The Determinants and Information Content of the Forward-looking Statements in Corporate Filings—a Naïve Bayesian Machine Learning Approach *

Feng Li

Ross School of Business, University of Michigan

701 Tappan St., Ann Arbor, MI 48109

Phone: (734) 936-2771

E-mail: Feng@umich.edu

First Draft: January 2008
This Draft: August 18, 2008

*I would like to thank Phil Berger, John Core, Ilia Dichev, Raffi Indjejikian, Mark Lang, Roby Lehavy, Russell Lundholm, Venky Nagar, Marlene Plumlee, Cathy Shakespeare, Terry Shevlin, Doug Skinner, K. R. Subramanyam, Hal White, and the participants of the 2008 University of Michigan Traverse City Conference and the 2008 North American Summer Meeting of the Econometric Society for helpful comments. I also thank Mike Chen, Qian Fang, Qian Li, Chris Liong, Doris Lu, Amy McMahon, Bharat Modi, Qimin Ni, Darin Styles, Sarah Xia, Quan Yuan, Kathy Wang, Tina Wei, and especially Rencheng Wang and Winnie Wei for their excellent research assistance and the Stephen M. Ross School of Business and the Accounting Department at the University of Michigan for financial support.

Abstract

This paper examines the tone and content of the forward-looking statements ("FLS") in corporate 10-K and 10-Q filings using a Naïve Bayesian machine learning algorithm. I first manually categorize 30,000 sentences of randomly selected FLS extracted from the MD&As along two dimensions: (1) tone (i.e., positive versus negative tone); and (2) content (i.e., profitability, operations, and liquidity etc.). These manually coded sentences are then used as training data in a Naïve Bayesian machine learning algorithm to classify the tone and content of about 13 million forward-looking statements from more than 140,000 corporate 10-K and 10-Q MD&As between 1994 and 2007.

I find that firms with better current performance, lower accruals, smaller size, lower market-to-book ratio, and less return volatility tend to have more positive forward-looking statements in MD&As. The average tone of the forward-looking statements in a firm's MD&A is positively associated with future earnings and liquidity, even after controlling for other determinants of future performance and there is no systematic change in the information content of MD&As over time. Finally, the evidence indicates that financial analysts do not fully understand the information content of the MD&As in making their forecasts.

1 Introduction

This paper analyzes the determinants and information content of the forward-looking statements in the MD&A section of corporate 10-K and 10-Q filings by examining the tone and content of the statements. The Securities and Exchange Commission mandated in 1980 that public companies' annual reports include a Management's Discussion and Analysis of Financial Condition and Results of Operations (MD&A) section that assesses the enterprises' liquidity, capital resources and operations in a way many investors can understand. One of the goals was to make public information about predictable future events and trends that may affect future operations of the businesses.

Yet, whether the MD&A disclosures are informative remains an open empirical question. On the one hand, consistent with the SEC's intention that the MD&A disclosures should provide relevant information to investors, MD&A is arguably the most read and most important component of the financial section (Tavcar (1998)) and of all the disclosure items of the annual report, sell-side financial analysts in the U.S. most frequently use MD&A when preparing their analyst reports (Knutson (1993) and Rogers and Grant (1997)). The safe harbor provisions of the Private Securities Litigation Reform Act of 1995 also intend to elicit more forward-looking information from reporting companies (Grundfest and Perino (1997)) and therefore should make the MD&A disclosures more informative.

On the other hand, companies may not have incentives to disclose in the MD&As because of concerns over proprietary costs (Verrecchia (1983)), uncertainties about the judicial interpretation of safe harbor protection (Grundfest and Perino (1997)), and the lack of mandatory auditing requirement for the MD&As (Hüfner (2007)). The SEC also argues that MD&As provide substantial boilerplate disclaimers and disclosures, generic language, and immaterial detail without much information content (SEC (2003)). Consistent with these arguments, Pava and Epstein (1993) show that the 25 randomly selected companies they studied did a good job of describing historical events, but very few provided useful and accurate forecasts in MD&As.

In this paper, I study the determinants and information content of the forward-looking statements in the MD&A section, which are likely to be the most informative part of the section. First, I explore the variations in the tone of the forward-looking statements in MD&As and study its economic determinants. Second, I examine whether the forward-looking statements contain (incremental) information about future profitability and liquidity. I then explore whether the information content of MD&As changed over time, especially after 2003, when SEC issued new guidelines for preparing the MD&As and the Sarbanes-Oxley Act enhanced disclosure requirements in MD&As(Bainbridge (2007)). Finally, I examine whether financial analysts understand and fully utilize the information in MD&As in their forecasts.

I rely on a Naïve Bayesian learning algorithm, instead of the dictionary-based approach (e.g., Kothari and Short (2006)), to do the content analysis. First, I manually categorize about 30,000 sentences of randomly selected forward-looking statements extracted from the MD&A section of the 10-Q filings along two dimensions: (1) tone (i.e., positive versus negative tone); and (2) content (i.e., product market condition, consumer demand, cost, margins, liquidity, and litigation risk, etc.). These manually coded sentences are then used as training data in the Naïve learning algorithm to categorize the tone and content of other forward-looking statements in 10-Q and 10-K filings.

N-fold cross-validation tests (with N varying from 3 to 50) indicate that the algorithm can predict the tone (out of four possible tones: positive, neutral, negative, and uncertain) and content category (out of 12 possible categories) of the forward-looking statements in 10-Q and 10-K filings with a success rate of about 59% and 63%, respectively. To provide a benchmark for this result, I show that it is much better than doing tone analysis using a general dictionary-based approach—content analysis based on the Harvard General Inquirer and LIWC dictionary can only classify the tone of the training data with a success rate of about 40%.

I then use the Bayesian learning algorithm to categorize the tone and content of about 13 million forward-looking statements from more than 140,000 corporate 10-Q and 10-K filings between 1994 and 2007. The results indicate that firms with better current performance, lower accruals, smaller size, lower market-to-book ratio, and less return volatility tend to have more positive forward-looking statements in MD&As.

I find that the average tone of the forward-looking statements in a firm's MD&A is positively correlated with its future earnings and liquidity and has explanatory power in addition to other variables that can predict future performance. For instance, the return on assets in the next year of firms with perfectly positive MD&A tones is 8 percentage points higher than that of firms with perfectly negative tones. An inter-quartile change in MD&A tones implies a difference in annual return on assets of 1.5 percentage points. These effects are found after controlling for current earnings, stock returns, accruals, and other factors that may affect future performance. Examining the information content of the MD&As over time shows that, despite SEC's continuous effort on strengthening MD&A disclosures and the passage of the Sarbanes-Oxley Act, there is no systematic change in the informativeness of MD&As over time.

The results also indicate that sell-side financial analysts' consensus earnings forecasts are higher when MD&A FLS are more optimistic. It appears, however, that they do not fully utilize the information in the MD&A tones—even after controlling for the latest analysts' forecasts before the announcement of future earnings, the tone of the MD&A forward-looking statements still has significant predictive power for future earnings and liquidity.

Finally, I decompose the tones of the forward-looking statements in MD&As into three dimensions: the tones of profitability-related FLS (i.e., revenue, cost, and operations), liquidity-related FLS (i.e., capital resources, financing, and investment), and other FLS, where the categories are generated by the learning algorithm classification. The empirical results show that both profitability-related FLS and liquidity-related FLS have information contents about future fundamentals, but the liquidity-related FLS have stronger implications both economically and statistically.

This paper contributes to the literature in several ways. This study is the first to use statistical learning methodology to analyze corporate disclosures and the empirical evidence shows that this approach performs significantly better than content analysis based on general dictionaries (e.g., the Harvard General Inquirer or the LIWC software) for analyzing corporate filings. Because the empirical analyses in the paper are joint tests of the machine learning methodology and the economic hypotheses, the results in the paper show that the statistical learning algorithm used widely in other research areas (e.g., Mitchell (2006)) can be successfully applied to corporate financial statements settings and could be useful for future research in disclosures.

This paper is also the first large-sample study on the forward-looking statements made by managers in corporate 10-Q and 10-K filings.[1] This extends the literature on management disclosures of forward-looking information (Patell (1976), Penman (1980), Pownell, Wasley, and Waymire (1993), Skinner (1994), Dietrich, Kachelmeier, Kleinmuntz, and Linsmeier (1997), Miller and Piotroski (2000), Hutton, Miller, and Skinner (2003), and Lang and Lundholm (2003)).

Some prior small-sample studies based on content analysis by human coders also find that there is information content in MD&As about future firm performance (Bryan (1997), Barron, Kile, and O'Keefe (1999), and Callahan and Smith (2004)). The innovations of this paper over the prior studies are several folds. First, none of the prior studies examines the determinants of the MD&A tones, and in this paper, I test hypotheses about economic factors that may explain the variations in MD&A tones. Second, perhaps due to their small sample sizes, neither do prior papers provide any empirical evidence on the informativeness of MD&As over time, even though the SEC has strengthened the requirements on MD&A disclosures significantly through releases and rules in the last ten years. This study provides evidence on this and thus sheds light on the SEC regulation. Finally, I empirically examine whether analysts fully understand the information in MD&As, which is not tested in the prior literature.

The rest of the paper proceeds as follows. Section 2 discusses the nature of MD&A

[1] A contemporaneous study (Muslu, Radhakrishnan, Subramanyam, and Lim (2008)) also examines the information content of the forward-looking statements in MD&As, but the focus of that paper is on the intensity of the forward-looking information, rather than the tone.

disclosures and hypotheses. Section 3 presents the details of the Naïve Bayesian learning algorithm and Section 4 discusses its empirical implementation, including the validation test results. Section 5 presents the empirical results and Section 6 concludes.

2 Literature review and research questions

2.1 MD&A and forward-looking statements

Item 303 of Regulation S-K ("Reg. S-K") presents the specific SEC rules for the MD&A and many SEC releases provide more detailed instructions and interpretive guidance. Since 1968, firms had to discuss unusual (non-recurring) components of earnings in MD&A (SEC (1968)). Later, firms were required to analyze certain trends associated with operations (SEC (1974)). Dissatisfied with the disclosures firms were providing, the SEC granted protection under safe harbor rules in 1979, and then issued, in the following year, a revised requirement which is still in force (SEC (1980)). According to SEC (2003), the MD&A requirements have three principal objectives: (1) to provide investors with a narrative explanation of the financial statements of a company; (2) to increase overall company disclosure and to supply the contextual basis for investors on which they can analyze financial information; (3) to provide information about the quality and potential variability of the earnings and cash flows of a company.

The SEC also encourages direct forward-looking information on circumstances such as known material trends, events, commitments, and uncertainties. However, the requirements in the rules and releases are not set out in concrete, objectified terms. They leave the decision on prospective disclosure (or not) to the discretion of management with regard to three different aspects. Whether the disclosure of forward-looking information is mandatory depends cumulatively on management's assessment (1) whether an above-mentioned circumstance is "presently known," (2) on the judgment of whether such a circumstance is "reasonably likely," and (3) whether "material effects" are to be expected. However, neither the term "reasonably likely" nor the term "material" are clearly defined, let alone sufficiently objectified in the respective SEC statements or other SEC sources. Therefore, with regard to providing direct forward-looking information in the MD&A, management's discretion is given considerable leeway.

In order to encourage companies to give forward-looking statements in their MD&As, any predictive information provided is explicitly covered by the safe-harbor rule for projections (Reg. S-K Item 303 (a) Instr. 7). Even with the safe harbor provisions of the Reform Act, however, companies may be avoiding disseminating forward-looking information because of uncertainty regarding judicial interpretation of the safe harbor and because of fear regarding state court litigation where, plaintiffs will argue, no such safe harbor is available (Grundfest and Perino (1997)).

It is not mandatory for companies to have their MD&A reports audited. At best, auditors have a professional responsibility, which is supported by the American Institute of Certified Public Accountants, to check a company's MD&A information for material inconsistencies against the respective financial statements of that company. A company may not be held liable under the federal securities laws for projections and other forward-looking statements if the forward-looking statement is accompanied by meaningful cautionary statements identifying important factors that could cause actual results to differ materially from those in the forward-looking statement, or the plaintiff fails to prove that the forward-looking statement was made with actual knowledge that the statement was false or misleading (Hüfner (2007)).

2.2 Literature review

There is extensive research on the economic implications of voluntary corporate disclosures in the accounting literature.[2] Conceptually, there are at least three characteristics of the disclosures that are interesting to researchers: the level ("how much you say"), the meaning

[2]There is also a substantial literature on the determinants of corporate disclosures (e.g., Lang and Lundholm (1993) and Miller (2002)), the frequency of voluntary disclosures (e.g., Botosan and Harris (2000)), management forecast (e.g., Waymire (1984)), and the consequences of mandatory disclosures (e.g., Leuz and Verrecchia (2000)), which is discussed in details in Healy and Palepu (2001) and Core (2001).

or the tone ("what do you mean"), and the transparency ("how you say it").

Many of the studies can be categorized along two dimensions: the dependent variables and the independent variables (i.e., the disclosure measures) examined, as illustrated in the following table.

Sample papers on the implications of corporate disclosures

	Independent variable (Disclosure measures)		
Dependent variable	(1) "How much you say" (e.g., Level / Amount)	(2) "What you mean" (e.g., Meaning / Tone)	(3) "How you say it" (e.g., Transparency / Voice)
Cost of capital	Botosan (1997) Botosan and Plumlee (2002)*	Kothari and Short (2006)	
Future earnings		Bryan (1997) Miller and Piotroski (2000) Callahan and Smith (2004) Davis, Piger, and Sedor (2005)	Li (2008) Mayew and Venkatachalam (2008)
Analyst behavior	Lang and Lundholm (1996)* Barron, Kile, and O'Keefe (1999)		
Other		Levine and Smith (2006)	

*Note: The AIMR score used by these papers to measure disclosure quality, which mainly covers the details (i.e., "level") of disclosures ranked by analysts, also partly covers the "candor" of the disclosures and can also be regarded as a measure of disclosure transparency.

The majority of the prior studies fit into column (1) in the table above—they examine the level or the amount of disclosures, i.e., which topics/subjects the managers discuss in the disclosures ("how much you say"). With the exception of the AIMR scores, most studies rely on manual coding of the level of disclosures, perhaps due to the difficulty in creating a large-sample measure of disclosure quality. A few papers examine "how you say it" by examining the readability (Li (2008)) and vocal tones of disclosures (Mayew and Venkatachalam (2008)).

This paper fits into the second column in the table above (i.e., studying "what you mean" in corporate disclosures). Most papers in this area carry out content analysis with human coders and have small sample sizes. For example, Bryan (1997) studies 250 MD&As and Callahan and Smith (2004) examine 71 firms and 420 firm-years. With the obvious advantage of being more precise, manual coding has two disadvantages (Core (2001)): small sample sizes because of cost considerations and difficulty to replicate by other researchers due to the subjectivity in the coding process. In particular, small sample sizes may limit the scope of empirical tests. For instance, to study the change in the information content of MD&A disclosures over time, researchers need a relatively large panel data. One approach to solve the small sample problem is to rely on dictionary-based content analysis to understand corporate disclosures. The difference between this approach and this paper is discussed in great detail in Section 3.

The empirical results on the information content of MD&As remain mixed. Bryan (1997) finds that the discussions of future operations and planned capital expenditures are associated with future short-term performance. Callahan and Smith (2004) find that their disclosure index based on content analysis provides incremental explanatory power in predicting future firm performance and market valuation while controlling for current income and other related factors. On the other hand, Pava and Epstein (1993) test the SEC's contention that MD&As were not reaching the intended goal by examining 25 randomly companies and show that while most companies did a good job of describing historical events, very few provided useful and accurate forecasts (i.e., many companies make practically no predictions.). They also find a strong bias in favor of correctly projecting positive trends, while negative trends tended to be either ignored or not fully reported. A contemporaneous study by Muslu, Radhakrishnan, Subramanyam, and Lim (2008) examines the intensity of the forward-looking information in MD&As and finds that more intense discussion of forward-looking information makes stock returns incorporate future earnings more timely and reduce analysts' forecast errors.

The second broad stream of literature into which this paper fits is the research in accounting, finance, and other social science fields that analyzes the content of textual documents using computer algorithms. For example, there is extensive research on the information content of corporate earnings release (Davis, Piger, and Sedor (2005)), accounting policy disclosures (Levine and Smith (2006)), audit opinions (Butler, Leone, and Willenborg (2004)), financial news (Tetlock (2007), Tetlock, Saar-Tsechansky, and Macskassy (2007), and Core, Guay, and Larcker (2008)), Internet stock message board (Antweiler and Frank (2004)), multiple sources of financial text (Kothari and Short (2006)), presidential election campaign (Pennebaker and Stone (2001)), and art history (Martindale (1990)).

2.3 Hypotheses

2.3.1 Determinants of the MD&A tone

I explore several factors to explain the cross-sectional variations in the tone of MD&A FLS:

Current firm performance. While there is a substantial theoretical and empirical literature on the amount of disclosure, little theoretical work exists on the direction of the disclosure (i.e., tones). Some of the arguments on the relation between the level of disclosure and firm performance can apply to the MD&A tones. Litigation concern may encourage firms with good current performance to be more cautious in discussing future events in MD&A. Momentum in firm performance also suggests that the MD&A FLS may be more optimistic for firms with good current performance. On the other hand, earnings are mean-reverting (i.e., there is a negative relation between current earnings innovation and future earnings change) implies a more cautious and pessimistic tone in forward-looking statements for firms with better current performance.

Accruals. It is well documented that accruals are negatively associated with future firm performance and investors seem to under-react to this information. One explanation of the accrual anomaly is that managers manipulate the accrual component of earnings (Sloan (1996)). If this is true, it implies that managers would know the implications of the accruals for future performance and a negative correlation between accruals and the FLS tones is expected. Alternatively, accruals may simply proxy for firms' economic conditions (e.g., distress) and as insiders managers are likely to understand (at least partially) the implications of accruals for future earnings. Therefore, both arguments point to a negative relation between accruals and the MD&A tones. On the other hand, if managers are overconfident or fixated on current earnings, then they may not understand the implications of accruals for future performance and hence no relation between accruals and MD&A tone exists.

Firm size. Size captures many aspects of a firm's operational and business environment. The accounting literature has used firm size as a proxy for a firm's political cost (Watts and Zimmerman (1986)). Holding current performance constant, bigger firms may have more cautious forward-looking statements because of the higher political and legal cost due to their visibility.

Market-to-book ratio. High market-to-book firms are different from low market-to-book firms in many aspects, including the investment opportunity set and growth potential. To the extent that growth firms face more uncertain future economic conditions, a negative relation between market-to-book ratio and MD&A tones is expected.

Volatility of operations. Firms with more volatile business environments may be more cautious in discussing future events because of information uncertainty with regard to future performance. Alternatively, firms with high performance variability are likely to have more severe information asymmetry between managers and investors. Finally, performance variability may be related to the tones of MD&As because of its effect on firms' vulnerability to legal action. These factors all posit a negative relation between volatility of business and the tone in MD&As.

Reporting quarter. Prior research has shown that the accounting numbers behave systematically differently across different reporting quarters (Das and Shroff (2002)). To examine the potential implications of the reporting quarters on the MD&A tones, I include three quarterly dummies ($Q2$, $Q3$, and $Q4$), which are dummies for the 2nd, 3rd, and 4th quarter, respectively.

2.3.2 Information content of the MD&As

Next, I examine whether the forward-looking statements in corporate MD&A disclosures contain information about future profitability and liquidity, i.e., when managers are more optimistic in discussing future outlook, the future does turn out to be "brighter." Based on the argument in the Introduction section, there are many factors that may lead to non-informative FLS disclosures. Ex ante, it's not clear whether on average the FLS disclosures contain economically meaningful information about future firm performance.

I also examine whether there is a systematic change in the information content of the MD&As over time with a particular focus on pre- and post-2003 comparison. The SEC issued new guidelines on MD&A disclosures in 2003 and encouraged a significant increase in its information content and a reduction in boilerplate languages (SEC (2003)). The Sarbanes-Oxley Act also enhanced the MD&A disclosures requirement (Bainbridge (2007)): First, MD&A section is chose as the vehicle for more complete disclosure of off-balance-sheet transactions; Second, the Sarbanes-Oxley Act requires CEOs and CFOs to certify that their financial statements, including the MD&A section, fairly presents the financial conditions and results of operations of the issuer. A test of the change in the information content of MD&As over time will shed light on the effectiveness of the regulations.

2.3.3 Financial analysts forecasts and the MD&As

Lastly, I examine whether sell-side financial analysts fully understand the implications of the MD&As for future performance. Prior research has shown that financial analysts revise their forecasts upward when the MD&As are more optimistic (Bryan (1997)). The result is not surprising given that financial analysts in the U.S. most frequently use MD&A when preparing their analyst reports (Knutson (1993) and Rogers and Grant (1997)). The more interesting question is whether they utilize the information in MD&As efficiently. In this paper, I test this hypothesis by examining whether the tones of the MD&A FLS can predict future performance even after controlling for the latest analyst forecasts before earnings announcement.

3 Naïve Bayesian algorithm and text classification

There are two general approaches to do content analysis: rule-based approach versus statistical approach. The first approach uses the "mapping" algorithm, in which a computer program reads the text and classifies the words (or phrases) in the text into different categories based on some pre-defined rules or categories (i.e., dictionary). For instance, the General Inquirer, published by Harvard psychologist Philip J. Stone, and the Linguistic Inquiry and Word Count (LIWC) software by University of Texas psychologist James W. Pennebaker, are often used in content analysis.

The second approach, which is mostly pioneered by computer scientists and mathematicians, relies on statistical techniques to infer the content of documents and classify documents based on statistical inference (e.g., Manning and Schütze (1999) and Mitchell (2006)).[3] For instance, the algorithm may calculate the statistical correlation between the frequency of some keywords and the document type to draw inferences.

In this paper, I carry out the content analysis using the statistical approach, which offers several advantages. First, there is no readily available dictionary that is built for the setting of corporate filings—as a result, the dictionary-based approach may have low power for corporate filings. For instance, take the sentence "In addition, the Company has experienced attrition of its Medicare and commercial business in 1998 and 1999 and expects additional attrition." According to the General Inquirer, which can be invoked through http://www.webuse.umd.edu:9090/, the sentence above has 2 or 10.53% positive words ("expect" and "experience") and no negative words, even though it is obvious that this sentence has a negative tone. Second, the simple dictionary-based approach does not take into consideration the setting of a sentence. For instance, if a sentence is about the revenues, then "increase" should be treated as positive words; it is, however, likely to be of negative tone if the topic is "cost."

Third, the rule-based approach generally ignores any prior knowledge that researchers may have about the text. For example, if we know that most of the sentences that appear in MD&A reports are neutral, then, unless there is a strong evidence that a sentence is of negative tone, it might be more efficient to classify a random sentence as neutral tone. This point is especially salient when the topic of interest is managerial disclosures, because

[3]The only study that I am aware of in the economics, finance, and accounting field using this approach is Antweiler and Frank (2004).

managers have incentives to disclose strategically.[4] Finally, the statistical approach typically will provide a natural way to validate the classification efficiency using the training data. The training data are human coded and could be used to test the effectiveness of the algorithm.

In this paper, I rely on a specific type of statistical learning method—the Bayesian algorithm—to do content analysis. Under this method of document classification, a given sentence is first reduced to a list of words with each word weighted (by, say, its frequency in the sentence), *words*, and our goal is to classify it into a specific category *cat* from a set of all possible categories (*cats*). For instance, we may want to classify a sentence into a category from the set of four possible categories $cats = (positive, negative, neutral, uncertain)$. The Naïve Bayesian algorithm would choose the best category by solving the following problem:

$$cat^* = \underset{cat \in cats}{\operatorname{argmax}} \frac{P(words|cat)P(cat)}{P(words)}.$$

Since $P(words)$ doesn't change over the range of categories, we can get rid of it. So our new problem is:

$$cat^* = \underset{cat \in cats}{\operatorname{argmax}} P(words|cat)P(cat).$$

Finally, we note that if $w_1, w_2, ..., w_n$ are the words in the document and we further assume that their probability of showing up in a sentence is independent, then this expression is equivalent to:

$$cat^* = \underset{cat \in cats}{\operatorname{argmax}} P(w_1|cat) * P(w_2|cat) * ... * P(w_n|cat) * P(cat),$$

which is the formula used in the document categorization algorithm of this paper.

The last step is the only non-rigorous one in the derivation, and this is the "naïve" part of the Naïve Bayesian technique. It assumes that the probability of each word appearing in a document is unaffected by the presence or absence of each other word in the document. We assume this even though we know this isn't true: For example, the word "iodized" is far more likely to appear in a document that contains the word "salt" than it is to appear in a document that contains the word "subroutine"; likewise, in our setting, the words "adverse

[4] On the other hand, if the subject of analysis is news articles, there may not be a strong prior.

effect" are more likely to show up together with the word "material" in a sentence from a corporate annual report. Luckily, as it turns out, making this assumption even when it isn't true may have little effect on our results.[5] The independence assumption simplifies the computation and avoids the "curse of dimensionality" problem (Bellman (1961)).

4 Empirical implementation

4.1 Data preparation

I first download all the 10-Ks and 10-Qs filed between 1994 and 2007 from the SEC Edgar Web site and remove all the HTML tags if necessary. I then extract the MD&A section of the filings.[6] Next, I split the MD&A text into sentences using the Lingua::EN::Sentence module in Perl, which takes into consideration many acronyms in the splitting process.[7]

The next task is to extract all the forward-looking statements from the 10-Qs and 10-Ks. I define the forward-looking statements as all those sentences that contain: "will," "should," "can," "could," "may," "might," "expect," "anticipate," "believe," "plan," "hope," "intend," "seek," "project," "forecast," "objective," or "goal." I do not include the word "shall" in the searching process because it is usually associated with legal language and boilerplate disclosures. For instance, I randomly selected 5% (or 601 filings) of all the 10-Ks filed in 1998 by all public issuers in the U.S. with a file size greater than 10K bytes. These 601 filings contain 68,878 sentences with the word "shall" in it. On the other hand, these filings have 62,783 sentences that contain one of the "forward-looking" words listed above, which suggests that including "shall" in the search process will more than double the number of forward-looking statements search results. I then randomly checked 100 of the 68,878 sentences with the word "shall" and conclude that almost all of them can be classified as boilerplate disclosures and are not of interest to this paper. Hence, I "shall" not include

[5]See, for instance, http:///www.cs.washington.edu/homes/pedrod/mlj97.ps.gz.

[6]Based on a random check of 200 filings, the success rate of extracting MD&As is about 95% for 10-Qs and between 85% to 90% for 10-Ks. The details of the process are available upon request.

[7]For instance, in the string "FASB No. 123," the dot should not be treated as a delimiter for a sentence.

"shall" in the search for forward-looking statements.

I also exclude any sentence that contains the word "undersigned," sentences that consist of all capital letters, and sentences containing words such as "herein," "hereinafter," "hereof," "hereon," "hereto," "theretofore," "therein," "thereof," and "thereon," because they are almost certain to be legal boilerplate. Furthermore, I exclude all the sentences that contain "expected," "anticipated," "forecasted," "projected," or "believed" that follows "was," "were," "had," and "had been." Situations like these typically indicated a sentence that's not forward-looking in nature. The search for forward-looking statements has both type I and type II errors. I expect the type II errors to be small in this case; given the long list of words that are used in the search process, sentences that are not flagged as forward-looking are unlikely to be forward-looking. On the other hand, it's possible that some sentences are flagged as forward-looking may not be FLS.

I carry out the content and tone analysis at the sentence level. The disadvantage of sentence-level (rather than paragraph-level or document-level) analysis is that, in some occasions, the tone and content of a sentence depends on its context. For instance, without knowing the context, it's not possible to determine the tone of the sentence "The increase is primarily the result of lower-than-anticipated sales volume." This is because we don't know what (cost or other items) increases because of the low sales volume. However, there are several advantages to sentence-level analysis. First, it can significantly reduce the amount of labor in coding the text. Second, different sentences in a paragraph or article can have different tones and contents and bundling them together introduces noise. Sentence-level analysis can therefore increase the power of the classification.

Finally, before doing the Naïve Bayesian classification, I implement the "stemming" and "stopwording" processes to further clean up the text. Stemming is the process for reducing inflected (or sometimes derived) words to their base or root form (e.g., "dependent" to "depend") to increase the power of textual analysis. I use the Lingua::Stem::En module from Perl, which implements the Porter's stemming algorithm, invented by Martin Porter at Cambridge University and first described in Porter (1980). "Stopwords" are a class of words that are typically the short, frequently occurring words in a language. Typical stopwords usually have only a grammatical function within a sentence, and don't add to the meaning. Stopwords include articles, case particles, conjunctions, pronouns, auxiliary verbs, and common prepositions. Some examples of stopwords for English are: "the," "and," "it," "is," and "of." To do stopwords cleanup, I use the Lingua::EN::Stopwords modules from Perl. The stop-words list used in the Lingua::EN::StopWords has a list of 213 words (refer to "http://wiki.christophchamp.com/index.php/Perl/Modules/Lingua." for a complete list). I modify this list by deleting the following words from the stop words list: "cannot," "no," "none," "nor," and "not" (i.e., including them in the statistical analysis). The reason for this modification is that one of the goals of this paper is to categorize the tones of statements, and words like "cannot" can completely change the tone. Empirically, however, including or excluding these words in the analysis does not affect any of the results in this paper.[8]

4.2 Preparing the training dataset

To construct the training data, I manually classify 30,000 randomly selected forward-looking statements along the tone and content dimensions. First, every sentence is classified into one of four tones: positive, neutral, negative, and uncertain nature. The "uncertain" tone is added because prior research has shown that managers tend to convey negative information by using words like "risk" and "uncertainty" (Li (2007)). A typical uncertain tone sentence looks like the following: "Significant additional work will be required for the scaling-up of each new product prior to commercialization, and this work may not be completed successfully."

I also divide the contents of the forward-looking statements from corporate filings into 12 categories.

[8] This is likely due to the learning nature of the analysis. For example, suppose when managers talk about "material effect," most of the time they will include "no" or "not" in the same sentence, then, even if "no" and "not" are included in the stop-words list, the algorithm can still capture the positive tone by conditioning on the words "material effect."

- Category 1: Sales / revenues / market condition / market position / consumer demand / competition / pricing / new contract
- Category 2: Cost / expense / reserves for contingent liability / asset impairment / goodwill impairment
- Category 3: Profit / income / performance results / margin
- Category 4: Operations / productions / general business
- Category 5: Liquidity: interest coverage / cash balance / working capital conditions
- Category 6: Investment - general capital expenditure; M&A / divestiture / discontinued operation
- Category 7: Financing - debt / equity / dividend / repurchase
- Category 8: Litigation / lawsuit (material impact or not)
- Category 9: Employee relations / retention / hiring / union relations
- Category 10: Regulations (e.g., environment laws) / income tax / government relation
- Category 11: Accounting method / accounting estimation assumptions / auditing / internal control
- Category 12: Other: Boilerplate / legal statement / standard statement

The training data construction is done with the help of 15 research assistants; most of the research assistants are students of the Master of Accounting program, BBA program or MBA program at the University of Michigan.[9] To make sure that the training dataset is of high quality, I impose the following search criteria for research assistants: (1) Native English

[9] One student is from the School of Information at Michigan who took my Intermediate Financial Accounting course and got an A+; two students are Master of Accounting students at the Shanghai University of Finance and Economics with Canadian Certified General Accountants certificates.

speakers are given higher priority; (2) If a person is not a native English speaker, then she has to rank within the top 10% in TOEFL test or GMAT/GRE verbal test; and (3) the student should get A- or above in an intermediate financial accounting course or holds an American, British, or Canadian CPA certificate.

The manual classification process is challenging because many forward-looking statements cannot be easily categorized into a content or tone category. A particular challenge lies in the unobserved expectation of the financial statement readers. Suppose a manager discloses that the adverse effect of an environment liability will not be material for the firm. This can be positive news if the reader has been expecting a material effect or neutral if the immateriality is fully anticipated. As another example, an increase in capital expenditure may signal negative tone for shareholders if the company is already over-investing, but it could also be positive if the investment is part of a turnaround plan. Without a more specific context, it's almost impossible to figure out these subtle differences in the training data construction process. I ask the research assistants to keep a neutral prior—i.e., assuming that the reader has no information about the topic—in making the judgment. In the environment liability example above, the sentence would be classified as positive because I am assuming that the reader has no prior information about the potential environmental liability. In later empirical analysis, the contents and tones of the FLS are examined conditional on other observable variables such as contemporaneous stock returns. To the extent that any expectations of the readers/investors about the subject are reflected in these observable variables and, therefore, are controlled, my "uninformative prior" research design in the training process should be effective.

The descriptive statistics of the training data set are reported in Table 1. Of the 30,000 FLS that are manually classified, 19.59% of them are coded as of positive tone, 39.97% neutral, 17.82% negative, and 22.55% of uncertain nature. The distribution of the sentences across the four groups is not even—the largest group is the neutral tone group. Somewhat inconsistent with the findings in Pava and Epstein (1993), there is a fair amount of negative tone discussions in MD&As with a percentage close to the positive tone group. Prior research shows that managers tend to express their negative views using "risk" and "uncertain" type of words (Li (2007)). Therefore, the 22.55% sentences with uncertainty tones are also possibly of negative connotations. In the empirical analysis, I combine the uncertainty tone group into the negative tone group.

Table 1 also shows the percentages of the training data for each content category. The sum of the percentages across the twelve categories is greater than 100% because one sentence can be assigned to multiple categories. For instance, consider the statement "The Company believes these changes will help move existing inventory and reduce its cash flow concerns, but it is unlikely that these changes alone will provide sufficient capital to fund ongoing operations"; this sentence is about both operations and liquidity and, therefore, is classified as both category 4 and category 5.

The first four categories are about revenue, cost, profitability, and operations; combined together, 62.81% of the discussions are about these issues. Category 5 (liquidity issues) is discussed in 11.57% of the sentences, Category 6 (investment issues) 10.79%, and Category 7 (financing issues) 16.45%. These three categories are related to financing, liquidity, and capital resources and will be combined in later analysis. Other categories take a very small share of the MD&A discussions—ranging from Category 9 (employee relations, 1.41%) to Category 10 (regulation issues, 4.05%). The fact that most of the FLS are about profitability, liquidity, and capital resources—as evidenced by the combined 101.62% percentages in categories 1 to 7—is consistent with the SEC's intention that the MD&A section primarily serves to provide investors with information about capital resources and liquidity.

4.3 Computation

I use the Algorithm::NaiveBayes module in Perl to do the computation. I first convert the vector of words for each sentence after the stemming and stopwording processes into a hash variable in Perl. I then feed the hash variables from the 30,000 sentences that are manually coded into the Bayesian classifier module in Perl and run the training process. After this step, the algorithm goes on to predict the tone and category of all the FLS from 10-Ks and

10-Qs, which in total are about 13 million sentences.

4.4 Validation of the algorithm

There are three common methods to evaluate the effectiveness of a text classification algorithm:

- Training Error. The classifier is trained on a training data set and also evaluated on the same data set. It is a good idea to get this result, although it is obviously biased toward the training data. Therefore, this evaluation method can detect underfitting but not overfitting of the training data (i.e., this method has a tendency of overfitting the data).

- Train and test. The data are divided into two parts: training and testing part. The split is usually 90% for training and 10% for testing, but sometimes 2/3 of data is used for training and 1/3 for testing. This is an unbiased evaluation, which can detect underfitting as well as overfitting. To be sure that the evaluation is unbiased, it is important not to use testing data in any way, even to glance at it, if it may influence our decisions regarding classifier construction. With some methodologically generic methods, this is not an issue.

- N-fold cross-validation.[10] In this method, the data is randomly partitioned into N equal parts. N experiments are performed, and in each experiment, one part is taken as the testing data, while remaining $N-1$ parts are used for training. At the end, the results over the N experiments are averaged. This is unbiased testing that gives more statistical significance than train-and-test, but it is not applicable if we need to examine the training data during classifier construction.

[10] The theory of cross-validation was inaugurated by Geisser (1975). It is important in guarding against testing hypotheses suggested by the data ("Type III error"), especially where further samples are hazardous, costly, or impossible.

I first validate the effectiveness of the algorithm by calculating the training errors—carrying out both the training and predicting processes using all the 30,000 manually coded sentences. Untabulated results show that the training errors are in general less than 10%. This literally means that if the algorithm is used to predict the sentences in the training data set (after learning from the same data set), it'll correctly classify the categories and tones more than 90% of the time. This suggests that the chances of underfitting are small.

Since the "train and test" method is almost a special case of the N-fold cross validation test, I next report the empirical validation of the Naïve Bayesian learnings algorithm using the N-fold cross-validation method, with N varying from 3 to 50. For instance, when $N = 3$, I carry out the 3-fold cross-validation by randomly dividing the 30,000 sentences of training data into 3 equal parts with each part containing 10,000 sentences. Three tests are then performed with one part (10,000 sentences) used as the learning data to classify the other two parts (20,000 sentences). In the end, the average success rate of the three tests is reported as the 3-fold cross-validation result.

Table 2 reports the average correct classification rate for the different level of N. In row (1) of Table 2, the results show that when we do the 3-fold cross-validation test (i.e., $N = 3$) to predict four-category tones (i.e., positive, neutral, negative, and uncertain), the learning algorithm classified the testing sentences correctly 59.15% of the time. Suppose we use an informed guessing approach—i.e., we calculate the percentage of each of the four tones using the learning data and apply this probability naively to the sentences in the testing data—the correct classification rate is only 32.44%. This shows that the Bayesian learning algorithm almost doubles the success rate from the informed guessing strategy. If we combine the negative tone and uncertain tone into one category (so that the task is to classify the sentences into 3 categories), the Bayesian algorithm has a success rate of 66.95% in the 3-fold cross-validation test, and that of the informed guessing is 40.47%. The results are quite stable when N increases from 3 to 50—the success classification rate of the 4-category tone by the Bayesian algorithm is about 59% and that of the 3-category tone (i.e., positive, neutral, negative/uncertain tone) is about 67%, both are much higher than the informed guessing strategy (about 32% and 40% respectively).

To classify every FLS into a content category out of 12 possible categories as defined in Section 4.4, the Bayesian algorithm has a success rate of about 63% in the N-fold cross-validation, while that of the informed guessing strategy is only about 15%. If we combine the 12 content categories into 3 groups—profit (categories 1 to 4), liquidity (categories 5 to 7), and other (categories 8 to 12), the Bayesian algorithm can achieve a success rate of more than 82%, with the informed guessing rate being about 44%. Overall, the N-fold cross-validation tests show that the Bayesian learning algorithm achieves a good classification rate compared with a naive random strategy.

Next, to compare the effectiveness of the machine learning algorithm with the general dictionary-based approach (e.g., Kothari and Short (2006) and Davis, Piger, and Sedor (2005)), I take the 30,000 training sentences and feed them into both the General Inquirer dictionary software and LIWC software, two of the most popular general dictionaries for content analysis, to examine the classification success rates of these general dictionaries.

Table 3 shows the success rates of LIWC and GI for 1000, 2000, 5000, 10,000, 15,000 and 30,000 sentences of the training data. LIWC and GI analyze a block of text, which is one sentence in this case, and calculate the percentages of positive and negative words in the sentence; unlike the learning algorithm, the dictionaries in general do not classify each sentence into positive or negative tones directly. To classify the FLS sentences into different tone categories using the dictionary approach, I implement the following rules: A sentence is classified as positive tone if the percentage of positive tone words is greater than the negative tone words plus $BAND$, neutral if the percentage of positive words is in the interval [percentage of negative words $-BAND$, percentage of negative words $+BAND$], and negative if the percentage of positive words is less than the percentage of negative words minus $BAND$. $BAND$ is a variable that defines the "neutrality" zone, and I report the classification results by varying $BAND$ between 0 and 50. The classification of the dictionaries is then compared with the human coder to determine the success rate.

The results in Table 3 indicate that when $BAND = 0$ and we use all 30,000 sentences, the General Inquirer has a success rate of 32.56%, below the informed guessing strategy reported in Table 2, while LIWC has a success rate of 40.26%, comparable to informed guessing. As $BAND$ increases—meaning more and more sentences are now classified as neutral—the General Inquirer does much better. For instance, when $BAND = 50$ (i.e., when the percentage of positive words is between the percentage of negative words plus or minus 50, the sentence is classified as neutral), GI has a success rate of 39.99%. Overall, however, both the General Inquirer and LIWC yield a classification rate comparable to the informed guessing strategy. This is likely due to the fact that they both use very general dictionaries (good for, say, news media coverage) and may not work well for the corporate filings.

5 Information content of the MD&A FLS

5.1 Descriptive statistics on the MD&A tones and contents

To be included in the final data for further analysis, a firm-quarter has to have the following data: (1) a Central Index Key that can be matched with the GVKEY from Compustat and PERMNO from CRSP; (2) quarterly earnings (item 69 in the Quarterly file) and cash flows from operations (item 108) from Compustat; (3) stock returns in CRSP; and (4) at least five sentences of forward-looking statements in the filing. The requirement of at least five sentences of FLS is arbitrary, and the purpose is to make sure that the empirical measures derived from the filing are not due to some random noise. Varying this requirement (e.g., requiring at least 10 FLS sentences per filing) does not alter any of the empirical results.

For every forward-looking sentence k, I define its tone as the following: $TONE_k = 1$ if the learning algorithm predicts the sentence to be positive; $TONE_k = 0$ if neutral; and $TONE_k = -1$ if negative or uncertain. I combine the uncertain tone into the negative tone, because there are a large number of uncertain statements and are of negative implications.

For every firm i in quarter j, I then define the tone of a firm's MD&A forward-looking statements as the average tone of all the K forward-looking sentences in its 10-Q or 10-K filing for that quarter as predicted by the learning algorithm.

$$TONE_{ij} = \frac{1}{K}\sum_{k=1}^{K} TONE_{ij,k} \tag{1}$$

By construction, $TONE_{ij}$ is a variable that's between -1 and 1, with 1 being complete optimism and -1 being completely pessimistic tone. The more positive it is, the more optimistic the tone of the forward-looking statements made by firm i in quarter j's 10-Q or 10-K filing.

Table 4A shows the descriptive statistics of the MD&A tones and other variables of my final sample of 145,479 firm-quarters. On average, the forward-looking statements in MD&A disclosures are negative, as indicated by the mean (median) of $TONE$ of -0.23 (-0.21). The mean of $TONE$ is significantly different from 0 with a p-value of 0.000 in t-test. Decomposing $TONE$ into three different components—$PROFIT\_TONE$ (the average tones of sentences in categories 1 to 4 as described in Section 4.4), $LIQUIDITY\_TONE$ ((the average tones of sentences in categories 5 to 7), and $OTHER\_TONE$ (the average tones of sentences in categories 8 to 12)—shows that the negative tones of MD&A FLS is mainly because of the profitability-related sentences. The mean of $PROFIT\_TONE$ is -0.42, $LIQUIDITY\_TONE$ 0.16, and $OTHER\_TONE$ -0.26, all of which are statistically significantly different from 0.

Table 4A also presents the descriptive statistics of the contents of the MD&A forward-looking statements. $PROFIT\_PCT$ is the percentage of contents devoted to profitability and operations (categories 1 to 4) as predicted by the learning algorithm, and $LIQUIDITY\_PCT$ is the percentage on liquidity and capital resources. The mean (median) of $PROFIT\_PCT$ and $LIQUIDITY\_PCT$ are 53.65% (55.32%) and 32.99% (29.63%).

In Panel B of Table 4, the Pearson correlations of the tones and contents are reported. There is a significant negative correlation between $TONE$ and $PROFIT\_PCT$ (-0.47), indicating that when managers devote more discussions to future profitability, the average tones are more pessimistic. This is consistent with the negative mean of $PROFIT\_TONE$

documented in Table 4A. The positive correlation between $TONE$ and $LIQUIDITY\_PCT$ (0.52) shows that managers will devote more discussions to liquidity and capital resources issues when the $TONE$ of the MD&As are more positive. Current earnings ($EARN$) is positively correlated with MD&A tones (Pearson correlation coefficient 0.142), and firms with more positive accruals tend to have more pessimistic MD&A statements, with the correlation between ACC and $TONE$ being -0.082.

Figure 1 plots the mean of the MD&A tones for firms sorted into quintiles. In quarter 0, all firms are sorted into five quintiles based on $TONE$ of the 10-K or 10-Q filed for that quarter, with quintile 1 firms having the most pessimistic tones and quintile 5 the most optimistic. The tones of these firms are then tracked for the next 48 quarters with the mean plotted for each quintile. From Figure 1, it can be seen that $TONE$ is mean-reverting—by quarter 10, the differences in tone between the quintiles are dramatically reduced.

Figure 2 plots the aggregate tones of MD&As of U.S. public firms over time. Every month, the tones of the MD&As in the 10-Ks and 10-Qs filed by firms with Compustat and CRSP coverage in the month are calculated and the average tone for that month is plotted. The figure shows substantial variation in the aggregate tones. Three vertical lines indicate the three possible dates of interest. The first vertical line indicates March 2000, when the NASDAQ index peaks at the level of 4572.83—the figure shows that contrary to the dramatic increase in the index in the late 1990s, the average tone in the MD&A sections is actually going down. Several possible reasons may explain this phenomenon. First, because of the anxiety over the Y2000 issues, managers may devote more discussions on these issues in an uncertain tone, which leads to less positive aggregate tones. Second, managers might have been anticipating the equity market downturn and become more cautious in communicating to stock investors. A more careful examination of these explanations will be very interesting and beyond the scope of this paper.

The second vertical line in Figure 2 shows that after the September 11 attack, there seems to be a downward trend in management's tones, although it's not clear whether this reflects the revision in management expectation or it's simply a continuation of a previous trend. The third vertical line shows that the passage of the Sarbanes-Oxley Act does not seem to have an immediate impact on management's forward-looking tones.

5.2 Determinants of the MD&A tones and contents

Table 5 reports the OLS regression results of $TONE$ on its hypothesized determinants: $EARN$ (current earnings), RET (contemporaneous stock returns), ACC (accruals), $SIZE$ (the logarithm of market value of equity), MTB (market-to-book ratio), $RETVOL$ (return volatility), and three reporting quarter dummies $Q2$, $Q3$, and $Q4$. Year and 2-digit SIC industry fixed effects are also included in the regression. Since there are likely to be correlations for the same firm over time in the error terms, the standard errors are clustered at the firm-level.

The results indicate that the tones of FLS are positively related to current performance (coefficient on $EARN$=0.12 with a t-statistics of 8.29 and that on RET is 0.03 with t being 15.35), confirming the univariate correlation in Table 3. This means that when a firm is performing well in a quarter, managers tend to discuss future outlook in a more positive tone. Accruals are significantly negatively related to $TONE$ (coefficient on $ACCRUALS$ being -0.06 with t=-5.99)—suggesting that when current accruals are very positive, management's discussions of future outlook are more negative. Given that accruals are negatively related to future performance, this suggests that managers perhaps understand, at least to some extent, the implications of accruals for future performance.

Bigger firms have more negative MD&A tones, as indicated by the significant negative coefficient on $SIZE$ (-0.005 with a t-statistic of -4.29). This is consistent with the hypothesis that large firms are more cautious in disclosures due to political and legal costs concerns. Firms with high market-to-book ratio have less positive tones in MD&As, consistent with the hypothesis that growth firms have more uncertain information environment and are more conservative in discussing future events. More volatile firms tend to have less optimistic discussion about future outlook (the coefficient on $RETVOL$ is -0.39 with a t-value of -

20.84).[11] This suggests that firms with more volatile business environments may be more cautious in forward-looking disclosures because of the information uncertainty with regard to future performance or because of legal concerns. Two of the three reporting quarter dummies ($Q2$ and $Q4$) show up insignificantly—both statistically and economically. In particular, the variable $Q4$ captures the tone difference in the 4th quarter MD&A FLS discussions relative to the 1st quarter (coefficient 0.00 with a t-value of 0.24). This suggests that even though 10-Ks are in general much longer and more complicated than 10-Qs, their tones are not systematically different from other quarters. The third quarter dummy $Q3$ has a coefficient of -0.005 (t-statistic=-3.89), which is small in economic magnitude. Overall, most of the economic factors show up significantly in explaining MD&A tones in the hypothesized directions.

5.3 FLS tones and future earnings and liquidity

In this section, I examine the implications of the FLS tones generated by the Naïve Bayesian algorithm for firm's future performance beyond that contained in numeric financial information. First, I check for the link between $TONE$ and future profitability.

Table 6A shows the regression results of the earnings in the next four quarters scaled by the book value of assets at the end of this quarter ($EARN(t{+}1)$, $EARN(t{+}2)$, $EARN(t{+}3)$, and $EARN(t + 4)$) on $TONE$ and other control variables, which include all the variables examined as the determinants of $TONE$ and year and industry-fixed effects. The residuals from the four equations are likely to be correlated with each other. In unreported Seemingly Unrelated Regressions, when these cross-equation correlations are taken care of, the coefficients only become more significant statistically.[12]

In column (1), the coefficient on $TONE$ is 0.010 with t-statistic of 7.72, suggesting that

[11] This is also consistent with the findings in Dichev and Tang (2007).

[12] The coefficients remain the same because all three equations contain the same explanatory variables. The statistical significance in the SUR setting, however, is much bigger. This is likely due to the fact that, in an SUR setting, it's not easy to implement the usual adjustment for within-firm correlation in error terms and for heteroskedasticity.

when managers are more optimistic in discussing future outlook in the MD&As, the earnings in the next quarter are significantly higher. The economic magnitude of this effect is quite substantial—next quarter's earnings scaled by book value of assets of firms with extremely optimistic FLS (i.e., $TONE = 1$) is higher than that of firms with extremely pessimistic FLS (i.e., $TONE = -1$) by two percentage points (0.010×2). This translates into an annual difference in ROA of eight percentage points—a huge difference for a return-on-asset metric.[13] A different and more realistic angle to look at this effect is based on the inter-quartile range of $TONE$. From Table 4A, the 25th percentile and 75th percentile of $TONE$ are -0.41 and -0.03 respectively—thus, an inter-quartile change in $TONE$ implies a difference in annual ROA of 1.5 percentage points, even after controlling for the level of current stock returns and accruals. The same positive relation exists for $TONE$ when it is used to predict earnings in the next four quarters. In columns (2) to (4), the coefficients on $TONE$ are 0.009 (t=6.17), 0.009 (t=5.41), and 0.005 (t=2.71), respectively, showing that $TONE$ has predictive power for future earnings at least four quarters after current quarter although this effect becomes smaller economically and statistically over time. Not surprisingly, the coefficients on current earnings and stock returns are both positive in predicting the earnings in the next four quarters. Current quarter accruals, on the other hand, have a significantly negative relation with future performance, consistent with prior literature (Sloan (1996)).

In Table 6B, the dependent variable is the change (rather than level) in earnings in the next four quarters. For instance, in column (1), the dependent variable is $DEARN(t+1)$, which is calculated as the earnings in the next quarter minus this quarter's earnings scaled by the book value of assets at the end of this quarter. The implications of $TONE$ for the change in future earnings are quite substantial both economically and statistically. For example, the coefficient on $TONE$ in column (1) of Table 6B is 0.010 (t=7.27), the same as that in Table 6A.

Next, I examine whether $TONE$ is systematically related to future liquidity. There are many financial ratios that measure firm liquidity. In the analysis, I primarily focus on cash

[13]Technically, this is not ROA because the numerator is the net income, rather than unleveraged income.

flow ratio, defined as operating cash flows divided by current liabilities. Other common ratios include current ratio and interest coverage ratio. The disadvantage of using current ratio is that, as a balance sheet measure, it tends to be quite sticky and is not the ideal measure of what "happens" in a future quarter. The disadvantage of the interest coverage ratio, on the other hand, is that fewer firms on Compustat have interest expense reported and small interest expense may also lead to a scaling problem. The empirical results using current ratio and interest coverage ratio to measure liquidity, however, are almost identical and are not reported here.

In column (1) of Table 7, where the dependent variable is $CFRATIO(t+1)$, the operating cash flows in the next quarter divided by the current liabilities at the end of the next quarter, the coefficient on $TONE$ is 0.280 with a t-statistics of 9.85. The economic magnitude of this coefficient is substantial—the inter-quartile range of the cash flow ratio is about 0.50 and increasing $TONE$ from the 25th to the 75th percentile implies an increase in next quarter's cash flow ratio by 0.11, after controlling for other economic factors that may affect future liquidity. In columns (2) to (4) of Table 7, where the cash flow ratios from the next three quarters are the dependent variables, there is still significantly positive association between $TONE$ and future liquidity with substantial economic magnitude, although the size of the coefficient on $TONE$ is decreasing when the forecasting horizon increases.

The control variables in predicting future cash flow ratio do not include the current cash flow ratio, because current earnings and current accruals are included in the regression as explanatory variables and further including a cash flow ratio makes it difficult to interpret the coefficients. In unreported results, I include the current cash flow ratio instead of accruals as a control variable, and the results remain essentially the same. Of all the control variables, current earnings and stock returns can predict future liquidity positively, current earnings are negatively associated with future liquidity. Bigger firms also tend to have better cash flow ratios, and firms with more volatile business have lower cash flow ratios.

Do positive and negative tones have symmetric implications about future economic performance? Or are the implications asymmetric? The results in Table 8 shed light on this question—in Table 8, $TONE$ is interacted with a dummy variable $PTONE$, which is set to one if $TONE \geq 0$ and zero otherwise. The table only shows the empirical results using next quarter's earnings and liquidity, but the unreported results based on the next three quarters' numbers are qualitatively similar.

The results in Table 8 indicate that there is no non-linear relation between $TONE$ and future earnings, as evidenced by the insignificant coefficient on $TONE \times PTONE$ (-0.001 with -0.45). This suggests that the positive and negative tones of MD&As have about the same implications for future earnings. In column (2), the coefficient on $TONE \times POS$ is -0.155 with a t-statistic of -2.42 and the sum of the coefficients on $TONE$ and $TONE \times PTONE$ is 0.143 with F-value of 10.23 (p-value=0.007). The results thus suggest that while both the positive tones and negative tones FLS have information content about future liquidity, the implications of positive tones FLS for future liquidity are about half the size of those of negative tone FLS.

5.4 Information content of MD&A over time

To assess whether the information content of MD&As changed over time, Tables 9A shows the regression of future earnings and liquidity on $TONE$ and its interaction with year. A significant positive (negative) interaction term indicates that MD&As have become more (less) informative over time. Of the 8 regressions, 7 show a negative interaction term and column (6) is the only one with a positive interaction term. However, none of the interaction coefficients is statistically significant. This result suggests that, despite the continuous effort by the SEC to strengthen the disclosure requirements for MD&A, there is no systematic change in its information contents over time.

In Table 9B, $TONE$ is interacted with a dummy variable $POST2003$, which is equal to 1 if the report is filed in or after 2003 and 0 otherwise. This test is designed to capture any systematic change in the information content of MD&As after the new SEC guidelines on MD&A and the Sarbanes-Oxley Act, which significantly enhanced MD&A disclosure. Six of the eight interaction terms of $POST2003$ with $TONE$ are negative and two of them are statistically significant. The other two interaction terms are positive but insignificant. Overall, this shows that there is no significant increase in the informativeness of MD&As after 2003.

5.5 Analysts forecasts and MD&A tones

In this section, I examine whether sell-side financial analysts understand the tones of the forward-looking statements in MD&As. I obtain the analysts' forecasts of quarterly earnings from IBES. To examine whether financial analysts understand the tones of MD&As, I regress $ANALYST\_FCST$, the latest earnings forecasts by analysts before the announcement of next quarter's earnings, on $TONE$ and other control variables. Results in Table 10 show that analysts' forecasts, indeed, are higher if the tones of the MD&A FLS are more positive, as indicated by the coefficient in column (1) on $TONE$ (0.006 with a t-statics of 7.21).

To examine the more interesting question of whether analysts fully understand the MD&As, I regress the realized future earnings on $TONE$ and the analysts' forecast of the earnings number. If $TONE$ remains significantly positive, then the results indicate some inefficiency of analysts' forecasts of utilizing the information in MD&As. The results in Table 10 confirm this hypothesis. In column (2), I include both $TONE$ and $ANALYST\_FCST$ to explain next quarter's earnings, where $ANALYST\_FCST$ is the latest consensus analyst forecast before the announcement of next quarter's earnings. Note that this actually gives several months for analysts to respond to the information in MD&As, because the latest forecasts are at most one month before the announcement date, while the filing of this quarter's 10-Q or 10-K is likely to be several months before next quarter's earnings announcement. Not surprisingly, the coefficient on $ANALYST\_FCST$ is positive and significant (coefficient 0.663 with a t-value of 12.16). $TONE$, however, remains positive and significant—its coefficient is 0.005, about half of the case when analysts forecasts are not controlled, with a t-statistics of 4.40. The results in column (3) show similar results with next quarter's liquidity level. Therefore, the evidence seems to suggest that sell-side analysts do understand the implications of MD&As for future earnings, but not to a full extent—it seems that they understand about 50% of the implications.

5.6 Refining FLS tones analysis

In this section, I first examine whether conditioning the tones of MD&As on the content (i.e., profitability versus liquidity) could make a difference in forecasting future performance. I include the tones of the profitability-related tones ($PROFIT\_TONE$), liquidity-related tones ($LIQUIDITY\_TONE$), and the tones of other sentences ($OTHER\_TONE$) separately in the statistical analysis. Table 11 reports the regression of $EARN(t+1)$ and $CFRATIO(t+1)$ on this quarter's tones and other control variables. In column (1), the coefficient on $PROFIT\_TONE$ is 0.003 (t=4.20), $LIQUIDITY\_TONE$ 0.013 (t=10.44), and $OTHER\_TONE$ -0.001 (t=-0.96). This suggests that both profitability-related FLS and liquidity-related FLS have implications for future earnings, while other types of FLS do not contain much information. What's surprising is that in predicting future earnings, the effect of liquidity-related FLS is much bigger than that of profitability-related FLS. In column (2), where the dependent variable is $CFRATIO(t+1)$, similar effects are observed: Liquidity-related FLS have more predictive power in forecasting future liquidity situations than profitability-related FLS.

Second, I compare the information content of 10-K MD&As with those of 10-Qs by separating the sample into 10-Qs and 10-Ks in Table 12. Column (1) reports the results using the 10-Q sample—the coefficient on $TONE$ in predicting next quarter's earnings is 0.008 (t=5.51). In column (2), where the 10-K filed after the fourth quarter is used, the coefficient on $TONE$ is 0.005 (t=3.19). Thus, it appears that the economic magnitude of the implications of 10-Q FLS is slightly bigger than that of the 10-K filings. The difference, however, is not statistically significant. Column (4) reports the regression results of $CFRATIO(t+1)$ on $TONE$ and other variables using 10-Q filings only, and the coefficient on $TONE$ is 0.248 with a t-statistic of 7.83. In column (5), where 10-K filings sample is used, the coefficient is 0.092 (t=4.76), which is of substantial economic implications but is much lower than that in column (4). Column (6) shows that this difference is also statistically significant. Overall, it appears that, while both have substantial information content about future earnings and liquidity, 10-Q MD&As tend to have more information content than those from 10-K filings. One explanation for the finding is that in annual reports, management may be more likely to discuss longer-term events and the implications for short-term earnings and liquidity are therefore not as strong.

6 Conclusions

This paper examines the implications of the forward-looking statements from the MD&A section of corporate 10-Q and 10-K filings for future performance. I use a Bayesian machine learning algorithm to categorize the tone and content of the forward-looking statements from more than 140,000 corporate 10-Q and 10-K filings between 1994 and 2007. I find that the tone of the forward-looking statements is a function of current performance, accruals, size, market-to-book, and return volatility. The tone of the forward-looking statements is positively correlated with future performance and has explanatory power incremental to other variables, but the informativeness of MD&As has not changed systematically over time despite continous efforts from the SEC to strengthen MD&A disclosures. Financial analysts understand partially the information content of MD&As for future performance. Methodologically, the machine learning approach seems to perform better in analyzing corporate filings than a general dictionary-based approach.

References

Antweiler, Werner, and Murray Z. Frank, 2004, Is all that talk just noise? The information content of internet stock message boards, *Journal of Finance*.

Bainbridge, Stephen M., 2007, *A Complete Guide to Sarbanes-Oxley: Understanding How Sarbanes-Oxley Affects Your Business* (Adams Media) first edn.

Barron, Orie E., Charles O. Kile, and Terrence B. O'Keefe, 1999, MD&A quality as measured by the SEC and analysts' earning forecasts, *Contemporary Accounting Research* 16, 75–109.

Bellman, Richrad E., 1961, *Adaptive Control Processes* (Princenton University Press: Princenton, NJ).

Botosan, Christine A., 1997, Disclosure level and the cost of equity capital, *The Accounting Review* 72, 323–349.

———, and Mary S. Harris, 2000, Motivations for a change in disclosure frequency and its consequences: An examination of voluntary quarterly segment disclosures, *Journal of Accounting Research* 38, 329–353.

Botosan, Christine A., and Marlene A. Plumlee, 2002, A re-examination of disclosure level and the expected cost of equity capital, *Journal of Accounting Research* 40, 21–40.

Bryan, Stephen H., 1997, Incremental information content of required disclosures contained in management discussion and analysis, *The Accounting Review* 72, 285–301.

Butler, Marty, Andrew J. Leone, and Michael Willenborg, 2004, An empirical analysis of auditor reporting and its association with abnormal accruals, *Journal of Accounting and Economics* 37, 139–165.

Callahan, Carolyn M., and Rodney E. Smith, 2004, Firm performance and management's discussion and analysis disclosures: An industry approach, Working paper, University of Arkansas-Fayetteville.

Core, John, Wayne Guay, and David Larcker, 2008, The power of the pen and executive compensation, *Journal of Financial Economics* 88, 1–25.

Core, John E., 2001, A review of the empirical disclosure literature: Discussion, *Journal of Accounting and Economics* 31, 441–456.

Das, Somnath, and Pervin K. Shroff, 2002, Fourth quarter reversals in earnings changes and earnings management, Working paper, University of Minnesota.

Davis, Angela K., Jeremy M. Piger, and Lisa M. Sedor, 2005, Beyond the numbers: An analysis of optimistic and pessimistic language in earnings press releases, Working paper, Washington University.

Dichev, Ilia, and Wei Tang, 2007, The link between earnings volatility and earnings predictability, Working paper, University of Michigan.

Dietrich, J. Richard, Steven J. Kachelmeier, Don N. Kleinmuntz, and Thomas J. Linsmeier, 1997, An experimental investigation of forward-looking non-financial performance disclosures, Working paper, University of Illinois at Urbana-Champaign and University of Texas at Austin.

Geisser, Seymour, 1975, The predictive sample reuse method with applications, *Journal of The American Statistical Association* 70, 320–328.

Grundfest, Joseph A., and Michael A. Perino, 1997, Ten things we know and ten things we don't know about the Private Securities Litigation Reform Act of 1995: Testimony Before the Subcommittee on Securities of the Committee on Banking, Housing, and Urban Affairs of the United States Senate, http://securities.stanford.edu/research/articles/19970723sen1.html.

Healy, Paul M., and Krishna G. Palepu, 2001, Information asymmetry, corporate disclosure, and the capital markets: A review of the empirical disclosure literature, *Journal of Accounting and Economics* 31, 405–440.

Hüfner, Bernd, 2007, The SEC's MD&A: Does it meet the informational demands of investors? A conceptual evaluation, *Schmalenbach Business Review* 59, 58–84.

Hutton, Amy, Gregory Miller, and Douglas Skinner, 2003, The role of supplementary statements with management earnings forecasts, *Journal of Accounting Research* 41, 867–890.

Knutson, Peter H., 1993, *Financial Reporting in 1990's and Beyond* (NY: Association for Investment Management and Research: Charlottesville, VA).

Kothari, S.P., and James E. Short, 2006, The effect of disclosures by management, analysts, and financial press on the equity cost of capital: A study using content analysis, *Working paper* MIT.

Lang, Mark, and Russell Lundholm, 1993, Cross-sectional determinants of analyst ratings of corporate disclosures, *Journal of Accounting Research* Autumn, 246–271.

———, 1996, Corporate disclosure policy and analyst behavior, *Accounting Review* 71, 467–492.

———, 2003, Voluntary disclosure and equity offerings: Reducing information asymmetry or hyping the stock?, *Contemporary Accounting Research* 17, 623–662.

Leuz, Christian, and Robert E. Verrecchia, 2000, The economic consequences of increased disclosure, *Journal of Accounting Research* 38, 91–124.

Levine, Carolyn B., and Michel J. Smith, 2006, Critical accounting policy disclosures, Working paper, Carnegie Mellon University and Boston University.

Li, Feng, 2007, Do stock market investors understand the risk sentiment of corporate annual reports?, University of Michigan Working Paper.

——— , 2008, Annual report readability, current earnings, and earnings persistence, *Journal of Accounting and Economics.*

Manning, Chris, and Hinrich Schütze, 1999, *Foundations of Statistical Natural Language Processing* (The MIT Press) first edn.

Martindale, Colin, 1990, *The Clockwork Muse: The Predictability of Artistic Change* (Basic Books: New York).

Mayew, William J., and Mohan Venkatachalam, 2008, The power of voice: Managerial affective states and future firm performance, Working paper, Duke University.

Miller, Gregory S., 2002, Earnings performance and discretionary disclosure, *Journal of Accounting Research* 40, 173–204.

——— , and Joseph D. Piotroski, 2000, Forward-looking earnings statements: Determinants and market response, Working paper, Harvard University and University of Chicago.

Mitchell, Tom M., 2006, The discipline of machine learning, Working paper, Carnegie Mellon University CMU-ML-06-108.

Muslu, Volkan, Suresh Radhakrishnan, K. R. Subramanyam, and Dongkuk Lim, 2008, Causes and consequences of forward looking disclosures in the management discussion and analysis (MD&A) sections: An exploratory study, Working paper, University of Taxas at Dallas and University of Southern California.

Patell, J., 1976, Corporate forecasts of earnings per share and stock price behavior: Empirical tests, *Journal of The Accounting Research* pp. 246–276.

Pava, Moses L., and Marc J. Epstein, 1993, How good is MD&A as an investment tool?, *Journal of Accountancy* 175, 51–53.

Penman, Stephen, 1980, An empirical investigation of the voluntary disclosure of corporate earnings forecasts, *Journal of Accounting Research* pp. 132–160.

Pennebaker, James W., and Lori D. Stone, 2001, Contribution to a CATA workshop presented at the 51st annual meeting of the International Communication Association, .

Porter, Martin F., 1980, An algorithm for suffix stripping, program, *Program* 14, 130–137.

Pownell, Grace, Charles Wasley, and Gregory Waymire, 1993, The stock price effect of alternative types of management earnings forecasts, *Accounting Review* pp. 869–912.

Rogers, Rodney K., and Julia Grant, 1997, Content analysis of information cited in reports of sell-side financial analysts, *Journal of Financial Statement Analysis* 3, 17–30.

SEC, 1968, Securities Act Release No. 4936 (December 9), .

———, 1974, Securities Act Release No. 5520 (August 14), .

———, 1980, Securities Act Release No. 6231 (September 2), .

———, 2003, Commission Guidance Regarding Management's Discussion and Analysis of Financial Condition and Results of Operations; Interpretation; Rule, .

———, 2008, General Instructions for Form 10-K, .

Skinner, Douglas J., 1994, Why firms voluntarily disclose bad news, *Journal of Accounting Research* 32, 38–60.

Sloan, Richard, 1996, Do stock prices fully reflect information in accruals and cash flows about future earnings?, *Accounting Review* 71, 289–315.

Tavcar, Lawrence R., 1998, Make the MD&A more readable, *CPA Journal Online* http://nysscpa.org/cpajournal/1998/0198/newsViews/0198NV7.htm (1.12.2005).

Tetlock, Paul, 2007, Giving content to investor sentiment: The role of media in the stock market, *Journal of Finance*.

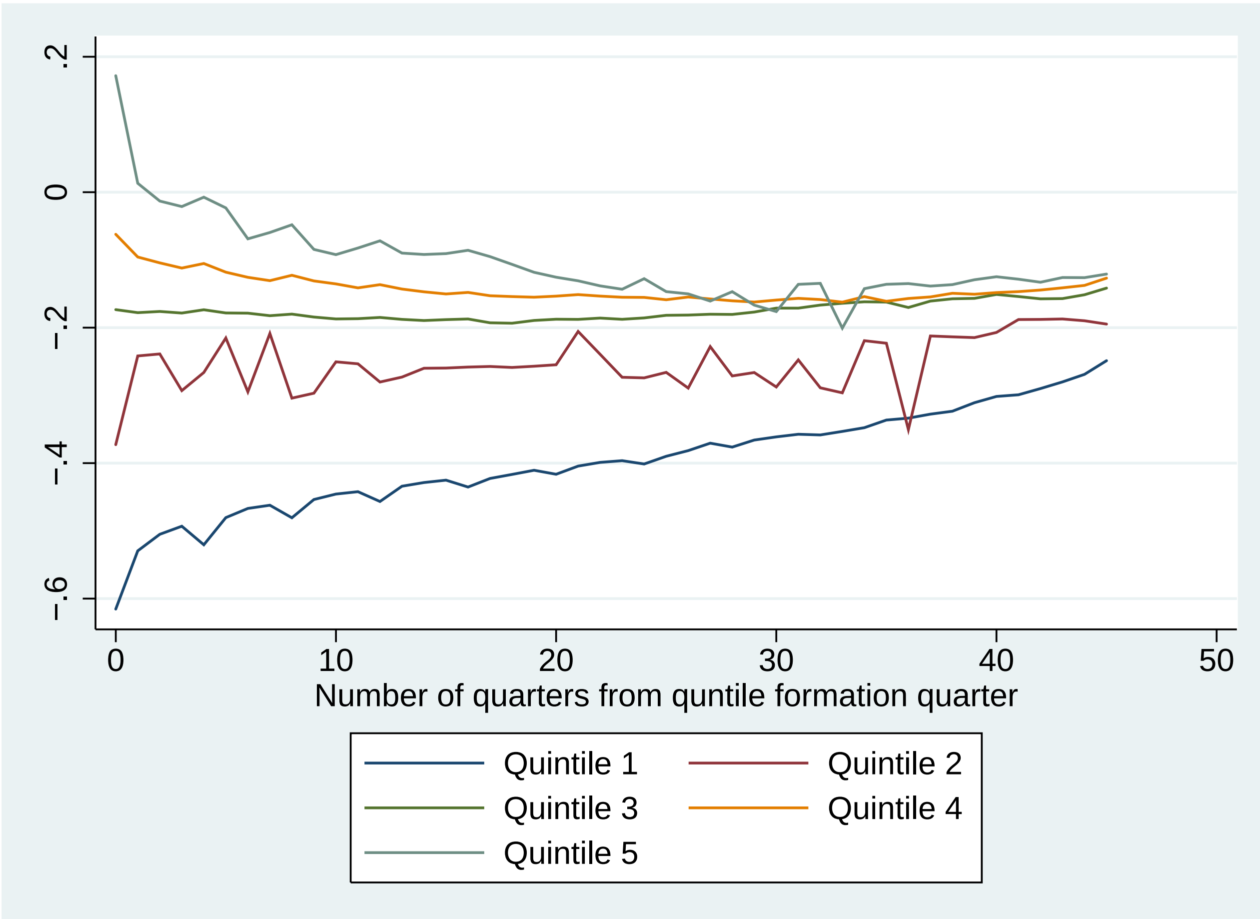
———, M. Saar-Tsechansky, and S. Macskassy, 2007, More than words: Quantifying language to measure firm's fundamentals, *Journal of Finance*.

Verrecchia, Robert E., 1983, Discretionary disclosure, *Journal of Accounting and Economics* 5, 179–194.

Watts, Ross L., and Jerry L. Zimmerman, 1986, *Positive Accounting Theory* (Prentice-Hall: Englewood Cliffs, NJ).

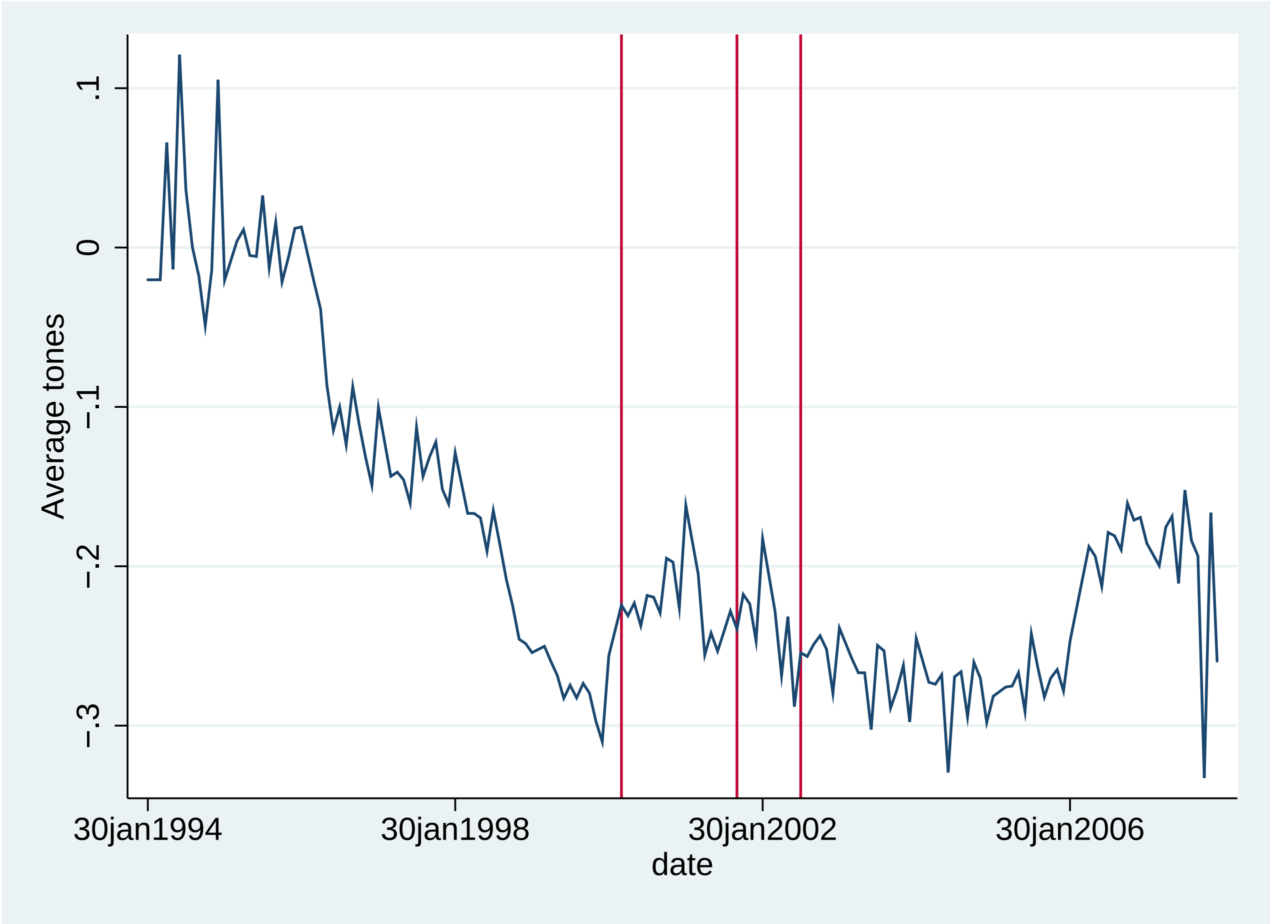
Waymire, Gregory, 1984, Additional evidence on the information content of management earnings forecasts, *Journal of Accounting Research* 22, 703–719.

Figure 1: MD&A tones over time by quintile



Notes: This figure shows the average tones of all MD&As of the 10-Ks and 10-Qs formed in quintiles. In quarter 0, firms are sorted into 5 quintiles (quintile 1 to quntile 5, with quintile 1 firms having the most negative MD&A tones and quintile 5 firms having the most positive tones) and the average tones of the five quintile firms are plotted over time for the next 48 quarters.

Figure 2: Aggregate tones of MD&As over time



Notes: This figure shows the average tones of all MD&As of the 10-Ks and 10-Qs filed in a given month between January 1994 and December 2007. The first vertical line indicates the month of March 2000 (when the NASDAQ index peaks); the second indicates the month of September 2001 (the terrorist attack event); the third indicates the month of July 2002 (Sarbanes-Oxley Act passage).

Table 1: Percentage distributions of MD&A FLS tones and contents

Positive tone	19.59	Cetegory 1: Revenues	15.06
Neutral tone	39.97	Category 2: Cost	10.45
Negative tone	17.82	Category 3: Profits	8.72
Uncertain tone	22.55	Category 4: Operations	28.58
		Sum of 1-4	*62.81*
		Category 5: Liquidity	11.57
		Category 6: Investing	10.79
		Category 7: Financing	16.45
		Sum of 5-7	*38.81*
		Category 8: Litigation	2.14
		Category 9: Employees	1.41
		Category 10: Regulation	4.05
		Category 11: Accounting	2.78
		Category 12: Other	3.32
		Sum of 8-12	*13.70*

Notes: This table shows the percentage distributions of the 30,000 sentences (i.e., the training data) that are manually coded into different tones and contents categories.

Table 2: N-fold cross-validation

Bayesian learning				
	Tone		Content	
N	4 categories	3 categories	12 categories	3 categories
3	59.15	66.95	62.52	82.31
5	59.30	67.00	62.76	82.37
10	59.31	67.02	62.91	82.42
25	59.27	66.99	62.88	82.40
50	59.37	67.11	63.02	82.46

Informed guessing				
	Tone		Content	
N	4 categories	3 categories	12 categories	3 categories
3	32.44	40.47	15.21	44.64
5	32.05	40.17	15.54	44.50
10	32.25	40.19	15.47	44.51
25	32.22	40.26	15.92	44.37
50	31.77	39.80	15.32	44.25

Notes: This table reports the N-fold cross-validation test results for the machine learning algorithm. For each N, the 30,000 training sentences that are manually coded are randomly divided into N equal parts. N experiments are then carried out, with $N-1$ parts used as learning data to classify the remaining data, and the average percentage of correct classifications is reported for each N. For the tones, the algorithm classifies each sentence into 4 possible categories (positive, neutral, negative, and uncertain) and the N-fold tests for the 4 categories are reported under "Tone—4 categories"; Under "Tone—3 categories," the negative and uncertain categories are combined together. For the contents, the algorithm classify each sentence into 12 possible categories (details in Section 4.4) and the N-fold tests for the 12 categories are reported under "Content—12 categories"; Under "Contents—3 categories," Categories 1 to 4 are combined together as "profitability" category; Categories 5 to 7 are combined together as "liquidity" category; Categories 8 to 12 are combined together as "other" category.

For each N-fold cross-validation tests, a benchmark success rate is calculated using an informed guessing strategy. In this strategy, the program calculates the probability for each tone/content category using the learning data (i.e., $(N-1)/N$ of the 30,000 sentences) and randomly classifies each data in the predicting set (i.e., $1/N$ of the 30,000 sentences) following these probabilities.

Table 3: Percentage rate of success of General Inquirer and LIWC in classifying the training data

General Inquirer								
N	band=0	band=1	band=3	band=5	band=10	band=20	band=30	band=50
1000	33.62	32.74	35.32	37.71	38.64	38.50	39.46	37.04
2000	33.18	33.27	33.12	37.51	39.45	40.70	40.35	40.86
5000	32.52	32.88	33.18	36.65	40.21	39.68	39.65	40.20
10000	31.93	32.69	34.09	36.73	39.36	39.72	40.36	39.80
15000	33.09	33.21	33.98	36.47	39.38	39.85	40.08	39.75
30000	32.56	32.61	33.62	36.39	39.27	39.96	39.98	39.99

LIWC								
N	band=0	band=1	band=3	band=5	band=10	band=20	band=30	band=50
1000	37.69	41.33	39.60	39.37	38.47	41.31	38.41	40.26
2000	41.12	39.91	40.57	41.04	38.91	41.91	38.17	39.32
5000	38.99	39.82	39.42	39.20	40.86	39.86	39.48	40.33
10000	39.95	39.80	39.76	39.63	39.21	40.58	39.93	40.02
15000	40.24	40.35	40.31	40.06	39.77	39.90	40.38	40.45
30000	40.26	40.28	40.29	40.10	39.95	39.98	39.99	39.99

Notes: This table shows the success rate of using the General Inquirer and LIWC softwares to classify the 30,000 training sentences into different tones. A sentence is considered to be classified as positive tone by General Inquirer or LIWC if the percentage of positive words is greater than the percentage of negative words plus $BAND$, which varies between 0 and 50 in different columns; neutral tone if the percentage of positive words is between the interval of the percentage of negative words minus $BAND$ and the percentage of negative words plus $BAND$; and negative tone if the percentage of positive words is less than the percentage of negative words minus $BAND$. The classification by the General Inquirer or LIWC of a sentence is then compared with the classification by the human coder to calculate the percentage of successful classification. N is the number of sentences that are chosen randomly from the 30,000 sentences for the analysis: This exercise is performed for all the 30,000 training sentences, and for 1000, 2500, 5000, 10000, and 15000 randomly selected sentences from the 30,000 sentences.

Table 4A: Descriptive statistics

N	Mean	Pr(=0)	P5	P25	Median	P75	P95	STDEV
tone	-0.23	0.000	-0.75	-0.41	-0.21	-0.03	0.23	0.29
profit_tone	-0.42	0.000	-0.94	-0.67	-0.44	-0.21	0.14	0.35
liquidity_tone	0.16	0.000	-0.38	-0.03	0.13	0.33	1.00	0.35
other_tone	-0.26	0.000	-1.00	-0.50	-0.13	0.00	0.20	0.40
profit_pct (%)	53.65	-	18.18	40.00	55.32	68.09	84.00	19.93
liquidity_pct (%)	32.99	-	8.33	18.37	29.63	44.44	69.23	19.03
other_pct (%)	13.35	-	0.00	4.17	12.00	20.00	33.33	11.47
earn	-0.02	-	-0.17	-0.02	0.01	0.02	0.05	0.13
ret	.04	-	-0.44	-0.14	0.01	0.16	0.59	0.38
cfratio	0.00	-	-1.76	-0.12	0.11	0.39	1.15	1.16
acc	-0.02	-	-0.17	-0.07	-0.02	0.01	0.16	0.16
size	5.54	-	2.41	4.10	5.46	6.86	9.02	2.01
mtb	2.17	-	0.80	1.08	1.46	2.30	5.79	2.47
retvol	0.16	-	0.05	0.09	0.13	0.20	0.37	0.12
	P-value							
Test of profit_tone=liquidity_tone	0.000							
Test of profit_tone=other_tone	0.000							
Test of liquidity_tone=other_tone	0.000							

Notes: This table shows the descriptive statistics for 145,479 sample firm-quarters. *tone* is the average tone of the forward-looking statements of a firm-quarter. A forward-looking sentence's tone has a value of 1 if the learning algorithm classifies the sentence as positive, 0 if neutral, and -1 if negative or uncertain. *profit_tone* is the average tone of the forward-looking statements of a firm-quarter that are about profits or operations (i.e., the statements that are classified as categories 1 to 4 as defined in Section 4.4). *liquidity_tone* is the average tone of the forward-looking statements of a firm-quarter that are about liquidity or capital resources (i.e., the statements that are classified as categories 5 to 7 as defined in Section 4.4). *other_tone* is the average tone of the forward-looking statements of a firm-quarter that are about other topics (i.e., the statements that are classified as categories 8 to 12 as defined in Section 4.4). *profit_pct* is the percentage of the forward-looking statements of a firm-quarter that are about profits or operations (i.e., the statements that are classified as categories 1 to 4 as defined in Section 4.4). *liquidity_pct* is the percentage of the forward-looking statements of a firm-quarter that are about liquidity or capital resources (i.e., the statements that are classified as categories 5 to 7 as defined in Section 4.4). $other_pct$ is the percentage of the forward-looking statements of a firm-quarter that are about other topics (i.e., the statements that are classified as categories 8 to 12 as defined in Section 4.4).

earn is the quarterly earnings scaled by book value of assets, winsorized at -3 and 3. *ret* is the contemporaneous stock returns in the fiscal quarter. *cfratio* is the quarterly cash flows from operations scaled by book value of current liability. *acc* is the accruals scaled by the book value of assets. *size* is the logarithm of the market value of equity at the end of the quarter. *mtb* is the market value of equity plus the book value of total liabilities scaled by the book value of total assets. *retvol* is the stock return volatility calculated using twelve months of monthly data before the fiscal quarter ending date.

Table 4B: Correlation matrix (p-values in parentheses)

Variables	tone	profit_tone	liquidity_tone	other_tone	profit_pct	liquidity_pct	earn	ret	acc	size	mtb	retvol
tone	1.000											
profit_tone	0.779	1.000										
	(0.000)											
liquidity_tone	0.541	0.279	1.000									
	(0.000)	(0.000)										
other_tone	0.486	0.262	0.231	1.000								
	(0.000)	(0.000)	(0.000)									
profit_pct	-0.467	-0.206	0.028	-0.140	1.000							
	(0.000)	(0.000)	(0.000)	(0.000)								
liquidity_pct	0.520	0.230	0.008	0.215	-0.828	1.000						
	(0.000)	(0.000)	(0.003)	(0.000)	(0.000)							
earn	0.142	0.103	0.159	0.065	-0.028	0.017	1.000					
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)						
ret	-0.005	-0.008	-0.001	-0.000	-0.013	-0.002	0.077	1.000				
	(0.043)	(0.003)	(0.841)	(0.873)	(0.000)	(0.495)	(0.000)					
acc	-0.082	-0.066	-0.078	-0.045	0.020	0.001	0.118	0.004	1.000			
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.604)	(0.000)	(0.126)				
size	0.040	0.062	-0.015	-0.013	-0.042	-0.026	0.213	0.103	-0.107	1.000		
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)			
mtb	-0.169	-0.122	-0.114	-0.078	0.097	-0.087	-0.200	0.208	0.150	0.152	1.000	
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)		
retvol	-0.263	-0.204	-0.176	-0.091	0.154	-0.110	-0.252	0.164	0.110	-0.323	0.198	1.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	

Table 5: Determinants of MD&A tones

COEFFICIENT	TONE
earn	0.119***
	(8.29)
ret	0.034***
	(15.35)
acc	-0.062***
	(-5.99)
size	-0.005***
	(-4.29)
mtb	-0.007***
	(-7.03)
retvol	-0.393***
	(-20.84)
q2	0.001
	(0.74)
q3	-0.005***
	(-3.89)
q4	0.001
	(0.24)
Observations	145479
R^2	0.22

Notes: Year and 2-digit SIC industry fixed effects are included in the regressions, but are not reported. $Q2$ ($Q3$ or $Q4$) is a dummy variable which is set to 1 if the current reporting quarter is the second (third or fourth) fiscal quarter. The dependent and other independent variables are as defined in Table 4A. T-statistics clustered at firm level are reported in parentheses. *** means $p < 0.01$; ** $p < 0.05$; * $p < 0.1$.

Table 6A: Future earnings and MD&A tones

COEFFICIENT	(1) EARN(t+1)	(2) EARN(t+2)	(3) EARN(t+3)	(4) EARN(t+4)
tone	0.010***	0.009***	0.009***	0.005***
	(7.72)	(6.17)	(5.41)	(3.24)
earn	0.388***	0.348***	0.356***	0.362***
	(21.69)	(22.13)	(17.04)	(19.41)
ret	0.011***	0.011***	0.006***	0.005***
	(7.25)	(6.07)	(4.20)	(2.78)
acc	-0.185***	-0.179***	-0.165***	-0.188***
	(-21.19)	(-22.41)	(-14.83)	(-18.60)
size	0.003***	0.003***	0.003***	0.003***
	(12.26)	(10.33)	(8.82)	(9.59)
mtb	-0.003***	-0.004***	-0.005***	-0.006***
	(-4.52)	(-4.51)	(-5.71)	(-8.57)
retvol	-0.062***	-0.067***	-0.060***	-0.053***
	(-10.73)	(-9.94)	(-7.76)	(-7.83)
q2	-0.000	-0.009***	0.010***	0.001
	(-0.67)	(-12.02)	(10.38)	(1.25)
q3	-0.009***	0.000	0.010***	-0.001
	(-10.64)	(0.28)	(10.94)	(-1.40)
q4	0.001	0.001	0.009***	-0.009***
	(0.79)	(0.98)	(7.95)	(-9.21)
Observations	122919	118100	112116	109597
R^2	0.35	0.30	0.27	0.27

Notes: The dependent variables are the earnings in the next four quarters scaled by the book value of assets at the end of this quarter. Year and 2-digit SIC industry fixed effects are included in the regressions, but are not reported. $Q2$ ($Q3$ or $Q4$) is a dummy variable which is set to 1 if the current reporting quarter is the second (third or fourth) fiscal quarter. Other independent variables are as defined in Table 4A. T-statistics clustered at firm level are reported in parentheses. *** means $p < 0.01$; ** $p < 0.05$; * $p < 0.1$.

Table 6B: Future earnings changes and MD&A tones

COEFFICIENT	(1) DEARN(t+1)	(2) DEARN(t+2)	(3) DEARN(t+3)	(4) DEARN(t+4)
tone	0.010***	0.009***	0.009***	0.005***
	(7.27)	(5.80)	(5.44)	(2.92)
earn	-0.617***	-0.651***	-0.640***	-0.633***
	(-32.58)	(-40.31)	(-33.37)	(-31.74)
ret	0.011***	0.011***	0.006***	0.005***
	(6.85)	(7.17)	(4.39)	(2.73)
acc	-0.187***	-0.186***	-0.176***	-0.200***
	(-17.72)	(-24.25)	(-14.51)	(-16.99)
size	0.004***	0.003***	0.003***	0.003***
	(11.01)	(11.09)	(9.41)	(9.36)
mtb	-0.003***	-0.004***	-0.005***	-0.006***
	(-3.78)	(-5.02)	(-6.99)	(-8.89)
retvol	-0.061***	-0.066***	-0.058***	-0.051***
	(-8.96)	(-9.37)	(-8.32)	(-6.94)
q2	-0.000	-0.009***	0.010***	0.001
	(-0.38)	(-11.81)	(10.40)	(0.86)
q3	-0.009***	0.000	0.010***	-0.002*
	(-10.38)	(0.10)	(10.70)	(-1.70)
q4	0.001	0.001	0.009***	-0.009***
	(0.69)	(0.83)	(7.67)	(-9.54)
Observations	122919	118100	112116	109597
R^2	0.46	0.44	0.36	0.37

Notes: The dependent variables are the earnings in the next four quarters minus the earnings in this quarter scaled by the book value of assets at the end of this quarter. Year and 2-digit SIC industry fixed effects are included in the regressions, but are not reported. $Q2$ ($Q3$ or $Q4$) is a dummy variable which is set to 1 if the current reporting quarter is the second (third or fourth) fiscal quarter. Other independent variables are as defined in Table 4A. T-statistics clustered at firm level are reported in parentheses. *** means $p < 0.01$; ** $p < 0.05$; * $p < 0.1$.

Table 7: Future liquidity and MD&A tones

COEFFICIENT	(1) CFRATIO(t+1)	(2) CFRATIO(t+2)	(3) CFRATIO(t+3)	(4) CFRATIO(t+4)
tone	0.280***	0.278***	0.241***	0.194***
	(9.85)	(9.97)	(9.01)	(7.71)
earn	2.643***	2.469***	2.423***	2.654***
	(16.00)	(15.97)	(14.03)	(15.97)
ret	0.081***	0.071***	0.045***	0.035**
	(5.76)	(5.27)	(3.32)	(1.97)
acc	-2.094***	-1.453***	-1.635***	-2.433***
	(-14.43)	(-12.12)	(-13.41)	(-18.60)
size	0.044***	0.050***	0.046***	0.032***
	(9.00)	(9.65)	(8.95)	(6.60)
mtb	-0.015*	-0.019**	-0.014*	-0.004
	(-1.90)	(-2.26)	(-1.71)	(-0.47)
retvol	-0.875***	-0.972***	-0.964***	-0.882***
	(-10.78)	(-11.89)	(-11.53)	(-10.46)
q2	0.019***	0.062***	-0.074***	0.005
	(3.24)	(8.85)	(-4.36)	(0.94)
q3	0.102***	-0.009	-0.070***	0.028***
	(12.60)	(-0.63)	(-5.18)	(2.93)
q4	0.015	0.000	-0.061***	0.088***
	(1.30)	(0.03)	(-5.51)	(7.37)
Observations	105626	101410	96272	94029
R^2	0.33	0.28	0.28	0.33

Notes: The dependent variables are the cash flows from operations in the next four quarters scaled by the book value of current liabilities at the end of that quarter. Year and 2-digit SIC industry fixed effects are included in the regressions, but are not reported. $Q2$ ($Q3$ or $Q4$) is a dummy variable which is set to 1 if the current reporting quarter is the second (third or fourth) fiscal quarter. Other independent variables are as defined in Table 4A. T-statistics clustered at firm level are reported in parentheses. *** means $p < 0.01$; ** $p < 0.05$; * $p < 0.1$.

Table 8: Tests of non-linearity effects of MD&A tones

	(1)	(2)
COEFFICIENT	EARN(t+1)	CFRATIO(t+1)
tone	0.007***	0.298***
	(3.73)	(6.66)
ptone	0.003***	0.013
	(3.51)	(0.83)
tone_ptone	-0.001	-0.155**
	(-0.45)	(-2.42)
earn	0.388***	2.642***
	(21.69)	(16.00)
ret	0.011***	0.081***
	(7.23)	(5.77)
acc	-0.185***	-2.093***
	(-21.19)	(-14.42)
size	0.003***	0.044***
	(12.27)	(8.96)
mtb	-0.003***	-0.015*
	(-4.53)	(-1.89)
retvol	-0.062***	-0.872***
	(-10.72)	(-10.76)
q2	-0.000	0.019***
	(-0.65)	(3.19)
q3	-0.009***	0.102***
	(-10.60)	(12.52)
q4	0.001	0.014
	(0.93)	(1.20)
Observations	122919	105626
R^2	0.36	0.33

Notes: The dependent variables are the earnings in the next quarter scaled by the book value of assets at the end of this quarter in column (1) and the cash flows from operations in the next quarter scaled by the book value of current liabilities at the end of next quarter in column (2). Year and 2-digit SIC industry fixed effects are included in the regressions, but are not reported. $PTONE$ is a dummy variable that equals 1 if $TONE \geq 1$ and 0 otherwise. $Q2$ ($Q3$ or $Q4$) is a dummy variable which is set to 1 if the current reporting quarter is the second (third or fourth) fiscal quarter. Other independent variables are as defined in Table 4A. T-statistics clustered at firm level are reported in parentheses. *** means $p < 0.01$; ** $p < 0.05$; * $p < 0.1$.

Table 9A: Information content of MD&A as a function of year

COEFFICIENT	(1) EARN(t+1)	(2) EARN(t+2)	(3) EARN(t+3)	(4) EARN(t+4)	(5) CFRATIO(t+1)	(6) CFRATIO(t+2)	(7) CFRATIO(t+3)	(8) CFRATIO(t+4)
tone	0.031	0.432	0.815	0.429	3.374	-3.969	3.222	16.801
	(0.05)	(0.67)	(1.12)	(0.52)	(0.27)	(-0.32)	(0.27)	(1.48)
year	-0.000	0.000	0.000*	0.000***	0.005***	0.007***	0.007***	0.006***
	(-0.52)	(1.21)	(1.94)	(3.91)	(2.77)	(3.66)	(3.30)	(2.84)
tone*year	-0.000	-0.000	-0.000	-0.000	-0.002	0.002	-0.002	-0.008
	(-0.04)	(-0.65)	(-1.11)	(-0.51)	(-0.25)	(0.34)	(-0.26)	(-1.47)

Table 9B: Information content of MD&A before and after 2003

COEFFICIENT	(1) EARN(t+1)	(2) EARN(t+2)	(3) EARN(t+3)	(4) EARN(t+4)	(5) CFRATIO(t+1)	(6) CFRATIO(t+2)	(7) CFRATIO(t+3)	(8) CFRATIO(t+4)
tone	0.010***	0.009***	0.008***	0.005***	0.229***	0.233***	0.208***	0.181***
	(7.73)	(5.96)	(5.08)	(2.72)	(9.02)	(9.09)	(8.50)	(7.51)
post2003	-0.000	0.001**	0.003***	0.004***	0.025*	0.023*	0.018	0.008
	(-0.08)	(2.10)	(3.47)	(5.58)	(1.89)	(1.68)	(1.31)	(0.64)
tone*post2003	-0.004*	-0.003	-0.002	0.002	0.002	-0.011	-0.032	-0.075**
	(-1.88)	(-1.42)	(-0.69)	(0.54)	(0.05)	(-0.27)	(-0.82)	(-2.00)

Notes: The dependent variables are the earnings in the next four quarters scaled by the book value of assets at the end of this quarter in columns (1) to (4) and the cash flows from operations in the next four quarters scaled by the book value of current liabilities at the end of that quarter in columns (5) to (8). Year and 2-digit SIC industry fixed effects are included in the regressions, but are not reported. Control variables are the same as those in Table 6A and Table 7, but are not reported. T-statistics clustered at firm level are reported in parentheses. *** means $p < 0.01$; ** $p < 0.05$; * $p < 0.1$.

Table 10: Analyst forecast and MD&A tones

	(1)	(2)	(3)
COEFFICIENT	ANALYST_FCST	EARN(t+1)	CFRATIO(t+1)
tone	0.006***	0.005***	0.162***
	(7.21)	(4.40)	(5.35)
earn	0.365***	0.164***	2.488***
	(19.18)	(6.15)	(10.95)
ret	0.002***	0.009***	0.039***
	(3.41)	(8.72)	(3.72)
acc	-0.182***	-0.067***	-2.171***
	(-21.36)	(-4.84)	(-13.31)
size	0.002***	0.001**	0.016***
	(10.42)	(2.47)	(3.16)
mtb	-0.001	0.000	0.010*
	(-1.45)	(1.06)	(1.75)
retvol	-0.031***	-0.048***	-0.728***
	(-6.60)	(-8.75)	(-6.84)
q2	-0.002***	-0.001*	0.049***
	(-5.93)	(-1.66)	(7.77)
q3	-0.004***	-0.005***	0.141***
	(-7.65)	(-6.61)	(15.28)
q4	-0.009***	-0.002**	-0.124***
	(-14.90)	(-2.21)	(-9.48)
analyst_fcst		0.663***	4.450***
		(10.87)	(9.68)
Observations	83613	71799	63457
R^2	0.54	0.47	0.43

Notes: The dependent variables are the latest financial analysts' consensus forecast of next quarter's earnings before its announcement scaled by the book value of assets (i.e., $ANALYST\_FCST$) in column (1), the earnings in the next quarter scaled by the book value of assets at the end of this quarter in column (2), and the cash flows from operations in the next quarter scaled by the book value of current liabilities at the end of next quarter in columns (4) and (5). Year and 2-digit SIC industry fixed effects are included in the regressions, but are not reported. $Q2$ ($Q3$ or $Q4$) is a dummy variable which is set to 1 if the current reporting quarter is the second (third or fourth) fiscal quarter. Other independent variables are as defined in Table 4A. T-statistics clustered at firm level are reported in parentheses. *** means $p < 0.01$; ** $p < 0.05$; * $p < 0.1$.

Table 11: Decomposing MD&A tones by contents

	(1)	(2)
COEFFICIENT	EARN(t+1)	CFRATIO(t+1)
profit_tone	0.003***	0.096***
	(4.20)	(5.64)
liquidity_tone	0.013***	0.250***
	(10.44)	(10.90)
other_tone	-0.001	0.006
	(-0.96)	(0.35)
Observations	122919	105626
R^2	0.36	0.33

Notes: The dependent variables are the earnings in the next quarter scaled by the book value of assets at the end of this quarter in column (1) and the cash flows from operations in the next quarter scaled by the book value of current liabilities at the end of next quarter in column (2). Control variables (results unreported) include $EARN$, RET, ACC, $SIZE$, MTB, $RETVOL$, $Q2$, $Q3$, and $Q4$. $Q2$ ($Q3$ or $Q4$) is a dummy variable which is set to 1 if the current reporting quarter is the second (third or fourth) fiscal quarter. Other independent variables are as defined in Table 4A. Year and 2-digit SIC industry fixed effects are included in the regressions, but are not reported. T-statistics clustered at firm level are reported in parentheses. *** means $p < 0.01$; ** $p < 0.05$; * $p < 0.1$.

Table 12: Information content of 10-K versus 10-Q

	EARN(t+1)			CFRATIO(t+1)		
	(1)	(2)	(3)	(4)	(5)	(6)
COEFFICIENT	10-Q	10-K	Diff.	10-Q	10-K	Diff.
tone	0.008***	0.006***	insig	0.248***	0.092***	sig
	(5.51)	(3.19)		(7.83)	(4.76)	
earn	0.513***	0.242***	sig	4.437***	0.801***	sig
	(20.11)	(15.52)		(12.15)	(12.65)	
ret	0.015***	0.002	sig	0.067***	0.006	sig
	(10.63)	(0.39)		(4.29)	(0.38)	
acc	-0.225***	-0.165***	sig	-4.273***	-0.640***	sig
	(-13.41)	(-16.73)		(-13.55)	(-12.35)	
size	0.003***	0.002***	sig	0.039***	0.013***	sig
	(10.39)	(4.32)		(6.67)	(3.78)	
mtb	-0.002***	-0.002**	insig	-0.003	-0.003	insig
	(-3.32)	(-2.23)		(-0.40)	(-0.42)	
retvol	-0.064***	-0.035***	sig	-0.971***	-0.324***	sig
	(-10.67)	(-2.91)		(-8.70)	(-6.28)	
Observations	92130	30787		79148	26485	
R^2	0.39	0.36		0.46	0.24	

Notes: The dependent variables are the earnings in the next four quarters scaled by the book value of assets at the end of this quarter (columns (1) and (2)) and the cash flows from operations in the next four quarters scaled by the book value of current liabilities at the end of that quarter (columns (3) and (4)). Year and 2-digit SIC industry fixed effects are included in the regressions, but are not reported. The independent variables are as defined in Table 4A. Columns (1) and (4) use the data from all 10-Q filings (i.e., filings in fiscal quarters 1 to 3), and Columns (2) and (5) use the data from all 10-K filings (i.e., filings in fiscal quarters 4). T-statistics clustered at firm level are reported in parentheses. Column (3) shows the statistical significance between the coefficients in column (1) and column (2). Column (6) shows the statistical significance between the coefficients in column (4) and column (5). "sig" ("insig") means that the difference is (not) significant at 0.10 level. *** means $p < 0.01$; ** $p < 0.05$; * $p < 0.1$.