Topic Tones of Analyst Reports and Stock Returns: A Deep Learning Approach*

Hitoshi Iwasaki[†] Ying Chen[‡] Qianqian Du[‡] Jun Tu[§]

Abstract

We disassemble the text of analyst reports into multiple pieces that represent three types of topics - opinion type (overall assessments of analysts), corporate fact type (helping investors to better digest corporate facts) and justification type (explaining quantitative numbers). We extract the tone for the text of each topic via a deep neural network supervised learning methodology. A baseline model without using text information has an adjusted R^2 of 2.3% in predicting the cumulative two-day abnormal returns. When we include the topic tones, the adjusted R^2 increases to 15.4%. This significant increase of R^2 is not much driven by the justification type topics since it is sort of redundant given the quantitative numbers (e.g., earnings forecasts). In contrast, the opinion type and the corporate fact type topics provide substantial information beyond the quantitative numbers and are the main drivers of the significant increase of R^2.

JEL Classification: C89, G11, G12, G14

Keywords: Textual Analysis; DNN Approach; Topic Tones; Information Content

**Declarations:* Funding: Singapore Ministry of Education Academic Research Fund Tier 1 at National University of Singapore and at Singapore Management University.

[†]Hitoshi Iwaski, PhD candidate, Department of Statistics & Applied Probability, National University of Singapore, 117546. Phone (65) 6516 4416. Email iwasaki@u.nus.edu.

[‡]Ying Chen, Associate Professor of Statistics, Department of Mathematics, National University of Singapore, 119076. Phone (65) 6601 3976. Email matcheny@nus.edu.sg.

[‡]Qianqian DU, Associate Professor of Finance, School of RIEM, Southwestern University of Finance and Economics, 611130. Phone (86) 87081672. Email duqq@swufe.edu.cn.

[§]Jun Tu,Associate Professor of Finance, Lee Kong Chian School of Business, Singapore Management University, Singapore, 178899. Phone: (65) 6828-0764, E-mail: tujun@smu.edu.sg.

1 Introduction

Financial analysts are widely considered to be crucial financial intermediaries in the capital market. Extensive research efforts have made to understand the analyst's information role based on quantitative forecasts (including analysts' stock recommendations, target prices and earnings forecasts) in the analyst reports (e.g., Abdel-Khalik and Ajinkya, 1982; Womack, 1996; Barber et al., 2010; Alon and Reuven, 2003; Bradley et al., 2014; Kothari et al., 2016). Besides the quantitative information, analyst reports also provide detailed qualitative analyses of the firm, which form the main body of analyst reports and cover a board range of topics, including the firm's current and future accounting performance, corporate governance, competition landscape, risk and macroeconomic condition, and so on. Recent studies emphasize the valuable role of the qualitative information by exploring the aggregate tone of analyst reports.[1] However, since the text in analyst reports can be complex and related to various different topics, we will not be able to know which topics are important and which topics do not matter much if the investigation is limited to the aggregate tone of the text. Moreover, the overall textual tone can be blurred by less important topics.[2]

In this study, we disassemble the potentially complex text of an analyst's report into multiple pieces corresponding to various topics. There is a hierarchy of topics for the text in analyst's reports. First, it is common for analysts to provide supportive arguments and justifications for their numerical forecasts in analyst reports, including stock

[1]Previous studies have documented that the aggregate textual tone is a crucial contributor for stock returns drift; for example, see the studies of corporate disclosures (Li, 2010), analyst reports (Huang et al., 2014), news articles (see Tetlock, 2007; Zhang et al., 2016), conference calls (Matsumoto et al., 2011) and Twitter (Sul et al., 2017).

[2]For instance, assume that there are two topics composing the whole text, with all of the valuable information in one topic and no information in the other. If the non-informative topic occupies a significant proportion of the overall text, then the aggregate tone will be dominated by the tone of the non-informative topic. If the tone of the non-informative topic is significantly different from the tone of the informative topic, the aggregate tone can be blurred by the tone of the non-informative topic and be misleading.

recommendation ratings, target prices and earnings forecasts. This part of the text may convey further information related to quantitative forecasts and hence it is defined as corresponding to the *Justification* topic (associated with the numerical forecasts). The rest of the text contains broad perspectives from analyst's subjective insights, and also includes their outlook to objective news and information about the corporate events, which we define as *Qualitative* topic and split into the text for the *Opinion* topic and the text for the *Corporate Fact* topic[3]. We further split these three topics into 10 more detailed topics, namely three *Justification* type topics: recommendation rating, target price, earnings forecast; three *Opinion* type topics: conclusion, general argument, discussion on risk; and, four *Corporate Fact* type topics: profit items, non-profit items, official guidance, and corporate action.

For the text of a given topic, we gauge its information content using the topic tone, which is constructed using an innovative deep neural network (DNN) supervised learning method. For each classified topic, the DNN approach processes every word in a sentence sequentially, it preserves the word order, and it then estimates (or learns) the parameters and hyperparameters based on a labeled training corpus. The state-of-the-art DNN learning approach has the strong advantage of extracting text tones at the sentence level. In contrast, the traditional Natural Language Processing (NLP) methods construct text tones at the word level. For instance, the two NLP methods that are widely used in finance and accounting studies are: 1) the standard dictionary method, which uses the Loughran and McDonald (2011) financial and accounting dictionaries; and 2) the bag-of-words method, which forms a document-term matrix to record the occurrence of words (e.g., the Naïve Bayes approach employed in Huang et al., 2014). Nevertheless, both of these methods break sentences down into individual words and they ignore the order of words. Therefore, they will produce the same tone for two sentences with the same set

[3] *Justification* topic refers to text explaining the quantitative numbers and *Qualitative* topic refers to text of analyst' opinions and corporate facts that do not explain quantitative numbers.

of words, no matter if the words are in the same order or not. For example, the following two sentences will have the same value of text tone based on methods at the word level:

"The sales exceed costs."

vs. *"The costs exceed sales."*

Although the two sentences express completely contradictory economic meanings, textual analyses based on the word level cannot differentiate them. Consequently, to better extract the information, it is necessary to build up a learning approach that is based on the sentence level rather than the word level.

In our study, the topic tone analysis is performed on a novel dataset containing 113,043 analyst reports in Japanese for the Japanese market, which is the third largest stock market in the world in terms of market capitalization.[4] We employ a supervised learning methodology using a cutting edge DNN to extract tone in each detailed topic to obtain topic tones. This new approach shows superior power in predicting the cumulative two-day abnormal returns starting from the analyst report issue date. In the main sample test, the value of R^2 improves from 2.3% (baseline model without using any information in the analyst report text) to 11.6% (including an overall aggregate DNN tone) and further to 15.4% (including multiple topic tones).

When we analyze the impact of the disassembled topic tones on stock market reaction, we find that the tone of *Justification* topics is less informative because it simply revisits and justifies the numerical information that is already reported in a summary table in an analyst report. In contrast, *Opinion* and *Corporate Fact* topics carry more information. We further show that the analyst's information value comes not only from their subjective analyses and discussion but also from helping investors to better understand and incorporate public-available corporate news and events.

[4]The Japanese language does not have clear morpheme boundaries, which makes tokenization quite difficult in the pre-processing for NLP. Therefore, we do not rely on ready-made financial lexicon or sentiment analyzers that are generally available only for English. Our pre-processing and learning process could be used in many other textual analyses, even for non-English languages.

We also investigate the informativeness of topic tones at firm-level analyses and we find significant market reaction to the topic tones, which is consistent with results at analyst report level. We also document that market reaction to topic tones is more pronounced for the smaller firms, where the information environment might be less transparent, and the firms with more growth opportunities that are measured by lower book to market (BTM). Furthermore, the results of the cumulative two-day abnormal returns at the portfolio analyses corroborate the strong informativeness of the topic tone.

We conduct an out-of-sample test to gauge the accuracy of the NLP models that are based on DNN, in which the parameters are learned from the training sample and frozen through the out-of-sample test. Our out-of-sample validation test is different from those conducted in the previous literature (see Li, 2010; Huang et al., 2014), where K-fold cross validations are used to test the "validation" performance[5]. In our out-of-sample stock returns prediction, without any hyperparameters updates, we find that the model incorporating topic tones delivers a significant improved adjusted R^2 to 12.9%.

Our research contributes to the long-standing literature on the analyst's information role by initiating supervised topic analysis. Despite the importance of these topics, most of the previous literature has focused on the aggregate textual tone (e.g., Tetlock et al., 2008; Li, 2010; Huang et al., 2014). Our investigation on a finer classification within labeled topics will not only deepen the understanding of the value of texts but it will also provide guidance to investors about how to quickly pinpoint the key aspects of professional advice. Our supervised topic analysis is also related to some recent studies (see Lowry et al., 2020; Huang et al., 2018) that use the LDA method to conduct unsupervised topic analysis. However, the LDA way does not require an annotated training corpus and it

[5]Li (2010) and Huang et al. (2014) conduct the K-fold validation test to illustrate the "out-of-sample" performance. Given that the sub-samples of the k-fold used for training and validation are portions of the whole sample period, strictly speaking, K-fold test inevitably induces a looking-ahead bias; for example, when the training sub-samples are from the later period. We conduct an out-of-sample test where the training sample contains historical data and the testing sample is from different periods

has no explicit topic labels. In the absence of labels, researchers need to use their own knowledge and intelligence to interpret the learned topics by looking at frequent words. Therefore, there is no guarantee that those topics are in line with the research interest.

Our study makes an important methodological contribution to the financing and accounting textual analysis literature by adopting the cutting-edge DNN supervised learning methodology. Because the supervised DNN approach is based on the sentence level analyses and provides an effective way to quantify the text information, it outperforms previous Natural Language Processing (NLP) algorithms that are based on the word level; especially the widely used Dictionary method (e.g., Loughran and McDonald, 2011; Jiang et al., 2019).

The rest of this paper is structured as follows. Section 2 presents the dataset of Japanese analyst's reports and the data preprocessing; Section 3 details the method of extracting topic tones from analyst reports; Section 4 implements the main empirical results; Section 5 conducts further analyses; Section 6 carries out the out-of-sample test; and Section 7 concludes.

2 Data and Pre-processing

We manually collect analyst reports of the firms listed on the Tokyo Stock Exchange and Osaka Exchange from January 2016 to June 2018. The reports are written in Japanese. We remove the analyst certificates and disclaimers during the language preprocessing, and we only consider reports associated with a single stock. Our final sample includes $113,043$ analyst reports. We match the analyst reports to the database Nikkei QUICK to obtain the analyst's numerical forecasts, and also the firm's accounting and return information. Where matched quantitative variables are lacking, we keep the reports in the DNN supervised learning but drop them for empirical tests. We split the analyst reports sample into the two parts: the main sample has 87,598 reports from January 2016 to December 2017, and the sample reserved for the out-of-sample test has 25,445 reports from January 2018 to June 2018.

The text of an analyst's report usually has three parts: the first part is the *Justification* topic, which provides supportive arguments and justifications corresponding to the numerical forecasts in the analyst report, and the (subjective) *Opinion* topic and (objective) *Corporate Fact* topic; the second part is the *Opinion* category, which focuses on the analyst's own views on certain issues or events; the third part is the *Corporate Fact* category, which deals more with the exposition and propagation of important news and information. We further disassemble each group into a few topics and this results in 10 detailed topic classifications, which are presented in Table 1.

[Insert Table 1 here]

As displayed in Table 1, the *Justfication* topic is split into three topics, labeled from Topic 1 to Topic 3, which correspond to the numerical measures on analyst rating, target prices and earnings forecasts, respectively. The *Opinion* topic is further categorized into Topic 4 to Topic 6, and the *Corporate Fact* topic is classified into Topic 7 to Topic 10. Topic 4 is related to a conclusive opinion that captures an analyst's overall impression. It is possible that an analyst may have a negative opinion in terms of Topic 4, while positive in Topic 1; namely recommendation.[6] Topic 5 is for more general discussion and argument. Topic 6 assesses the risks and uncertainties that are associated with the firm's business outlook. Topic 7 to Topic 10 are in the category of *Corporate Fact* type that are related to news and contain little interpretation or discussion. The description of actual profits is classified as Topic 7, while the information about the non-profit accounting items

[6]This happens, for example, when a new event on earnings announcement somewhat disappoints the analyst, yet they may still consider the stock price attractive and thus are confident with a buy rating. In such a case, we could see a positive tone for the sentence of Topic 1 and a negative tone for the sentence of Topic 4 in the analyst's report.

is labeled as Topic 8. Topic 9 consists of information on reports of official guidance and its revision, which are required to be released in the Japanese equity market. Information about the corporate actions, such as capital structure, repurchase and acquisition, is classified into Topic 10.

In our study, we customize the NLP algorithms to capture the properties of the Japanese language. Shift-JIS is used to encode the multi-byte texts of Japanese analyst's reports. Tokenization is conducted with MeCab (Kudo et al., 2004), which simultaneously performs the segmentation, part-of-speech tagging, and lemmatization. For the specific purpose of financial text analysis, we employ a customized dictionary in which some of the key financial terms are registered so that the tokenizer can recognize these specific words. For example, we pre-registered word 営業利益 (operating profit), which would otherwise be mistakenly broken into two separate words: 営業 (operating) and 利益 (profit) and downgrade the quality of language processing. Finally, stop words and symbols (such as brackets and parentheses) are filtered out. Each sentence is converted to a processed one-hot vector to represent the sequence of words. The details of the NLP are presented in Appendix A1.

3 Constructing the Topic Tones

In this section, we present the details of how to construct the topic tones using DNN supervised classifications. DNN supervised learning is performed at the sentence level, where topic and tone annotations are given to each sentence. Specifically, in topic classification, a sentence is classified into a certain topic among the 10 topic candidates. In tone classification, a sentence is assigned to a certain tone of positive, negative, or neutral. Each of the 10 topic tones of an analyst's report depends on the tones of the sentences categorized to the topic.

The quality of the topic tone depends on how to efficiently proceed the information extraction. Traditional document representation, known as bag-of-words, runs at the word level and creates a matrix to represent the frequency of words in the collection of documents. However, this representation cannot capture the order of the words, which may result in misclassification.[7] We perform the NLP algorithms at the sentence level and we learn the unknown parameters of the DNN classifiers with a training corpus that is composed of 3,000 sentences that are randomly chosen from the main sample.

Within the training corpus, each sentence is labeled with topics and a tone. In the training procedure, the unknown parameters of the classification models are estimated to maximize the topic and tone classification accuracy. After training, the parameters and hyperparameters are fixed without any update for both in-sample and out-of-sample analysis in the further study.

Instead of generating an overall tone to represent the aggregate information of an analyst's reports, we look for finer topic tones to reveal and evaluate the detailed relevance of various aspects of the texts. It is possible that one sentence contains texts for multiple topics. For example, the following sentence conveys information for three topics, namely Topic 1: Analyst rating, Topic 3: Earnings forecast, and Topic 9: News on official guidance:

"Given weaker than expected official guidance for FY2018, we lowered our earnings forecast but kept the buy rating unchanged based on our view that management was overly conservative about the competitive landscape."

We adopt the binary relevance method (Tsoumakas and Katakis, 2009) to transform the multi-label problem into separate binary problems. In particular, each sentence is processed 10 times binary classifications, sequentially for each of the 10 topics; as illustrated

[7]The limitation of the bag-of-words algorithm can be illustrated by the example in the introduction that cannot differentiate the following two expressions: "the sales exceed costs" vs."the costs exceed sales".

in Figure 1.

[Insert Figure 1 here]

In the classifications, we apply DNN algorithms, including n-grams Convolutional Neural Network (CNN)[8], Long Short Term Memory (LSTM) and bi-directional LSTM. We further implement the classifications using Naïve Bayes method to compare with the existing works on analyst reports, such as Huang et al. (2014). For each classification task, we choose the method that has the best performance based on the K-fold cross-validation accuracy. We report the detailed technical setup and numerical results in Appendix A2. Table A.2 shows that the algorithms of the DNN method consistently have better performance than Naïve Bayes approach. For example, in Topic 4, CNN (3-4-5) algorithm is selected for subsequent tasks that include topic and tone classifications because its cross validation accuracy is 0.958, which is higher than other three DNN algorithms and Naïve Bayes method. Table A.2 further compares the performance of the topic models that consist of 10 binary classifiers with the tone model, which is a single trinary classifier.

After the sentence-level topics and tones are evaluated, they are aggregated to form the topic tones at the report level. For topic $i = 1, \cdots, 10$, we calculate the topic tone at report level:

$$TT_i \quad = \quad Tone_i^{pos} - Tone_i^{neg}, \tag{1}$$

where $Tone_i^{pos}$ ($Tone_i^{neg}$) is computed as the percentage of positive (negative) sentences within a specific topic in a report. We report the distribution of our constructed topic tones for the main sample in Table 2.A. We find that Topic 1 has higher proportion of

[8]The numbers in the parentheses at CNN (n-gram) indicate the window sizes of the filters used in the architecture.

positive sentences than other topics, and Topic 6 has the highest proportion of negative sentences and lowest standard deviation among the topics. We also present the summary statistics of the main variables in Table 2.B. We find that 9 of 10 topic tones, which capture the difference between positive and negative sentences in each detailed topics, have positive mean value; this is consistent with the well-documented problem of the analyst's optimism.

[Insert Table 2 here]

4 Empirical Analyses

In this section, we investigate the information value of analyst reports by analyzing informativeness of DNN tones on the stock market reactions. First, we compare the performance of the overall DNN tone with the revisions of numerical forecasts and the tone based on the Naïve Bayes approach in predicting subsequent market returns. Furthermore, we study how the detailed topic tones influence following cumulative returns.

4.1 Informativeness of the Overall DNN Tones

Previous research has documented that the revisions of the numerical forecasts are informative to the market (e.g., Jegadeesh et al., 2004; Barber et al., 2001; Huang et al., 2014) and that the overall textual tone based on Naïve Bayes approach can significantly predict following returns (Huang et al., 2014). To examine whether the overall tone based on DNN approach has better performance in explaining following returns, we estimate the following multivariate regression[9]:

[9] We use the revisions of analyst's numerical forecasts rather than the level of numerical forecasts in the main tests because the literature documents that analyst numerical revisions are more informative than the level of forecasts (Jegadeesh et al., 2004). We obtain consistent results when we include both the level and the revisions of the analyst's numerical forecasts into the regression.

$$
\begin{aligned}
CAR\ &= \alpha_0 + \beta_1 Tone\_DNN + \beta_2 Tone\_NB + \gamma_1 REC\_Rev + \\
&\gamma_2 TP\_Rev + \gamma_3 EPS\_Rev + \sum\nolimits_j \delta_j Controls_j + \epsilon.
\end{aligned} \tag{2}
$$

CAR refers to the cumulative two-day abnormal returns starting from the analyst report issue date, where the abnormal returns are calculated as the difference between the raw stock return and the benchmark index return (i.e., TOPIX return). $Tone\_DNN$ is the overall tone extracted by DNN classifier and $Tone\_NB$ is the overall tone based on Naïve Bayes approach. REC_Rev is the analyst recommendation revision, measured as the current report's recommendation value minus the analyst's last recommendation value, where analysts' recommendations are converted into the numerical scale from one to five. TP_Rev represents the target price revision and EPS_Rev is the earnings forecast revision, defined similarly to REC_Rev. We include CAR_Prior, which is the cumulative 20-day abnormal return before the analyst report date, to control for a possible short-term return drift, and we further control book-to-market ratio (BTM) and logarithm of the market value of equity (Size) in Equation (2). Industry and year fixed-effects are controlled for all of the regressions in the paper, and the standard errors are estimated using a two-dimension cluster control at the firm and analyst's brokerage firm levels.

We report the estimation results for model (2) in Table 3. Specification (1) presents the result of the baseline regression, which only includes numerical information. We find that all the revisions of the numerical forecasts are significantly correlated with the following cumulative abnormal returns, with adjusted R^2 of 2.3%. In specification (2), we add $Tone\_NB$ into the baseline regression and find that adjusted R^2 is 7.3%. In specification (3), when we include $Tone\_DNN$ into the baseline regression, the estimated coefficient is 0.014 with t-statistics of 14.74, and the adjusted R^2 is 11.6%, much larger than 7.3%, the R^2 under $Tone\_NB$. In specification (4), including both $Tone\_DNN$ and $Tone\_NB$ into the baseline regression. Compared with the results in (2), we show that both the significance and coefficient magnitude of $Tone\_NB$ decrease dramatically after $Tone\_DNN$ is included into the regression (the loading has decreased by 70%). Moreover, by comparing specification (3) and (4), the adjusted R^2 only increases marginally from 11.6% to 11.8%, which indicates that limited incremental information is contributed by $Tone\_NB$, once $Tone\_DNN$ is controlled. These findings indicate that at aggregate tone level, the tone based on the supervised DNN learning approach brings more pronounced incremental predictability for following returns beyond the tone based on Naïve Bayes method or traditional numerical forecasts.

[Insert Table 3 here]

4.2 The Informativeness of Topic Tones

The aggregate textual tone conveys analysts' overall opinion. However, We are not be able to detect which topics are more essential and which topics do not play important roles if the investigation is limited to the aggregate textual tone analysis. Furthermore, the aggregate textual tone can be blurred by less important topics. To explore the full spectrum of analysts' information value, we investigate market reactions to the detailed topic tones. As described in Section 2, analyst reports have a hierarchy of information, and we construct 10 detailed topic tones and categorize them into different categories. We expect that market will react divergently to the different topic tones. For instance, *Qualitative* topics which provide incremental information to investors beyond the numerical forecasts might have stronger market reaction than *Justification* topics. We modify the model (2) by replacing the overall textual tones with detailed topic tones and different categories of tones. We then run the following multivariable regression:

$$\begin{aligned} CAR \ &= \alpha_0 + \sum\nolimits_i \beta_i DetailedTone_i + \gamma_1 REC\_Rev + \gamma_2 TP\_Rev + \\ &\gamma_3 EPS\_Rev + \sum\nolimits_j \delta_j Controls_j + \epsilon. \end{aligned} \tag{3}$$

$DetailedTone$ stands for TT or other categories of tone, such as $Tone\_Justif$, $Tone\_Quali$, $Tone\_Opn$ or $Tone\_Fact$. TT is the topic tone in each of the 10 detailed topics, which are constructed by disassembling the comprehensive textual information of analyst reports into different topics and applying DNN classifier to extract the tone within each topic. $Tone\_Justif$ ($Tone\_Quali$) is the aggregate topic tone from Topic 1 to Topic 3 (Topic 4 to Topic 10), while $Tone\_Opn$ ($Tone\_Fact$) is aggregate topic tone from Topic 4 to Topic 6 (Topic 7 to Topic 10). All the topic tones are standardized to have mean zero and unit variance for the convenience of comparing the magnitude of the estimated loading. REC_Rev, TP_Rev, EPS_Rev and control variables are defined as in Section 4.1.

In panel A of table 4, we investigate how investors react to different categories of topics by regressing the $Tone\_Justif$, $Tone\_Quali$, $Tone\_Opn$ and $Tone\_Fact$ separately with control variables. In specification (1), the estimated coefficient of $Tone\_Justif$ is 0.007 and the adjusted R^2 is 4.3%. In specification (2), including $Tone\_Quali$ into the baseline regression, the estimated coefficient of $Tone\_Quali$ is 0.015[10] and the adjusted R^2 (i.e., 13.6%) is surprising larger than the adjusted R^2 in specification (3) of table 3 (i.e, 11.6%), which includes the overall tone $Tone\_DNN$ into the baseline regression, indicating that the aggregate tone can be blurred by less important topics. In specification (3) and (4), $Tone\_Opn$ and $Tone\_Fact$ show comparable predictability for following returns, with the adjusted R^2 of 10.7% and 11.0% respectively.

We conduct the horse-racing tests between different topics in the panel B of the table 4. In specification (1), when we include $Tone\_Justif$ and $Tone\_Quali$ to the baseline the regression, the estimate coefficient of $Tone\_Justif$ has decreased sharply to 0.002 (compared to 0.007 in specification (1) of the panel A in table 4). In the specification (2),

[10]The results indicate that one standard deviation increase in the favorableness of the *Qualitative* opinion results in an additional two-day abnormal return of 150 basis points. Under a similar setting in Huang et al. (2014), the magnitude of economic significance of the tone when applying Naïve Bayes approach is 41 basis point and when employing Loughran and McDonald (2011) Dictionary is 33 basis point.

Tone_Quali is decomposed into *Tone_Opn* and *Tone_Fact*, and we find that both of their estimated coefficients are larger the coefficient of *Tone_Justif*. Meanwhile, the adjusted R^2 is 13.7%, only slightly larger than R^2 of 13.6% in specification (2) of Panel A in table 4 that only includes *Tone_Quali* into the baseline regression. The result indicates that the *Justfication* type topics are sort of redundant given the *Opinion* topics, *Corporate Fact* topics and the quantitative numbers, although according to table 2.A, the sentence of *Tone_Justif* that covers Topic 1 to 3 account for as much as 26.19% of the total number of sentences of analysts reports. Moreover, if the tone of this topic is different from the tone of the other topics, the overall tone can be blurred by the tone of these sort of redundant sentences as discussed in the above paragraph.

The regression result in the specification (3) of panel B provides a clear description of informativeness of the 10 detailed topics. The adjusted R^2 improves to 15.4%, which is significantly larger than using alternative tones, especially the aggregate NB tone (with R^2 of 7.3%) and the aggregate DNN tone (with R^2 of 11.6%). Furthermore, the results show that 8 of the 10 topic tones are significantly and positively correlated with following abnormal return. Topic 1 that discusses the rationale for the analyst numerical rating and Topic 6 that analyzes the risk and uncertainties are insignificant[11]. In addition, the estimated coefficient of TT_4 has larger magnitude than other topic tones, suggesting that the topic related to the analyst's conclusion is the most informative source to investors. Furthermore, the significant estimated coefficients of TT_7 to TT_{10} indicates the information related to the news and corporate events are still informative to the market, even though most of them are publicly available. The result is consistent with gradual information diffusion theory in Hong and Stein (1999), which claims that information diffuses

[11]According to the summary statistics in Table 2, Topic 1 has higher proportion of positive sentences than other topics, while Topic 6 has the highest proportion of negative sentences and lowest standard deviation among the topics. The relatively optimistic tone in Topic 1 may be driven by the analyst's incentive(e.g., Lin and McNichols (1998)), and the comparatively pessimistic and less variant tone in Topic 6 might be induced by the analyst's cliché about the risk and uncertainty, which are less informative to investors.

slowly among investors because of the investor's limited ability to incorporate news, and in our setting, the detailed topics in the analyst reports accelerate the information incorporating process differently. Overall, the findings suggest that analysts' information value comes not only from analysts' conclusive analyses and discussion but also from helping investors incorporate corporate news.

[Insert Table 4 here]

5 Further Analyses

5.1 Firm-level Analyses

The report-level analyses in Section 4 show the strong informativeness of the topic tones. In our main sample, 47.8% of the reports are single reports for a certain stock on a particular day, while the rest of the analyst reports are issued together with other reports for the same firm on the same day. As a robustness check, we investigate the informativeness of the topic tones at the firm level. Therefore, we conduct the following regression:

$$\begin{aligned} CAR_t = \alpha_0 + \sum_{i=1}^{10} \beta_i \frac{1}{m_t} \sum_{s=1}^{m_t} TT_{i,t,s} + \gamma_1 \frac{1}{m_t} \sum_{s=1}^{m_t} REC\_Rev_{t,s} + \gamma_2 \frac{1}{m_t} \sum_{s=1}^{m_t} TP\_Rev_{t,s} + \\ \gamma_3 \frac{1}{m_t} \sum_{s=1}^{m_t} EPS\_Rev_{t,s} + \sum\nolimits_j \delta_j Controls_{j,t} + \epsilon, \end{aligned} \tag{4}$$

where m_t denotes the number of reports for the same stock issued on the same day t, and $\mathrm{TT}_{i,t,s}$ refers to the tone of topic i in report s. REC_Rev, TP_Rev, EPS_Rev and control variables are constructed at the firm level.

The firm-level regression results are presented in Table 6. We find that the value of R^2

improves from 2.8% (baseline model with only numerical information) to 9.6% (baseline model with an overall aggregate DNN tone) and further to 15.4% (baseline models with multiple topic tones), advocating that detailed topic tones are more informative to the market than the overall textual tone which can be blurred by less important topics. In the specification (3), the estimate of the interaction of the $Tone\_DNN$ and size is -0.003 (with t-statistics of -13.38). This indicates that the market reaction to the aggregate DNN tone is more pronounced for smaller firms whose information environment is less transparent. Likewise, the interaction of $Tone\_DNN$ and BTM is also negative of -0.002 (with t-statistics of -2.55). BTM is widely used as a proxy of firm growth (e.g., Gaver and Gaver, 1993; Lang et al., 1996) and it is known to be highly correlated with subsequent realized growth (Kallapur and Trombley, 1999). The findings indicate that investors pay more attention to analyst aggregate tones, seeking for implication on the firms' business outlook, particularly for growth-oriented firms.

[Insert Table 5 here]

5.2 Topic Tones and Future Earnings Growth

Textual information contains fundamental analysis that is value-relevant for a firm's future growth (Abarbanell and Bushee, 1997). To examine the predictive power of the topic tones for future earnings growth, we conduct the following regression analysis:

$$\begin{aligned} Growth_{t+n} &= \alpha_0 + \beta_1 Tone\_DNN_t + \beta_2 Tone\_NB_t + +\gamma_1 REC\_Rev_t + \gamma_2 TP\_Rev_t + \\ &\gamma_3 EPS\_Rev_t + \sum\nolimits_j \delta_j Controls_{j,t} + \epsilon, \end{aligned} \tag{5}$$

where $Growth_{t+n}$ is estimated as the difference between operating income from year $t+n$ and year t, scaled by total assets in year t, and n equals 1, 2 and 3, respectively, in the specification (1), (2) and (3) of Table 7. We include ROA, Size and BTM in year t as control variables and add industry and year fixed-effects in the regression.

The regression results of model (6) are reported in Table 7. In specification (1) both the aggregate DNN topic tone and the overall tone based on Naïve Bayes approach significantly predict earnings growth for year t+1, while the coefficient magnitude of $Tone\_DNN_t$ is three times larger than that of $Tone\_NB_t$. In specification (2) and (3), the aggregate DNN topic tone continues to significantly predict future earnings growth the subsequent year 2 and year 3, while the overall tone based on Naïve Bayes approach is no longer significant indicating the predictability of $Tone\_NB_t$ attenuates in the long term. Overall, the results corroborate the significant informativeness of our DNN tone.

[Insert Table 6 here]

5.3 Portfolio Analyses

In this subsection, we investigate the CAR in the framework of portfolio analyses based on independent double sorting approach to see the incremental value of the topic tones. Based on the regression result of model (4), we find that TT_4, which is the topic related to the analyst's conclusion, is the most informative topic of the analyst reports. Hence, we explore the incremental information value of in panel A of Table 8. We sort analyst reports into terciles independently according to TT_4 and REC_Rev. For each change of recommendation level, CAR increases monotonically with TT_4 tercile rankings, i.e., more favorable the TT_4, higher the following return. The differences of CAR between high and low TT_4 terciles ranges significantly from 2.1% to 4.5%, suggesting that the single topic tone TT_4 provides incremental information beyond the numerical analyst's forecast.

In panel B of Table 8, to examine the incremental informativeness of the aggregate tone based on Naïve Bayes approach, we conduct similarly double sorting based on $Tone\_NB$ and REC_Rev. Although the CARs between high and low $Tone\_NB$ terciles are significant in all REC_Rev tercile ranks, the magnitudes of CARs only ranges from 0.8%

to 2.3% that are significantly smaller than the corresponding CARs in panel A based on TT_4 and REC_Rev (with P-value <0.01). The result suggests that single topic tone show stronger informativeness than the overall tone based on Naïve Bayes approach.

[Insert Table 7 here]

6 The Out-of-sample Test

The K-fold cross validation is usually used when sample size is small, where the data are split to training and validation to mimic the out-of-sample forecasting scenarios. In our study, we directly implement out-of-sample forecasting given that there are reasonable large training sample (42,818 reports) and validation sample (12,784 reports). Moreover, given that the training is only based on the past information only, it is practically more evident if the trained model is able to deliver high predictive power of stock returns.

First, we freeze the parameters of DNN algorithm learned from the main sample and we apply the frozen parameters to construct topic tones from January 2018 to June 2018. We test the predictive power of the topic tones derived from the frozen parameters using model (3) and we present the results in Table 9. In the specification (1), the out-of-sample adjusted R^2 of the model including aggregate DNN tone into baseline regression is 12.2%, while in the specification (2) the overall NB tone contributes limited incremental adjusted R^2 given the aggregate DNN tone. Meanwhile, the detailed topic model consistently generates higher out-of-sample adjusted R^2 (15.9%) than the aggregate DNN tone model, further supporting the stronger informativeness of the detailed topic tones.

[Insert Table 8 here]

Furthermore, we conduct the out-of-sample prediction in the sample from January 2018 to June 2018. We freeze the parameters learned from the main sample and we apply the fitted model to evaluate the topic tones and predict stock returns, without any parameter update. The forecasts based on model (3) are conducted as follows:

$$\begin{aligned} C\hat{A}R = \hat{\alpha}_0 + \sum_{i=1}^{10} \hat{\beta}_i \frac{1}{m_t} \sum_{s=1}^{m_t} TT_{i,t,s} + \hat{\gamma}_1 REC\_Rev + \\ \hat{\gamma}_2 TP\_Rev \ + \hat{\gamma}_3 EPS\_Rev + \sum\nolimits_j \hat{\delta}_j Controls_j + \epsilon, \end{aligned} \tag{6}$$

where $C\hat{A}R$ is the forecast stock return and the estimated coefficients are from model (4). The distribution of the forecast returns vs. the actual returns is displayed in Figure 2. In terms of accuracy, the out-of-sample adjusted R^2 is 12.9%, which is reasonably high for a stock returns prediction task. Both the metrics and the shape of the scatter plot indicate that the predictive model works well in the out-of-sample setting.

[Insert figure 2 here]

7 Conclusion

In this paper, we disassembled the potentially complex text of an analyst's report into multiple pieces corresponding to various topics according to a hierarchy structure. The text that conveys further information related to quantitative forecasts is defined to be corresponding to the *Justification* topic. The rest of the text contains broad perspectives from analyst's subjective insights and outlook to objective news and information about the corporate events, which we split into the text for the *Opinion* topic and the text for the *Corporate Fact* topic. We then further split the analyst's report text into 10 more detailed topics, namely three *Justification* type topics: recommendation rating, target price, earnings forecast; three *Opinion* type topics: conclusion, general argument, discussion on risk, and four *Corporate Fact* type topics: profit items, non-profit items, official guidance, and corporate action.

We apply a DNN supervised learning methodology to extract the tone for the text of each topic. Using a set of 113,043 analyst reports in Japanese, we find that a baseline model without topic tones has an adjusted R^2 of 2.3% in predicting the cumulative two-day abnormal returns starting from the analyst report issue date. However, the R^2 increases to 15.4% when the topic tones are also used. When we compare the impact of various topics, we find that the text related to the *Opinion* type topics convey extra information in explaining stock returns beyond the text related to the *Justification* type topics. Furthermore, the text related to the *Corporate Fact* type topics may contain information that can help the investors to better understand corporate news.

References

Abarbanell, J. S. and Bushee, B. J. (1997). Fundamental analysis, future earnings, and stock prices, *Journal of Accounting Research* **35**(1): 1–24.

Abdel-Khalik, A. R. and Ajinkya, B. B. (1982). Returns to informational advantages: The case of analysts' forecast revisions, *Accounting Review* **57**(4): 661–680.

Alon, B. and Reuven, L. (2003). An empirical analysis of analysts' target prices: Short-term informativeness and long-term dynamics, *Journal of Finance* **58**(5): 1933–1967.

Barber, B., Lehavy, R., McNichols, M. and Trueman, B. (2001). Can investors profit from the prophets? Security analyst recommendations and stock returns, *Journal of Finance* **56**(2): 531–563.

Barber, B. M., Lehavy, R. and Trueman, B. (2010). Ratings changes, ratings levels, and the predictive value of analysts' recommendations, *Financial Management* **39**(2): 533–553.

Bond, F., Baldwin, T., Inui, K., Ishizaki, S., Nakagawa, H. and Shimazu, A. (2016). *Readings in Japanese Natural Language Processing*, CSLI Publications, Stanford.

Bradley, D., Clarke, J., Lee, S. and Ornthanalai, C. (2014). Are analysts' recommendations informative? intraday evidence on the impact of time stamp delays, *Journal of Finance* **69**(2): 645–673.

Gaver, J. J. and Gaver, K. M. (1993). Additional evidence on the association between the investment opportunity set and corporate financing, dividend, and compensation policies, *Journal of Accounting and Economics* **16**(1): 125–160.

Hochreiter, S. and Schmidhuber, J. (1997). Long short-term memory, *Neural Comput.* **9**(8): 1735–1780.

Hong, H. and Stein, J. C. (1999). A unified theory of underreaction, momentum trading, and overreaction in asset markets, *Journal of Finance* **54**(6): 2143–2184.

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

Huang, A. H., Zang, A. Y. and Zheng, R. (2014). Evidence on the information content of text in analyst reports, *Accounting Review* **89**(6): 2151–2180.

Jegadeesh, N., Kim, J., Krische, S. D. and Lee, C. M. (2004). Analyzing the analysts: When do recommendations add value?, *Journal of Finance* **59**(3): 1083–1124.

Jiang, F., Lee, J., Martin, X. and Zhou, G. (2019). Manager sentiment and stock returns, *Journal of Financial Economics* **132**(1): 126–149.

Kallapur, S. and Trombley, M. A. (1999). The association between investment opportunity set proxies and realized growth, *Journal of Business Finance & Accounting* **26**(3-4): 505–519.

Kim, Y. (2014). Convolutional neural networks for sentence classification, *Proceedings of the EMNLP*, 1746–1751.

Kothari, S. P., So, E. and Verdi, R. (2016). Analysts' forecasts and asset pricing: A survey, *Annual Review of Financial Economics* **8**: 197–219.

Kudo, T., Yamamoto, K. and Matsumoto, Y. (2004). Applying conditional random fields to Japanese morphological analysis, *Proceedings of the EMNLP*, 230–237.

Kurohashi, S. and Nagao, M. (1998). Building a Japanese parsed corpus while improving the parsing system, *Proceedings of the NLPRS*, 719–724.

Lang, L., Ofek, E. and Stulz, R. (1996). Leverage, investment, and firm growth, *Journal of Financial Economics* **40**(1): 3–29.

Li, F. (2010). The information content of forward-looking statements in corporate filings—A naïve Bayesian machine learning approach, *Journal of Accounting Research* **48**(5): 1049–1102.

Lin, H.-w. and McNichols, M. F. (1998). Underwriting relationships, analysts' earnings forecasts and investment recommendations, *Journal of Accounting and Economics* **25**(1): 101–127.

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

Lowry, M., Michaely, R. and Volkova, E. (2020). Information revealed through the regulatory process: Interactions between the SEC and companies ahead of their IPO, *The Review of Financial Studies*, Forthcoming

Matsumoto, D., Pronk, M. and Roelofsen, E. (2011). What makes conference calls useful? The information content of managers' presentations and analysts' discussion sessions, *Accounting Review* **86**(4): 1383–1414.

Matsumoto, Y., Kitauchi, A., Yamashita, T., Hirano, Y., Matsuda, H., Takaoka, K. and Asahara, M. (2000). *Morphological Analysis System ChaSen Version 2.2*, Nara Institute of Science and Technology, Ikoma.

Sul, H. K., Dennis, A. R. and Yuan, L. I. (2017). Trading on twitter: Using social media sentiment to predict stock returns, *Decision Sciences* **48**(3): 454–488.

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

Tetlock, P. C., Saar-Tsechansky, M. and Macskassy, S. (2008). More than words: Quantifying language to measure firms' fundamentals, *Journal of Finance* **63**(3): 1437–1467.

Tsoumakas, G. and Katakis, I. (2009). Multi-label classification: An overview, *International Journal of Data Warehousing and Mining* **3**: 1–13.

Womack, K. L. (1996). Do brokerage analysts' recommendations have investment value?, *Journal of Finance* **51**(1): 137–167.

Zhang, J. L., Härdle, W. K., Chen, C. Y. and Bommes, E. (2016). Distillation of news flow into analysis of stock reactions, *Journal of Business & Economic Statistics* **34**(4): 547–563.

Figure 1 The multi-label issue and binary relevance method

This figure illustrates how we transform a multi-label problem into separate binary problems. We adopt the binary relevance method (Tsoumakas and Katakis, 2009), having each sentence processed 10 times binary classifications, sequentially for each of the 10 topics. The original sentence in “e.g.1” is in Japanese, and we translate it into English.

e.g.1 弊社予想より弱い2018年度業績ガイダンスを受け、我々は業績予想を見直し下方に修正するが、
会社側は業界の競争激化を過度に保守的に見積もっていると考えるため、投資判断は買い推奨を継続する。

Given weaker than expected official guidance for FY2018, we lowered our earnings forecast but kept the buy rating unchanged based on our view that management was overly conservative about the competitive landscape.

Sample sentence	Topic 1 Analyst rating	Topic 2 Target price	Topic 3 Earnings forecast	…	Topic 9 News of official guidance	…
1	✓		✓		✓	
2		✓				
3					✓	
4						
5	✓		✓			

Sample sentence	Topic 1 Yes	Topic 1 No
1	✓	
2		✓
3		✓
4		✓
5	✓	

Sample sentence	Topic 2 Yes	Topic 2 No
1		✓
2	✓	
3		✓
4		✓
5		✓

Sample sentence	Topic 3 Yes	Topic 3 No
1	✓	
2		✓
3		✓
4		✓
5	✓	

Sample sentence	Topic 9 Yes	Topic 9 No
1	✓	
2		✓
3	✓	
4		✓
5		✓

Figure 2 The out-of-sample test
This figure depicts the distribution of the forecast returns vs. the actual returns in the out-of-sample test, where horizontal axis displays the predicted CAR using model (4) shown in Section 4.3 and vertical axis displays the actual CAR. The test period is from Jan 2018 to Jun 2018 with the sample of 12,784 observations. The adjusted R^2 is 12.9%.

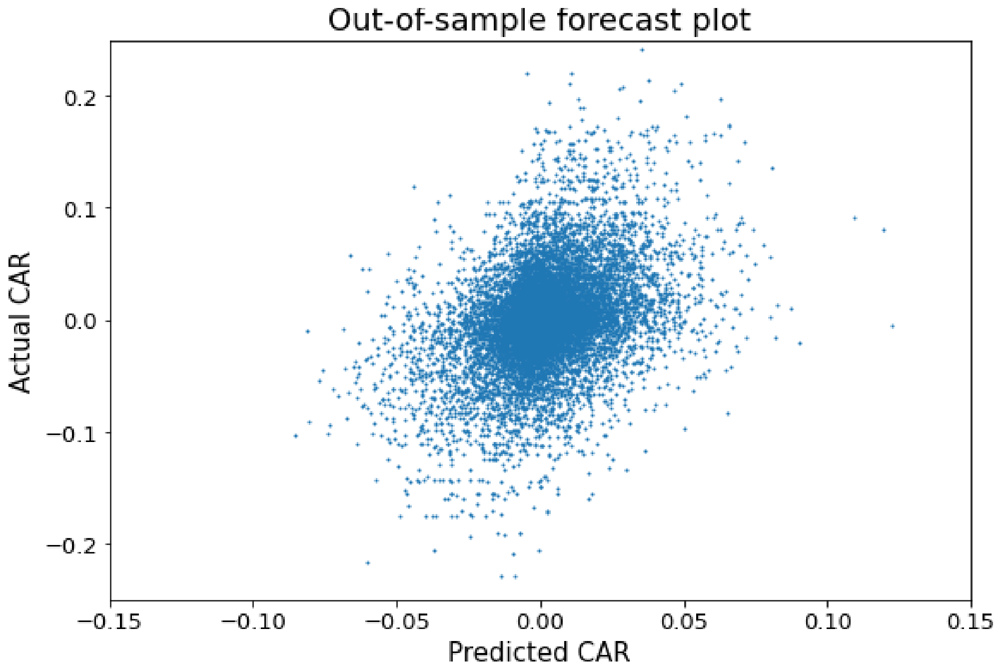

Table 1 The list of defined topics

This table displays a hierarchical summary of topics and the description for each detailed topic.

	Topic name	Description
Justification topic		
Topic 1	Analyst rating	Rationale for an analyst's stock recommendation
Topic 2	Target price and valuation	Discussion on a target price and valuation methodology
Topic 3	Earnings forecast	Explanation of an analyst's earnings forecast
Opinion topic		
Topic 4	Analyst conclusion	Conclusive opinion on the reported matter
Topic 5	Analyst general argument	Other arguments that support an analyst's opinion
Topic 6	Analyst discussion on risk	Examination on risk and uncertainty on upside and downside
Corporate Fact topic		
Topic 7	News on profits	News related with profits
Topic 8	News on non-profit items	News related with non-profit accounting items
Topic 9	News on official guidance	News on official guidance and comparison of actual result with guidance
Topic 10	News on corporate action	News on corporate actions such as dividend, share buy-back

Table 2.A The distribution of topic tones

This table displays the distribution of topic tones. The sum of the far right column, the proportion in the sample, is not equal to 100% due to sentences with multiple topic labels, as well as sentences without any assigned topics.

		% of positive	% of neutral	% of negative	# of detected sentences	Proportion in the sample
Topic 1	Analyst rating	50.23%	36.84%	12.93%	60,320	8.69%
Topic 2	Target price and valuation	32.68%	55.46%	11.86%	71,011	10.23%
Topic 3	Earnings forecast	32.22%	48.35%	19.42%	50,436	7.27%
Topic 4	Analyst conclusion	46.27%	29.30%	24.42%	65,970	9.51%
Topic 5	Analyst general argument	37.88%	50.15%	11.97%	108,940	15.70%
Topic 6	Analyst discussion of risk	25.85%	26.91%	47.24%	23,695	3.41%
Topic 7	News on profits	39.49%	42.14%	18.37%	42,193	6.08%
Topic 8	News on non-profit items	42.82%	36.32%	20.86%	75,769	10.92%
Topic 9	News on official guidance	32.30%	48.75%	18.95%	98,583	14.21%
Topic 10	News on corporate action	22.68%	71.92%	5.39%	27,308	3.94%

Table 2.B Summary statistics of the main sample

This table presents summary statistics for the variables in the main sample. *CAR* refers to the cumulative two-day abnormal returns starting from the analyst report issue date, where the abnormal returns are calculated as the difference between the raw stock return and the benchmark index return (i.e., TOPIX return). *REC_Rev* is the analyst recommendation revision, measured as the current report's recommendation value minus the analyst's last recommendation value, where the analyst recommendation values of Sell, Underperform, Hold, Buy and Strong Buy are expressed in the 1 to 5 rating system. *TP_Rev* represents the target price revision and *EPS_Rev* is the earnings forecast revision, defined similarly to *REC_Rev*. *CAR_Prior* is the cumulative 20-day abnormal return before the analyst report date, *BTM* is book-to-market ratio and Size is logarithm of the market value of equity. TT_i is the topic tone for specific topic.

	mean	std	min	max
CAR	0.002	0.043	-0.366	0.565
REC_Rev	-0.003	0.324	-4.000	4.000
TP_Rev	0.011	0.061	-0.220	0.368
EPS_Rev	0.000	0.005	-0.039	0.022
CAR_Prior	0.006	0.069	-0.504	1.072
BTM	0.770	0.450	0.010	4.000
Size	13.539	1.268	8.330	17.014
TT_1	0.038	0.095	-1.000	1.000
TT_2	0.026	0.084	-1.000	1.000
TT_3	0.010	0.071	-0.667	0.750
TT_4	0.023	0.114	-0.714	0.800
TT_5	0.040	0.093	-0.500	0.800
TT_6	-0.006	0.035	-0.500	0.333
TT_7	0.014	0.071	-0.556	0.636
TT_8	0.023	0.088	-0.615	0.833
TT_9	0.022	0.114	-1.000	1.000
TT_{10}	0.008	0.040	-0.500	0.750

Table 3 Informativeness of the overall DNN tone

This table presents regression results for the following regression:

$$CAR = \alpha_0 + \beta_1 Tone\_DNN + \beta_2 Tone\_NB + \gamma_1 REC\_Rev + \gamma_2 TP\_Rev + \gamma_3 EPS\_Rev + \sum_j \delta_j Controls_j + \epsilon,$$

where $Tone\_DNN$ is the overall tone extracted by DNN classifier and $Tone\_NB$ is the overall tone based on Naïve Bayes approach. Other variables are defined in table 2.b. Industry and year fixed-effects are controlled and the standard errors are estimated using a two-dimension cluster control at the firm and analyst's brokerage firm levels. All the tones are standardized to have mean zero and unit variance for the convenience of comparing the magnitude of the estimated loading. We report t-statistics in the parentheses.

	(1)	(2)	(3)	(4)
Tone_DNN			0.014	0.013
			(14.74)	(16.27)
Tone_NB		0.010		0.003
		(9.70)		(4.99)
REC_Rev	0.008	0.008	0.006	0.006
	(15.01)	(15.67)	(12.46)	(12.96)
TP_Rev	0.054	0.032	0.016	0.015
	(7.00)	(5.03)	(2.32)	(2.22)
EPS_Rev	0.531	0.419	0.320	0.316
	(5.21)	(5.01)	(3.91)	(3.93)
CAR_prior	-0.021	-0.027	-0.031	-0.031
	(-2.94)	(-4.14)	(-5.08)	(-5.16)
BTM	0.003	0.006	0.008	0.008
	(2.02)	(3.42)	(5.03)	(5.01)
Size	-0.001	-0.001	-0.001	-0.001
	(-1.45)	(-1.91)	(-1.97)	(-2.02)
Intercept	0.005	0.009	0.009	0.009
	(0.86)	(1.88)	(1.94)	(2.06)
Observations	42,818	42,818	42,818	42,818
Adjusted R^2	2.3%	7.3%	11.6%	11.8%

Table 4 Informativeness of topic tones

This table presents the results for the following regression:

$$CAR = \alpha_0 + \sum_i \gamma_i DetailedTone_i + \gamma_1 REC\_Rev + \gamma_2 TP\_Rev + \gamma_3 EPS\_Rev + \sum_j \delta_j Controls_j + \epsilon,$$

where $DetailedTone$ stands for TT or other categories of tone, such as $Tone\_Justif$, $Tone\_Quali$, $Tone\_Opn$ and $Tone\_Fact$. TT is the topic tone in each of the 10 detailed topics. $Tone\_Justif$ ($Tone\_Quali$) is the aggregate topic tone from Topic 1 to Topic 3 (Topic 4 to Topic 10), while $Tone\_Opn$ ($Tone\_Fact$) is aggregate topic tone from Topic 4 to Topic 6 (Topic 7 to Topic 10). The variables are defined in Table 2.b. Industry and year fixed-effects are controlled and the standard errors are estimated using a two-dimension cluster control at the firm and analyst's brokerage firm levels. All the topic tones are standardized to have mean zero and unit variance for the convenience of comparing the magnitude of the estimated loading. We report the t-statistics in the parentheses.

Panel A

	(1)	(2)	(3)	(4)
Tone_Justif	0.007			
	(5.99)			
Tone_Quali		0.015		
		(20.52)		
Tone_Opn			0.013	
			(15.00)	
Tone_Fact				0.013
				(17.20)
REC_Rev	0.007	0.008	0.008	0.008
	(14.28)	(14.88)	(14.34)	(15.24)
TP_Rev	0.020	0.045	0.043	0.050
	(2.55)	(7.08)	(6.00)	(8.07)
EPS_Rev	0.413	0.373	0.461	0.373
	(4.15)	(5.05)	(5.73)	(4.69)
CAR_Prior	-0.023	-0.028	-0.027	-0.026
	(-3.25)	(-4.71)	(-4.36)	(-4.15)
BTM	0.004	0.007	0.006	0.007
	(2.78)	(4.71)	(3.71)	(4.28)
Size	-0.001	-0.001	-0.001	-0.001
	(-1.57)	(-1.34)	(-1.78)	(-1.14)
Intercept	0.006	0.005	0.006	0.005
	(1.23)	(1.11)	(1.27)	(0.93)
Observations	42,818	42,818	42,818	42,818
Adjusted R^2	4.3%	13.6%	10.7%	11.0%

Panel B

	(1)	(2)	(3)
Tone_Justify	0.002	0.002	
	(3.32)	(3.33)	
Tone_Quali	0.015		
	(22.40)		
Tone_Opn		0.008	
		(14.08)	
Tone_Fact		0.009	
		(13.32)	
TT_1			-0.000
			(-0.39)
TT_2			0.001
			(2.93)
TT_3			0.001
			(3.73)
TT_4			0.011
			(13.94)
TT_5			0.001
			(3.78)
TT_6			0.000
			(0.60)
TT_7			0.003
			(3.86)
TT_8			0.002
			(6.34)
TT_9			0.003
			(7.78)
TT_{10}			0.001
			(4.56)
REC_Rev	0.007	0.007	0.008
	(14.37)	(14.31)	(14.33)
TP_Rev	0.038	0.037	0.042
	(6.58)	(6.44)	(7.90)
EPS_Rev	0.352	0.357	0.350
	(4.84)	(4.98)	(5.14)
CAR_Prior	-0.029	-0.029	-0.029
	(-4.74)	(-4.77)	(-4.87)
BTM	0.007	0.007	0.006
	(4.78)	(4.78)	(4.39)
Size	-0.001	-0.001	-0.001
	(-1.35)	(-1.39)	(-1.75)
Intercept	0.006	0.006	0.007
	(1.17)	(1.19)	(1.42)
Observations	42,818	42,818	42,818
Adjusted R^2	13.7%	13.7%	15.4%

Table 5 Firm-level analyses

This table presents the results for the firm-level regression:

$$CAR_t = \alpha_0 + \sum_{i=1}^{10} \beta_i \frac{1}{m_t} \sum_{s=1}^{m_t} TT_{i,t,s} + \gamma_1 \frac{1}{m_t} \sum_{s=1}^{m_t} REC\_Rev_{t,s} + \gamma_2 \frac{1}{m_t} \sum_{s=1}^{m_t} TP\_Rev_{t,s} + \gamma_3 \frac{1}{m_t} \sum_{s=1}^{m_t} EPS\_Rev_{t,s} + \sum\nolimits_j \delta_j Controls_{j,t} + \epsilon,$$

where m_t denotes the number of reports on the same stock issued on the same day t, and $\mathrm{TT}_{i,t,s}$ refers to the tone of topic i in report s. Industry and year fixed-effects are controlled. The standard errors are estimated using a two-dimension cluster control at the firm and analyst's brokerage firm levels. All the tones are standardized to have mean zero and unit variance for the convenience of comparing the magnitude of the estimated loading. We report the t-statistics in the parentheses.

	(1)	(2)	(3)	(4)
Tone_DNN		0.011	0.057	
		(30.76)	(16.19)	
Tone_DNN*Size			-0.003	
			(-13.75)	
Tone_DNN*BTM			-0.002	
			(-2.55)	
TT_1				0.000
				(0.27)
TT_2				0.001
				(2.24)
TT_3				0.001
				(2.63)
TT_4				0.010
				(27.20)
TT_5				0.001
				(3.23)
TT_6				-0.000
				(-0.53)
TT_7				0.003
				(8.25)
TT_8				0.002
				(6.06)
TT_9				0.003
				(8.02)
TT_{10}				0.001
				(3.80)
REC_Rev	0.009	0.007	0.007	0.008
	(17.28)	(13.20)	(13.92)	(16.00)
TP_Rev	0.046	0.013	0.015	0.036
	(12.47)	(3.60)	(4.07)	(9.88)
EPS_Rev	0.523	0.341	0.326	0.360
	(10.02)	(6.86)	(6.70)	(7.48)
CAR_Prior	-0.016	-0.024	-0.025	-0.023
	(-3.52)	(-5.55)	(-5.66)	(-5.48)
BTM	0.003	0.006	0.006	0.005
	(2.20)	(5.74)	(6.08)	(5.15)
Size	-0.001	-0.001	-0.001	-0.001
	(-3.18)	(-3.15)	(-2.72)	(-2.08)
Intercept	0.008	0.010	0.013	0.005
	(1.96)	(2.89)	(3.35)	(1.13)
Observations	27,586	27,586	27,586	27,586
Adjusted R^2	2.8%	9.6%	10.7%	15.4%

Table 6 The topic tone and future earnings growth

This table presents results for following regression:

$$Growth_{t+n} = \alpha_0 + \beta_1 Tone\_DNN_t + \beta_2 Tone\_NB_t + \gamma_1 REC\_Rev_t + \gamma_2 TP\_Rev_t + \gamma_3 EPS\_Rev_t + \sum_j \delta_j Controls_{j,t} + \epsilon,$$

where $Growth_{t+n}$ is estimated as the difference between operating income from year t+n and year t, scaled by total assets in year t, and n equals 1, 2 and 3 respectively in specification (1), (2) and (3). All the tones are standardized to have mean zero and unit variance for the convenience of comparing the magnitude of the estimated loading.

	(1) Growth_{t+1}	(2) Growth_{t+2}	(3) Growth_{t+3}
Tone_DNN	0.004	0.005	0.005
	(9.83)	(6.52)	(4.81)
Tone_NB	0.001	0.001	0.001
	(3.26)	(1.18)	(0.84)
REC_Rev	-0.001	-0.001	0.000
	(-1.69)	(-1.07)	(0.04)
TP_Rev	0.001	0.018	0.012
	(0.16)	(2.73)	(1.52)
EPS_Rev	0.373	0.304	0.332
	(7.52)	(3.54)	(3.23)
ROA	-0.159	-0.209	-0.254
	(-1.63)	(-1.56)	(-1.51)
BTM	-0.027	-0.037	-0.037
	(-5.59)	(-5.53)	(-4.04)
Size	0.000	0.001	0.001
	(0.17)	(0.37)	(0.57)
Intercept	0.046	0.062	0.024
	(4.32)	(3.64)	(1.01)
Observations	40,469	40,047	39,480
Adjusted R^2	22.5%	26.5%	22.4%

Table 7 CAR based on portfolio analyses

This table presents the CAR at the portfolio level analyses. In panel A (B), the analyst reports are sorted into terciles independently according to TT_4 ($Tone\_NB$) and REC_Rev. The t-statistics of the High-low or Upgrade-Downgrade are reported in the parentheses.

Panel A Double sorting based on TT_4 and REC_Rev

		Downgrade	Reiteration	Upgrade	Upgrade-Downgrade	T-statistics
TT4	Low	-0.036	-0.026	0.010	0.045	(5.13)
	Middle	-0.019	0.001	0.023	0.042	(20.45)
	High	-0.015	0.019	0.032	0.046	(9.13)
	High-Low	0.021	0.045	0.022		
	T-statistics	(4.08)	(59.76)	(2.48)		

Panel B Double sorting based on $Tone\_NB$ and REC_Rev

		Downgrade	Reiteration	Upgrade	Upgrade-Downgrade	T-statistics
Tone_NB	Low	-0.025	-0.01	0.015	0.040	(9.23)
	Middle	-0.018	0.003	0.022	0.040	(12.26)
	High	-0.017	0.013	0.029	0.046	(13.47)
	High-Low	0.008	0.023	0.014		
	T-statistics	(2.60)	(44.59)	(3.10)		

Table 8 The out of sample test

This table presents the predictive power of the topic tones for CAR in the out of sample period, where we freeze the parameters of DNN algorithm learned from the main sample and apply the frozen parameters to construct topic tones from January 2018 to June 2018. All the topic tones are standardized to have mean zero and unit variance for the convenience of comparing the magnitude of the estimated loading.

	(1)	(2)	(3)
Tone_DNN	0.013	0.012	
	(13.45)	(11.42)	
Tone_NB		0.002	
		(4.11)	
TT_1			0.000
			(0.62)
TT_2			0.001
			(2.42)
TT_3			0.001
			(1.63)
TT_4			0.010
			(7.96)
TT_5			0.001
			(1.99)
TT_6			0.000
			(0.28)
TT_7			0.003
			(4.42)
TT_8			0.002
			(4.94)
TT_9			0.004
			(4.04)
TT_{10}			0.000
			(0.79)
REC_Rev	0.007	0.007	0.008
	(5.88)	(5.86)	(8.52)
TP_Rev	0.013	0.012	0.038
	(1.87	(1.69)	(4.81)
EPS_Rev	0.303	0.307	0.340
	(2.77)	(2.82)	(3.42)
CAR_Prior	-0.015	-0.015	-0.014
	(-1.60)	(-1.65)	(-1.59)
BTM	0.007	0.007	0.006
	(3.26)	(3.29)	(2.95)
Size	-0.000	-0.000	-0.000
	(-0.43)	(-0.44)	(-0.22)
Intercept	0.009	0.010	0.001
	(1.01)	(1.11)	(0.15)
Observations	12,784	12,784	12,784
Adjusted R^2	12.2%	12.3%	15.9%

Appendix

A1. Uniqueness of Japanese

The Japanese writing system is formed of three different sets of characters: Hiragana, Katakana, and Kanji. It is also common to see Arabic numerals and Latin script in Japanese text (Bond et al., 2016). Similar to Chinese and Thai, Japanese does not have clear morpheme boundaries, which makes tokenization quite hard in the pre-processing for NLP. Given that the smallest unit for NLP is a word, it is crucial to properly decompose a sentence into words. Consequently, several novel morphological analyzers have been developed, as follows: JUMAN (Kurohashi and Nagao, 1998), ChaSen (Matsumoto et al., 2000), and MeCab (Kudo et al., 2004). All three perform the segmentation, part-of-speech tagging, and lemmatization simultaneously. In our environment (Windows), MeCab has a stable performance with R and Python and thus we extensively rely on MeCab for the tokenization.

A2. LSTM and CNN

In our NLP algorithms, we apply Convolutional Neural Network (CNN) (Kim (2014)) and Long Short Term Memory (LSTM) (Hochreiter and Schmidhuber (1997)). LSTM has advantage in processing input data sequentially and storing input information in hidden layers as memory, whereas CNN is better designed to capture certain local patterns by using several filters simultaneously scanning inputs in a sequential manner. For specific architectures and hyperparameter selection, we conduct five-fold cross validation. We pick the best-performing model and its configuration based on cross validation accuracy. In particular, we test standard and bi-directional LSTM, as well as different window sizes from three to seven words, which are also called n-grams where $n \in \{3, ..., 7\}$, for CNN. To allow multi-label classification, our topic classification model consists of 10 separate binary classifiers, while the tone model is a trinary classifier. We compare and report the cross validation accuracy of the NLP algorithms, including the four algorithms that we use in the DNN approach and Naïve Bayes approach, in Table A.2.

Table A.2 The performance of NLP algorithms

This table compares the cross validation accuracy of the NLP algorithms, including the four algorithms that we used in our DNN approach and Naïve Bayes approach. Five-fold cross validation accuracy, defined as follows:

$$\text{CV Accuracy} = \frac{1}{K}\sum_{i=1}^{K}\frac{\text{No. of correctly predicted samples in CV set}_i}{\text{No. of total samples in CV set}_i}$$

where CV is an abbreviation for "cross-validation" and we took $K = 5$ for the training. The best-performing architecture for each task is highlighted in boldface type, which is selected for subsequent tasks that include topic and tone classifications. The numbers in the parentheses at CNN indicate the window sizes of the filters (n-gram) that we used in the architecture. The topic models consist of 10 different binary classifiers and the tone model is a single trinary classifier.

CV accuracy of the topic model

		CNN (3-4-5)	CNN (5-6-7)	LSTM	Bidirectional LSTM	Naïve Bayes
Topic 1 model	Analyst rating	**0.985**	0.984	0.983	0.978	0.948
Topic 2 model	Target price and valuation	**0.981**	0.980	0.975	0.975	0.944
Topic 3 model	Earnings forecast	0.951	0.954	**0.959**	0.950	0.907
Topic 4 model	Analyst conclusion	**0.958**	**0.958**	0.949	0.946	0.910
Topic 5 model	Analyst general argument	0.900	0.904	**0.907**	0.903	0.817
Topic 6 model	Analyst discussion of risk	0.988	0.989	**0.993**	0.992	0.938
Topic 7 model	News on profits	**0.949**	0.947	0.945	0.943	0.886
Topic 8 model	News on non-profit items	0.901	**0.903**	0.894	0.894	0.856
Topic 9 model	News on official guidance	0.940	0.939	**0.945**	0.933	0.878
Topic 10 model	News on corporate action	0.968	0.969	**0.972**	0.970	0.927

CV accuracy of the tone model

	CNN (3-4-5)	CNN (5-6-7)	LSTM	Bidirectional LSTM	Naïve Bayes
Tone model	**0.745**	0.742	0.740	0.738	0.630