-1.91	-0.59	-0.03	0.55	1.94	4.96	-0.013	1.213
AR_{+5}	-4.40	-1.91	-0.58	-0.02	0.55	1.95	4.56	-0.008	1.209
CAR (0,1)	-6.1	-3.22	-0.92	-0.01	0.9	3.16	6.17	-0.009	2.042
CAR (2,5)	-7.05	-3.99	-1.28	-0.03	1.2	3.96	6.99	-0.032	2.500

Panel D: Percentage Abnormal Volume (AV)

$AV_{t,j}$	Min	5%	25%	50%	75%	95%	Max	Mean	Std. Dev.
%AV -5, j	-96.69	-56.94	-24.40	-3.04	21.01	73.84	144.14	0.8493	39.7426
%AV -4, j	-96.20	-56.35	-23.91	-2.57	21.79	75.13	145.16	1.5402	40.2035
%AV -3, j	-98.83	-55.89	-23.49	-2.10	22.69	76.57	147.01	2.2774	40.4615
%AV -2, j	-95.82	-55.04	-22.76	-1.22	24.02	81.61	155.43	3.6757	41.7100
%AV -1, j	-89.83	-54.27	-21.69	0.34	26.58	86.99	166.30	6.0379	43.5128
%AV 0, j	-91.38	-52.68	-20.89	1.06	27.95	94.22	175.17	7.6604	44.9093

Panel E: Summary statistics of market capitalization

	Market capitalization				**Total**
	QUARTILE 1	QUARTILE 2	QUARTILE 3	QUARTILE 4	
Number of Industries (SIC4dig)	98	56	46	23	**126**
Number of firms	137	78	61	31	**221**
mean (Billions of USD)	12	64	143	321	**135**
Std. Dev. (Billions of USD)	10	19	27	163	**144**
Max (Billions of USD)	33	96	189	1,090	**1,090**
Min (Billions of USD)	0	33	96	190	**0**

Table VI. Validation results using out-of-sample dataset and abnormal returns

This table reports the results from cross-sectional regression analyses to examine the links between the tone and topic of news headlines and contemporaneous abnormal returns using Equation 4. The dependent variable is Abnormal Returns on day 0 for firm *j* ($AR_{0,j}$), the day of publishing news headlines. All abnormal returns are reported in percentages. Definitions of the sentiment indexes are provided in Internet Appendix IA.C. There are 66,977 firm-day observations using the Close-to-Close window. The sample size drops to 66,877 when we consider AR, and to 63,428 when we include market capitalization. The results reported in Panel A focus on tone indexes, Panel B results focus on tone index weight adjusted by the number of news providers, Panel B results utilize the tone and the topic of news headlines indexes, Panel D results focus on media optimism index, and Panel E results focus on media optimism and the topic of news headlines indexes. The results are reported for four models based on different combinations of lagged AR, lagged sentiment index, year fixed effects, firm fixed effects, industry fixed effects using 4-digit SIC code, and annual market capitalization fixed effects using quartiles. We obtain daily AR from event study tests using the Fama-French-Momentum model and value weighted index. All abnormal returns are winsorized at 1% and 99% by year. The robust error term is clustered for industry 4-digit SIC code. t-statistics are in parentheses and *, **, *** represent significance levels of 0.1, 0.05, 0.01, respectively.

Dependent variable	AR $_{0, j}$			
Models	(1)	(2)	(3)	(4)
Panel A: Tone indices				
Negative index $_{0, j}$	-0.2718***	-0.2866***	-0.2829***	-0.2807***
	(-7.94)	(-8.17)	(-8.18)	(-8.13)
Positive index $_{0, j}$	0.0975***	0.1035***	0.1014***	0.1015***
	(4.49)	(4.46)	(4.38)	(4.37)
AR $_{-1, j}$	-0.0582*	--	0.0197**	0.0211***
	(-1.93)	--	(2.50)	(2.67)
AR $_{-2, j}$	0.0702**	--	0.0048	0.0067
	(2.26)	--	(0.61)	(0.86)
AR $_{-3, j}$	0.0738**	--	-0.0011	-0.0002
	(2.14)	--	(-0.13)	(-0.03)
AR $_{-4, j}$	0.0622**	--	0.0050	0.0057
	(2.04)	--	(0.60)	(0.69)
AR $_{-5, j}$	0.0017	--	-0.0045	-0.0038
	(0.05)	--	(-0.54)	(-0.45)
Negative index $_{-1, j}$	-0.0572**	--	-0.0661**	-0.0629**
	(-2.44)	--	(-2.18)	(-2.12)
Negative index $_{-2, j}$	0.0702**	--	0.0670**	0.0708**
	(2.26)	--	(2.11)	(2.34)
Negative index $_{-3, j}$	0.0738**	--	0.0728**	0.0788**
	(2.14)	--	(2.02)	(2.22)
Negative index $_{-4, j}$	0.0622**	--	0.0624*	0.0670**
	(2.04)	--	(1.89)	(2.11)
Negative index $_{-5, j}$	0.0017	--	0.0149	0.0194
	(0.05)	--	(0.48)	(0.61)
Positive index $_{-1, j}$	-0.0572**	--	-0.0408*	-0.0368
	(-2.44)	--	(-1.71)	(-1.51)
Positive index $_{-2, j}$	0.0120	--	0.0159	0.0216
	(0.67)	--	(0.81)	(1.12)
Positive index $_{-3, j}$	-0.0039	--	0.0006	0.0055
	(-0.18)	--	(0.03)	(0.24)
Positive index $_{-4, j}$	-0.0077	--	0.0009	0.0039
	(-0.40)	--	(0.05)	(0.20)
Positive index $_{-5, j}$	0.0087	--	0.0194	0.0233
	(0.42)	--	(0.86)	(1.02)
Constant	0.0006	-0.1682***	-0.1722***	0.0437
	(0.04)	(-6.49)	(-6.76)	(1.03)
Adj. R-sq	0.0048	0.0069	0.0074	0.0058

Panel B: Tone indices weighted by news providers				
Negative-Press index $_{0, j}$	-0.2836***	-0.2951***	-0.2918***	-0.2885***
	(-9.56)	(-9.31)	(-9.29)	(-9.22)
Positive-Press index $_{0, j}$	0.1047***	0.1069***	0.1070***	0.1089***
	(8.00)	(7.43)	(7.52)	(7.55)
Adj. R-sq	0.0087	0.0111	0.0114	0.0099

Table VI. (Continued)

Dependent variable	AR $_{0, j}$			
Models	(1)	(2)	(3)	(4)
Panel C: Tone and topic indexes				
Business-Negative index $_{0, j}$	-0.1657***	-0.1725***	-0.1692***	-0.1647***
	(-3.82)	(-3.73)	(-3.67)	(-3.61)
Business-Positive index $_{0, j}$	0.0380*	0.0407*	0.0393*	0.0422*
	(1.67)	(1.80)	(1.71)	(1.85)
Finance-Negative index $_{0, j}$	-0.3299***	-0.3488***	-0.3444***	-0.3405***
	(-5.85)	(-6.15)	(-6.06)	(-6.21)
Finance-Positive index $_{0, j}$	0.1802***	0.1922***	0.1894***	0.1879***
	(5.31)	(5.22)	(5.21)	(5.22)
Mixed-Negative index $_{0, j}$	-0.3238***	-0.3259***	-0.3217***	-0.3244***
	(-4.52)	(-4.37)	(-4.34)	(-4.35)
Mixed-Positive index $_{0, j}$	0.0991***	0.1084***	0.1054***	0.1053***
	(2.70)	(2.78)	(2.76)	(2.82)
Adj. R-sq	0.0040	0.0061	0.0066	0.0050
Panel D: Media optimism index				
Media-Optimism index $_{0, j}$	0.1601***	0.1703***	0.1677***	0.1663***
	(10.77)	(10.20)	(10.14)	(9.91)
Adj. R-sq	0.0041	0.0064	0.0067	0.0051
Panel E: Media optimism index and topics of news headlines indexes				
Business-Optimism index $_{0, j}$	0.0791***	0.0845***	0.0824***	0.0829***
	(4.79)	(4.65)	(4.46)	(4.49)
Finance-Optimism index $_{0, j}$	0.2284***	0.2418***	0.2388***	0.2363***
	(8.50)	(8.27)	(8.22)	(8.33)
Mixed-Optimism index $_{0, j}$	0.1571***	0.1677***	0.1653***	0.1646***
	(4.75)	(4.67)	(4.67)	(4.72)
Adj. R-sq	0.0036	0.0058	0.0061	0.0046
L(5) (AR $_{0, j}$)	Yes	No	Yes	Yes
L(5) (sentiment index $_{0, j}$)	Yes	No	Yes	Yes
Year FE	No	Yes	Yes	Yes
Firm FE	No	Yes	Yes	No
Industry FE	No	No	No	Yes
Market Cap. FE	No	Yes	Yes	Yes
Observations	66,877	63,428	63,428	63,428

Table VII. Validation results using out-of-sample dataset and *abnormal volume*

This table reports the results from cross-sectional regression analyses to examine the links between the tone and topic of news headlines and contemporaneous abnormal Volume. The dependent variable is Abnormal Volume on day 0 for firm *j* ($AV_{0,j}$), the day of publishing headlines. Definitions of the sentiment indexes are provided in the section 5.1 Variable Constructions. There are 66,977 firm-day observations using the Close-to-Close window. The sample size drops to 66,877 when we include AV and to 63,428 when we consider market capitalization. The results report in Panel A focus on tone indexes, Panel B results focus on tone index weight adjusted by the number of news providers, and Panel B results utilize the tone and the topic of news headlines indexes. The results are reported for four models based on different combinations of lagged AV, lagged sentiment index, year fixed effects, firm fixed effects, industry fixed effects using 4-digit SIC code, and annual market capitalization fixed effects using quartiles. We obtain daily AV from event study tests using the Fama-French-Momentum model, Abnormal Relative Volumes, and Log-Transformed Value-Weighted Volume Index. The volume Event Study is conducted only for stocks in Nasdaq and NYSE). All market variables are winsorized at 1% and 99% by year. The robust error term is clustered for industry 4-digit SIC code. t-statistics are in parentheses and *, **, *** represent significance levels of 0.1, 0.05, 0.01, respectively.

Dependent variable	$AV_{0, j}$			
Models	(1)	(2)	(3)	(4)
Panel A: Tone indexes				
Negative index $_{0, j}$	3.9743***	5.5633***	3.9204***	4.1001***
	(6.01)	(5.85)	(6.88)	(7.09)
Positive index $_{0, j}$	1.8161***	0.2586	1.4647***	1.5916***
	(2.98)	(0.37)	(2.62)	(2.78)
$AV_{-1, j}$	0.5865***		0.5696***	0.5739***
	(48.15)		(46.73)	(46.24)
$AV_{-2, j}$	0.0505***		0.0437***	0.0450***
	(4.75)		(4.17)	(4.32)
$AV_{-3, j}$	0.1166***		0.1086***	0.1103***
	(19.23)		(17.79)	(18.24)
$AV_{-4, j}$	-0.0980		-0.0578	-0.0532
	(-0.55)		(-0.29)	(-0.27)
$AV_{-5, j}$	-0.0582		-0.0245	-0.0235
	(-0.47)		(-0.19)	(-0.18)
Negative index $_{-1, j}$	-4.5549***		-2.8685***	-3.1662***
	(-7.20)		(-4.60)	(-5.11)
Negative index $_{-2, j}$	-1.3180*		0.2038	-0.0983
	(-1.77)		(0.26)	(-0.12)
Negative index $_{-3, j}$	-1.8982***		-0.4006	-0.6252
	(-2.93)		(-0.59)	(-0.92)
Negative index $_{-4, j}$	-1.5070**		0.0824	-0.1484
	(-2.09)		(0.11)	(-0.20)
Negative index $_{-5, j}$	-1.2269*		0.6427	0.3926
	(-1.93)		(1.08)	(0.66)
Positive index $_{-1, j}$	-3.0455***		-1.4868***	-1.8380***
	(-5.13)		(-2.65)	(-3.21)
Positive index $_{-2, j}$	-1.6826***		-0.2486	-0.5886
	(-4.49)		(-0.64)	(-1.49)
Positive index $_{-3, j}$	-1.7243***		-0.2593	-0.5757*
	(-4.91)		(-0.82)	(-1.82)
Positive index $_{-4, j}$	-2.1473***		-0.6844	-0.9970**
	(-5.12)		(-1.50)	(-2.18)
Positive index $_{-5, j}$	-2.4175***		-0.7905	-1.1365**
	(-4.78)		(-1.60)	(-2.25)
Constant	5.8798***	-3.5321**	-3.7927***	6.9725***
	(9.59)	(-2.17)	(-5.96)	(5.30)
Adj. R-sq	0.4351	0.0706	0.4424	0.4397
Panel B: Tone indexes weighted by news providers				
Negative-Press index $_{0, j}$	7.8922***	12.7256***	8.3627***	8.3538***
	(11.23)	(11.39)	(12.54)	(12.22)
Positive-Press index $_{0, j}$	4.7905***	7.0253***	5.0222***	4.9839***
	(12.46)	(11.98)	(13.77)	(13.04)
Adj. R-sq	0.4425	0.0857	0.4500	0.4473

Table VII. (Continued)

Dependent variable	$AV_{0,j}$			
Models	(1)	(2)	(3)	(4)
Panel C: Tone and topic indexes				
Business-Negative index $_{0,j}$	0.8433	1.8642	1.4117	1.3872
	(1.00)	(1.31)	(1.61)	(1.58)
Business-Positive index $_{0,j}$	-1.6199**	-6.6471***	-2.3506***	-2.1686***
	(-2.39)	(-8.29)	(-3.75)	(-3.43)
Finance-Negative index $_{0,j}$	7.8143***	11.8204***	7.4289***	7.7892***
	(6.15)	(6.80)	(6.38)	(6.71)
Finance-Positive index $_{0,j}$	3.8840***	5.5588***	3.8925***	4.0365***
	(4.64)	(5.17)	(4.88)	(4.77)
Mixed-Negative index $_{0,j}$	4.6511***	5.3743***	4.6012***	4.5593***
	(2.91)	(2.67)	(2.83)	(2.82)
Mixed-Positive index $_{0,j}$	1.2876	0.8844	1.0282	0.9970
	(1.53)	(0.83)	(1.24)	(1.24)
Adj. R-sq	0.4393	0.0744	0.4470	0.4443
L(5) ($AV_{0,j}$)	Yes	No	Yes	Yes
L(5) (sentiment index $_{0,j}$)	Yes	No	Yes	Yes
Year FE	No	Yes	Yes	Yes
Firm FE	No	Yes	Yes	No
Industry FE	No	No	No	Yes
Market Cap. FE	No	Yes	Yes	Yes
Observations	66,877	63,428	63,428	63,428

Table VIII. Reversal and delay market response results using Out-of-Sample dataset and Cumulative Abnormal Returns (CAR)

This table reports the results from cross-sectional regression analyses to examine the potential delayed market response or reversal in the initial change to market returns using Equation 5. The dependent variable is percentage $AR_{0,j}$, $AR_{1,j}$, $AR_{2,j}$, and Cumulative Abnormal Returns, CAR_j *(0,1) and* CAR_j *(2,5)* on firm j on trading days 0, 1, 2, 0 to 1, and 2 to 5 form the day of publishing headlines day 0. The definitions of the tone indexes are provided in Internet Appendix IA.C. There are 66,977 firm-day observations using the Close-to-Close window. The sample size drops to 66,877 when we consider AR, and to 63,428 when we include market capitalization. The results reported in Panel A focus on tone indexes and Panel B results utilize the tone and the topic of news headlines indexes. The results are reported for five models, the dependent variable in model 1 is AR on day 0, in model 2 is AR on day 1, in Model 3 is CAR (0,1) window, in Model 4 is AR on day 2, and Model 5 is CAR(2, 5) window, all for firm j. In all models we control for five lags of AR $_{0,j}$ and Sentiment Index $_{0,j}$. All models control year fixed effects, firm fixed effects, and annual market capitalization fixed effects using quartiles. We obtain daily AR from event study tests using the Fama-French-Momentum model and value weighted index. All abnormal returns are winsorized at 1% and 99% by year. The robust error term is clustered for industry 4-digit SIC code. t-statistics are in parentheses and *, **, *** represent significance levels of 0.1, 0.05, 0.01, respectively.

Dependent variable	AR $_{0,j}$	AR $_{1,j}$	CAR_j (0,1)	AR $_{2,j}$	CAR_j (2,5)
Models	(1)	(2)	(3)	(4)	(5)
Panel A: Tone indexes					
Negative index $_{0,j}$	-0.2829***	-0.0010	-0.2897***	0.0062	0.1452***
	(-8.18)	(-0.05)	(-7.41)	(0.27)	(3.22)
Positive index $_{0,j}$	0.1014***	0.0112	0.1146***	0.0094	0.0357
	(4.38)	(0.70)	(4.07)	(0.54)	(1.22)
Adj. R-sq	0.0074	0.0026	0.0073	0.0023	0.0071
Panel B: Tone and Topic Indexes					
Business-Negative index $_{0,j}$	-0.1692***	-0.0065	-0.1790***	-0.0292	0.0304
	(-3.67)	(-0.18)	(-3.07)	(-0.77)	(0.41)
Business-Positive index $_{0,j}$	0.0393*	0.0052	0.0452*	0.0208	0.0009
	(1.71)	(0.18)	(1.39)	(0.79)	(0.02)
Finance-Negative index $_{0,j}$	-0.3444***	-0.0225	-0.3737***	0.0134	0.2133**
	(-6.06)	(-0.53)	(-5.00)	(0.36)	(2.43)
Finance-Positive index $_{0,j}$	0.1894***	0.0641***	0.2572***	-0.0055	-0.0375
	(5.21)	(3.02)	(5.44)	(-0.23)	(-0.88)
Mixed-Negative index $_{0,j}$	-0.3217***	0.0264	-0.3017***	-0.0790	0.0548
	(-4.34)	(0.48)	(-2.99)	(-1.46)	(0.48)
Mixed-Positive index $_{0,j}$	0.1054***	0.0081	0.1156**	-0.0019	0.0217
	(2.76)	(0.26)	(2.47)	(-0.06)	(0.40)
Adj. R-sq	0.0066	0.0024	0.0071	0.0022	0.0069
L(5) (AR $_{0,j}$)	Yes	Yes	Yes	Yes	Yes
L(5) (sentiment index $_{0,j}$)	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes	Yes
Market Cap. FE	Yes	Yes	Yes	Yes	Yes
Observations	63,428	63,428	63,428	63,428	63,428

Table IX. Robustness Results using Out-of-Sample Dataset and Abnormal Returns from Close-to-Open Window

This table reports the results from cross-sectional regression analyses to examine the links between the tone and topic of news headlines published outside the trading session on the open stock prices using Equation 4. Open AR and Open sentiment index are used instead of Close-to-Close AR. We compute the return from the previous day close price to same day open price (i.e. $Open - Returns_{\,t} = \left(\frac{open\ price_{\,t}}{close\ price_{\,t-1}}\right) - 1$). We call this session Close-to-Open window. The dependent variable is the percentage of Abnormal Returns on day *0* for firm *j* using open returns (Open $AR_{0,j}$). The definitions of the open sentiment indexes are provided in section 5.2. There are 32,490 firm-day observations using Close-to-Open window. The sample size drops to 32,479 after considering Open AR and to 31,961 when we include all other variables. The results reported in Panel A and B focus on tone indexes and the tone and the topic of news headlines, respectively. The results are reported for four models using different combinations of lagged open AR, lagged open sentiment indexes, and different fixed effects controls. We obtain the percentage of daily open AR from event study tests using market model and value weighted S&P 500 index. All abnormal returns are winsorized at 1% and 99% by year. The robust error term is clustered for industry 4-digit SIC code. T-statistics are in parentheses and *, **, *** represent significance levels of 0.1, 0.05, 0.01, respectively.

Dependent variable	Open AR $_{0,\,j}$			
Models	(1)	(2)	(3)	(4)
Panel A: Tone index				
Open Negative index $_{0,\,j}$	-0.1231***	-0.1219***	-0.1237***	-0.1261***
	(-4.45)	(-4.50)	(-4.47)	(-4.60)
Open Positive index $_{0,\,j}$	0.0513***	0.0416***	0.0421***	0.0445***
	(3.21)	(2.82)	(2.78)	(2.88)
Adj. R-sq	0.0022	0.0018	0.0019	0.0013
Panel B: Tone and topic indexes				
Open Business-Negative index $_{0,\,j}$	-0.0425	-0.0347	-0.0367	-0.0387
	(-1.08)	(-0.88)	(-0.90)	(-0.97)
Open Business-Positive index $_{0,\,j}$	0.0410**	0.0382**	0.0381**	0.0405**
	(2.30)	(2.06)	(2.05)	(2.18)
Open Finance-Negative index $_{0,\,j}$	-0.3668***	-0.3670***	-0.3693***	-0.3681***
	(-3.48)	(-3.41)	(-3.39)	(-3.42)
Open Finance-Positive index $_{0,\,j}$	0.0673***	0.0616**	0.0632**	0.0630***
	(2.85)	(2.53)	(2.60)	(2.67)
Open Mixed-Negative index $_{0,\,j}$	0.0164	0.0340	0.0331	0.0317
	(0.22)	(0.43)	(0.42)	(0.40)
Open Mixed-Positive index $_{0,\,j}$	0.0103	0.0148	0.0139	0.0181
	(0.40)	(0.56)	(0.53)	(0.70)
Adj. R-sq	0.0030	0.0033	0.0029	0.0023
L(5) (Open AR $_{0,\,j}$)	Yes	No	Yes	Yes
L(5) (Open sentiment index $_{0,\,j}$)	Yes	No	Yes	Yes
Year FE	No	Yes	Yes	Yes
Firm FE	No	Yes	Yes	No
Industry FE	No	No	No	Yes
Market Cap. FE	No	Yes	Yes	Yes
Observations	32,479	31,961	31,961	31,961

Table X. Robustness results using Out-of-Sample dataset and abnormal returns from Open-to-Close window

This table reports the results from cross-sectional regression analyses to examine link between the tone and topic of news headlines published using Equation 4 and during the trading session (i.e. intraday) on the closing stock prices. In this session we compute the return from the same day open price to close price (i.e. $intraday\ return_{\ t} = \left(\frac{close\ price_{\ t}}{open\ price_{\ t}}\right) - 1$). We call this session Open-to-Close window. The dependent variable is the percentage of Abnormal Returns on day *0* for firm *j* using intraday returns *(Intraday $AR_{0,j}$)*.The definitions of the intraday sentiment indexes are provided in section 5.2. There are 45,620 firm-day observations using Open-to-Close window. The sample size drops to 45,606 after considering intraday AR and to 44,589 when we include other variables. The results reported in Panel A and B focus on tone indexes and tone and the topic of news headlines, respectively. The results reported for four models using different combinations of lagged intraday AR, lagged intraday sentiment indexes, and different fixed effects controls. We obtain the percentage of daily intraday AR from event study tests using market model and value weighted S&P 500 index. All abnormal returns are winsorized at 1% and 99% by year. The robust error term is clustered for industry 4-digit SIC code. t-statistics are in parentheses and *, **, *** represent significance levels of 0.1, 0.05, 0.01, respectively.

Dependent variable	Intraday AR $_{0,\ j}$			
Models	(1)	(2)	(3)	(4)
Panel A: Tone indexes				
Intraday Negative index $_{0,\ j}$	-0.1357***	-0.1343***	-0.1359***	-0.1279***
	(-3.57)	(-3.30)	(-3.32)	(-3.18)
Intraday Positive index $_{0,\ j}$	0.0573*	0.0641**	0.0663**	0.0651**
	(1.92)	(2.19)	(2.27)	(2.21)
Adj. R-sq	0.0033	0.0059	0.0080	0.0053
Panel B: Tone and topic indexes				
Intraday Business-Negative index $_{0,\ j}$	-0.1041**	-0.0903*	-0.0931*	-0.0922**
	(-2.23)	(-1.92)	(-1.98)	(-1.98)
Intraday Business-Positive index $_{0,\ j}$	-0.0163	-0.0153	-0.0102	-0.0126
	(-0.45)	(-0.44)	(-0.30)	(-0.38)
Intraday Finance-Negative index $_{0,\ j}$	-0.1523**	-0.1653**	-0.1654**	-0.1418**
	(-2.33)	(-2.57)	(-2.51)	(-2.11)
Intraday Finance-Positive index $_{0,\ j}$	0.1136**	0.1156***	0.1167***	0.1187***
	(2.60)	(2.97)	(2.99)	(2.89)
Intraday Mixed-Negative index $_{0,\ j}$	-0.2594***	-0.2451***	-0.2447***	-0.2516***
	(-2.96)	(-2.65)	(-2.69)	(-2.72)
Intraday Mixed-Positive index $_{0,\ j}$	0.0844**	0.1033**	0.1058**	0.0990**
	(2.25)	(2.33)	(2.40)	(2.23)
Adj. R-sq	0.0035	0.0060	0.0082	0.0056
L(5) (Intraday AR $_{0,\ j}$)	Yes	No	Yes	Yes
L(5) (Intraday sentiment index $_{0,\ j}$)	Yes	No	Yes	Yes
Year FE	No	Yes	Yes	Yes
Firm FE	No	Yes	Yes	No
Industry FE	No	No	No	Yes
Market Cap. FE	No	Yes	Yes	Yes
Observations	45,606	44,589	44,589	44,589

Table XI. Linear combination of coefficients of sentiment indexes across different regression models

This table presents the results from examining the linear combination of the coefficient of the sentiment indexes in the regression models reported in model A, Panel A of Tables VI, IX, and X. The results in Table VI focus on Close-to-Close, Table IX focus on close-to-open, and Table X discus open-to-close window. We examine this null hypothesis: $[coefficient\ of\ Open\ sentiment\ Index_{\ 0,j,Table\ IX}] + [coefficient\ of\ Intraday\ sentimentIndex_{\ 0,j,Table\ X}] - [coefficient\ of\ sentiment\ Index_{\ 0,j,Table\ VI}) = 0)$. The results are reported for *Negative index* and *Positive index* across 66,877 firm-day observations.

Tone	Coef.	Robust Std. Err.	z	P>|z|	[95% Conf. Interval]	
Negative index	0.0219	0.0388	0.56	0.573	[-0.0542,	0.0980]
Positive index	0.0081	0.0239	0.34	0.736	[-0.0387,	0.0548]

Table XII. Validation Robustness Tests Using a Post Period Sample

This table reports the results from cross-sectional regression analyses to examine the sensitivity of our validation results to the news headlines published between January and December 2019, immediately following the period covered by our main sample, i.e. 2014 to 2018. For the post period sample, we collect the news headlines published by the same 15 news providers for 125 firms randomly selected from the 221 firms in our sample of the company/sector dictionary, resulting in a total of 20,116 news headlines. The table shows the association between TASA tone signals and the market response using Equation 5. The dependent variable is percentage $AR_{0,j}$, $AR_{1,j}$, $AR_{2,j}$, and Cumulative Abnormal Returns, $CAR_j\ (0,1)$ and $CAR_j\ (2,5)$ on firm j on trading days 0, 1, 2, 0 to 1, and 2 to 5 form the day of publishing headlines day 0. The definitions of the tone indexes are provided in Internet Appendix IA.C. There are 6,591 firm-day observations using the Close-to-Close window. The sample size drops to 6,391 when we consider AR, and to 6,182 when we include market capitalization. The results reported in Panel A focus on tone indexes and Panel B results utilize the tone and the topic of news headlines indexes. The results are reported for five models, the dependent variable in model 1 is AR on day 0, in model 2 is AR on day 1, in Model 3 is CAR (0,1) window, in Model 4 is AR on day 2, and Model 5 is CAR(2, 5) window, all for firm j. In all models we control for five lags of $AR_{0,j}$ and Sentiment $\text{Index}_{0,j}$. All models control month fixed effects, firm fixed effects, and annual market capitalization fixed effects using quartiles. We obtain daily AR from event study tests using the Fama-French-Momentum model and value weighted index. All abnormal returns are winsorized at 1% and 99%. The robust error term is clustered for industry 4-digit SIC code. t-statistics are in parentheses and *, **, *** represent significance levels of 0.1, 0.05, 0.01, respectively.

Dependent variable	$AR_{0,j}$	$AR_{1,j}$	CAR_j (0,1)	$AR_{2,j}$	CAR_j (2,5)
Models	(1)	(2)	(3)	(4)	(5)
Panel A: Tone indexes					
Negative index $_{0,j}$	-0.3882***	0.1766**	-0.2141*	0.0462	0.4016**
	(-3.64)	(2.04)	(-1.84)	(0.35)	(2.09)
Positive index $_{0,j}$	0.1801*	0.0833	0.2647*	0.0016	0.1366
	(1.87)	(0.85)	(1.74)	(0.01)	(0.63)
Adj. R-sq	0.0207	0.0093	0.0304	0.0191	0.0514
Panel B: Tone and Topic Indexes					
Business-Negative index $_{0,j}$	-0.5956*	-0.0156	-0.6142*	-0.0087	0.6158
	(-1.84)	(-0.12)	(-1.76)	(-0.04)	(1.12)
Business-Positive index $_{0,j}$	-0.1286	0.1389	0.0095	0.1230	0.2717
	(-0.54)	(0.57)	(0.03)	(0.50)	(0.78)
Finance-Negative index $_{0,j}$	-0.1659**	0.4896***	0.3230	0.3418**	0.8154**
	(-2.61)	(3.90)	(1.03)	(2.15)	(2.12)
Finance-Positive index $_{0,j}$	0.2733***	0.0990	0.3742**	0.1407	0.1064
	(2.99)	(0.82)	(2.48)	(0.69)	(0.31)
Mixed-Negative index $_{0,j}$	-0.3743	0.1915	-0.1860	-0.2102	-0.1076
	(-1.63)	(1.39)	(-0.80)	(-1.38)	(-0.29)
Mixed-Positive index $_{0,j}$	0.2745	-0.0874	0.1883	-0.4196	-0.5485
	(0.82)	(-0.38)	(0.43)	(-1.66)	(-1.47)
Adj. R-sq	0.0202	0.0097	0.0304	0.0202	0.0519
L(5) ($AR_{0,j}$)	Yes	Yes	Yes	Yes	Yes
L(5) (sentiment index $_{0,j}$)	Yes	Yes	Yes	Yes	Yes
Month FE	Yes	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes	Yes
Market Cap. FE	Yes	Yes	Yes	Yes	Yes
Observations	6,182	6,182	6,182	6,182	6,182

Table XIII. Pearson's Correlation between tone assignments of sentiment models and contemporaneous abnormal returns using the test set (p-values are in parentheses, *p<.05)

This table reports the results from the Pearson correlation between tone assignments of a benchmark of sentiment models and contemporaneous abnormal returns. The benchmark of sentiment models includes the TASA, Hybrid Model 2, Automatic Model, and L&M BoW. $AR_{0,j}$ is Abnormal Returns on day 0 for firm *j* ($AR_{0,j}$), the day of publishing news headlines. There are 19,635 firm-day observations using the test set and Close-to-Close window. The test set and benchmark are explained in Section 4. The results reported in Panel A focus on Positive and Negative tone indexes and Panel B utilize Optimism tone Index. The definitions of all variables are provided in Internet Appendix IA.C. p-values are in parentheses and * represents significance level of 0.05.

Panel A: Tone indices

	L&M BoW negative index $_{0,j}$	L&M BoW neutral index $_{0,j}$	L&M BoW positive index $_{0,j}$	Automatic negative index $_{0,j}$	Automatic neutral index $_{0,j}$	Automatic positive index $_{0,j}$	Hybrid 2 negative index $_{0,j}$	Hybrid 2 neutral index $_{0,j}$	Hybrid 2 positive index $_{0,j}$	TASA negative index $_{0,j}$	TASA neutral index $_{0,j}$	TASA positive index $_{0,j}$
L&M BoW neutral index $_{0,j}$	**-0.7065***	**1**										
	(0)											
L&M BoW positive index $_{0,j}$	**-0.0054**	**-0.7038***	**1**									
	(0.4015)	**(0)**										
Automatic negative index $_{0,j}$	**0.0343***	**-0.0442***	**0.0280***	**1**								
	(0)	**(0)**	**(0)**									
Automatic neutral index $_{0,j}$	**-0.0507***	**0.0851***	**-0.0693***	**-0.5276***	**1**							
	(0)	**(0)**	**(0)**	**(0)**								
Automatic positive index $_{0,j}$	**0.0187***	**-0.0448***	**0.0446***	**-0.4528***	**-0.5185***	**1**						
	(0.0039)	**(0)**	**(0)**	**(0)**	**(0)**							
Hybrid 2 negative index $_{0,j}$	**0.1296***	**-0.0701***	**-0.0310***	**0.1168***	**-0.1458***	**0.0355***	**1**					
	(0)	**(0)**	**(0)**	**(0)**	**(0)**	**(0)**						
Hybrid 2 neutral index $_{0,j}$	**-0.0592***	**0.0936***	**-0.0728***	**-0.2081***	**0.4072***	**-0.2179***	**-0.3318***	**1**				
	(0)	**(0)**	**(0)**	**(0)**	**(0)**	**(0)**	**(0)**					
Hybrid 2 positive index $_{0,j}$	**-0.0734***	**-0.009**	**0.0864***	**0.0564***	**-0.1870***	**0.1395***	**-0.6651***	**-0.4837***	**1**			
	(0)	**(0.1631)**	**(0)**	**(0)**	**(0)**	**(0)**	**(0)**	**(0)**				
TASA negative index $_{0,j}$	**0.1429***	**-0.0577***	**-0.0618***	**0.0204***	**-0.0197***	**0.0002**	**0.2979***	**-0.0514***	**-0.2357***	**1**		
	(0)	**(0)**	**(0)**	**(0.0016)**	**(0.0023)**	**(0.9814)**	**(0)**	**(0)**	**(0)**			
TASA neutral index $_{0,j}$	**0.0118**	**0.0106**	**-0.0269***	**-0.0545***	**0.1396***	**-0.0916***	**-0.0416***	**0.2535***	**-0.1621***	**-0.2569***	**1**	
	(0.0674)	**(0.0999)**	**(0)**	**(0)**	**(0)**	**(0)**	**(0)**	**(0)**	**(0)**	**(0)**		
TASA positive index $_{0,j}$	**-0.1209***	**0.0356***	**0.0711***	**0.0311***	**-0.1050***	**0.0788***	**-0.1950***	**-0.1786***	**0.3223***	**-0.5537***	**-0.6625***	**1**
	(0)	**(0)**	**(0)**	**(0)**	**(0)**	**(0)**	**(0)**	**(0)**	**(0)**	**(0)**	**(0)**	
AR 0, j	**-0.0193***	**0.0105**	**0.0129**	**-0.0167***	**-0.0037**	**0.0208***	**-0.0242***	**-0.0016**	**0.0237***	**-0.0342***	**0.0005**	**0.0262***
	(0.009)	**(0.1148)**	**(0.0535)**	**(0.012)**	**(0.575)**	**(0.0018)**	**(0.0003)**	**(0.8082)**	**(0.0004)**	**(0)**	**(0.9448)**	**(0.0001)**

Panel B: Media optimism index

	L&M BoW optimism index $_{0,j}$	Automatic optimism index $_{0,j}$	Hybrid 2 optimism index $_{0,j}$	TASA optimism index $_{0,j}$
Automatic optimism index $_{0,j}$	0.0133*	1		
	(0.0399)			
Hybrid 2 optimism index $_{0,j}$	0.1238*	0.0528*	1	
	(0)	(0)		
TASA optimism index $_{0,j}$	0.1578*	0.0236*	0.3263*	1
	(0)	(0.0003)	(0)	
AR $_{0,j}$	**0.0184***	**0.0220***	**0.0263***	**0.0335***
	(0.013)	**(0.0009)**	**(0.0001)**	**(0)**

Table XIV. Comparison between TASA and other sentiment models in explaining contemporaneous AR using the test set

This table reports the results from joint cross-sectional regression analyses to compare the explanatory power of the TASA approach with a benchmark of sentiment models in explaining contemporaneous abnormal returns using Equation 4. The benchmark of sentiment models includes the Hybrid Model 2, Automatic Model, and L&M BoW. The dependent variable is Abnormal Returns on day 0 for firm *j* ($AR_{0,j}$), the day of publishing news headlines. There are 19,635 firm-day observations using the test set and Close-to-Close window. The test set and benchmark are explained in Section 4. The results reported in Panel A focus on Positive and Negative tone indexes and Panel B utilize Optimism tone Index. The definitions of all variables are provided in Internet Appendix IA.C. The results are reported for eight models. We control for (5) lags of $AR_{0,j}$ and sentiment $index_{0,j}$. Additionally, we control for year fixed effects, firm fixed effects, and annual market capitalization fixed effects using quartiles. We obtain daily AR from event study tests using the Fama-French-Momentum model and value weighted index. All abnormal returns are winsorized at 1% and 99% by year. The robust error term is clustered for industry 4-digit SIC code. t-statistics are in parentheses and *, **, *** represent significance levels of 0.1, 0.05, 0.01, respectively.

Dependent Variable				**AR $_{0, j}$**				
Model	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: Tone indices								
AR $_{-1, j}$	.008	.008	.007	.007	.008	.008	.007	.008
	(.729)	(.779)	(.704)	(.721)	(.812)	(.796)	(.712)	(.762)
AR $_{-2, j}$	-.012	-.012	-.012	-.012	-.011	-.011	-.012	-.011
	(-1.306)	(-1.262)	(-1.3)	(-1.311)	(-1.246)	(-1.207)	(-1.293)	(-1.246)
AR $_{-3, j}$	-.019*	-.018*	-.019*	-.019*	-.018*	-.018*	-.018*	-.018*
	(-1.768)	(-1.732)	(-1.759)	(-1.773)	(-1.699)	(-1.703)	(-1.753)	(-1.716)
AR $_{-4, j}$	-.006	-.006	-.006	-.006	-.006	-.006	-.006	-.006
	(-.751)	(-.683)	(-.735)	(-.728)	(-.705)	(-.712)	(-.76)	(-.689)
AR $_{-5, j}$	.006	.006	.005	.006	.006	.006	.005	.006
	(.522)	(.544)	(.505)	(.53)	(.519)	(.527)	(.496)	(.519)
L&M BoW negative index $_{0, j}$	-.015	-.014		-.012	-.037*			
	(-1.542)	(-1.464)		(-1.332)	(-1.823)			
L&M BoW positive index $_{0, j}$	.029	.024		.021	.028			
	(.39)	(.357)		(.334)	(.469)			
Automatic negative index $_{0, j}$	-.025*	-.026*		-.026*		-.040*		
	(-1.603)	(-1.655)		(-1.694)		(-2.002)		
Automatic positive index $_{0, j}$	.036	.035*		.033		.039**		
	(1.624)	(1.764)		(1.526)		(2.027)		
Hybrid 2 negative index $_{0, j}$	-.036*		-.025*	-.024*			-.047*	
	(-1.383)		(-.924)	(-.806)			(-2.093)	
Hybrid 2 positive index $_{0, j}$	.041*		.029	.026			.045**	
	(1.778)		(1.13)	(.958)			(2.171)	
TASA negative index $_{0, j}$		-.066**	-.058*	-.056*				-.069**
		(-2.281)	(-1.826)	(-1.764)				(-2.394)
TASA positive index $_{0, j}$		.047*	.041*	.039*				.05**
		(1.851)	(1.423)	(1.346)				(2.028)
R-squared	.006	.007	.006	.007	.004	.005	.006	.006
Panel B: Media Optimism index								
L&M BoW optimism index $_{0, j}$	.026	.025		.021	.027			
	(1.296)	(1.389)		(1.195)	(1.631)			
Automatic optimism index $_{0, j}$	.03***	.03***		.03**		.032***		
	(2.711)	(2.725)		(2.666)		(2.844)		
Hybrid 2 optimism index $_{0, j}$		.039***	.027**	.025**				.041***
		(3.535)	(2.245)	(2.122)				(3.693)
TASA optimism index $_{0, j}$	.056***		.048***	.046***			.058***	
	(6.172)		(4.614)	(4.539)			(6.171)	
R-squared	.007	.006	.006	.007	.005	.005	.006	.006
L(5) (AR $_{0, j}$)	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Market Cap. FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	19,635	19,635	19,635	19,635	19,635	19,635	19,635	19,635