The Information Content of Forward-Looking Statements in Corporate Filings—A Naïve Bayesian Machine Learning Approach

Author(s): FENG LI

Source: *Journal of Accounting Research*, DECEMBER 2010, Vol. 48, No. 5 (DECEMBER 2010), pp. 1049-1102

Published by: Wiley on behalf of Accounting Research Center, Booth School of Business, University of Chicago

Stable URL: https://www.jstor.org/stable/40929537

DOI: 10.1111/j.1475-679X.2010.00382.x
Journal of Accounting Research
Vol. 48 No. 5 December 2010
Printed in U.S.A.

The Information Content of Forward-Looking Statements in Corporate Filings—A Naïve Bayesian Machine Learning Approach

FENG LI*

Received 14 January 2009; accepted 8 June 2010

ABSTRACT

This paper examines the information content of the forward-looking statements (FLS) in the Management Discussion and Analysis section (MD&A) of 10-K and 10-Q filings using a Naïve Bayesian machine learning algorithm.

*Ross School of Business, University of Michigan. I would like to thank Phil Berger, Dave Burgstahler, John Core, Ilia Dichev, Weili Ge, Leslie Hodder, Jack Hughes, Raffi Indjejikian, Mark Lang, Dave Larcker, Charles Lee, Roby Lehavy, Russell Lundholm, Dawn Matsumoto, Brian P. Miller, Venky Nagar, Marlene Plumlee, Grace Powell, Jonathan Rogers, Haresh Sapra, Cathy Shakespeare, Terry Shevlin, Doug Skinner, Abbie Smith, K. R. Subramanyam, Brett Trueman, Jenny Tucker, Regina Wittenberg Moerman, Sarah Zechman, the participants of the 2008 University of Michigan Traverse City Conference, the 2008 North American Summer Meeting of the Econometric Society, the 2009 FARS mid-year conference, the New York University 2009 Summer Camp, and the Deutsche Bank Asset Management seminar, and the workshop participants at Chicago, Notre Dame, University of Washington, UCLA, Indiana, Florida, Emory, and Stanford for their helpful comments and Julie Miller for her editorial help. Special thanks go to Merle Erickson (the editor) and an anonymous reviewer. I also thank Mike Chen, Qian Fang, Qian Li, Chris Liong, Doris Lu, Amy McMahon, Bharat Modi, Qimin Ni, Darin Styles, Kathy Wang, Tina Wei, Sarah Xia, Quan Yuan, and especially Rencheng Wang and Winnie Wei for their excellent research assistance and Ernst, and Young Faculty Development Fund and the Harry Jones Endowment for Earnings Quality at the University of Michigan for financial support.

I find that firms with better current performance, lower accruals, smaller size, lower market-to-book ratio, less return volatility, lower MD&A Fog index, and longer history tend to have more positive FLSs. The average tone of the FLS is positively associated with future earnings even after controlling for other determinants of future performance. The results also show that, despite increased regulations aimed at strengthening MD&A disclosures, there is no systematic change in the information content of MD&As over time. In addition, the tone in MD&As seems to mitigate the mispricing of accruals. When managers "warn" about the future performance implications of accruals (i.e., the MD&A tone is positive (negative) when accruals are negative (positive)), accruals are not associated with future returns. The tone measures based on three commonly used dictionaries (Diction, General Inquirer, and the Linguistic Inquiry and Word Count) do not positively predict future performance. This result suggests that these dictionaries might not work well for analyzing corporate filings.

1. Introduction

This paper analyzes the information content of forward-looking statements (FLS) in the Management's Discussion and Analysis section of 10-K and 10-Q filings. In 1980, the Securities and Exchange Commission (SEC) mandated that public companies include in their annual reports a section for Management's Discussion and Analysis of Financial Condition and Results of Operations (MD&A). The MD&A is intended to assess an enterprise's liquidity, capital resources, and operations in a way that many investors can understand. One of the SEC's goals in mandating the MD&A was to make public the information about predictable future events and trends that may affect future operations of the business.

Despite this goal, whether MD&A disclosures are truly informative remains an open empirical question. Consistent with the SEC's intention, the MD&A is arguably the most read and most important component of the financial section (Tavcar [1998]). Indeed, of all the disclosure items of the annual report, sell-side financial analysts in the United States most frequently rely upon the MD&A when preparing their reports (Knutson [1993] and Rogers and Grant [1997]). Furthermore, the safe harbor provisions of the Private Securities Litigation Reform Act of 1995 encourage more forward-looking information (Grundfest and Perino [1997]) and should make MD&A disclosures more informative.

However, the MD&A might not be as informative as intended for several reasons. Companies have concerns over proprietary costs (Verrecchia [1983]) and uncertainties about the judicial interpretation of safe harbor protection (Grundfest and Perino [1997]). Also, the MD&As are not required to be audited (Hüfner [2007]) and many include substantial boilerplate disclaimers and disclosures, generic language, and immaterial detail (SEC [2003]). Pava and Epstein [1993] show that although most of the companies they study accurately describe historical events, very few provide useful and accurate forecasts in their MD&As.

In this paper, I study the determinants and information content of the forward-looking statements in the MD&As. Specifically, I explore variations in the tone of the FLS and study their economic determinants. I also examine whether the FLS contain incremental information regarding future profitability and liquidity. I then explore whether this information content changes over time, particularly after 2003, when the SEC issued new guidelines for preparing MD&As and the Sarbanes-Oxley Act increased disclosure requirements in MD&As (Bainbridge [2007]). Finally, I examine the implications of the MD&A tone for the accrual anomaly.

To assess the content and tone of FLS in MD&As, I rely on a Naïve Bayesian learning algorithm instead of a dictionary-based approach (e.g., Kothari, Li, and Short [2009]).[1] First, I manually categorize 30,000 sentences of randomly selected FLS extracted from the MD&A section of 10-Q filings along two dimensions: tone (i.e., positive versus negative) and content (e.g., profitability operations, liquidity, etc.). These sentences are then used as training data in the Naïve learning algorithm to categorize the tone and content of other FLS in 10-Q and 10-K filings. *N*-fold cross-validation tests (with *N* varying from 3 to 50) indicate that the algorithm predicts the tone (positive, neutral, or negative) and content (out of 12 possible categories) of FLS with a success rate of about 67% and 63%, respectively.

After developing the training data, I use the Bayesian learning algorithm to categorize the tone and content of about 13 million FLS from more than 140,000 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 (MTB) ratio, less return volatility, lower MD&A Fog index, and longer history tend to have more positive FLS in their MD&As.

I find that the average tone of the FLS in a firm's MD&A is positively correlated with its future earnings and liquidity and has explanatory power incremental to other variables in predicting future performance. For instance, the return on assets in the subsequent quarter is 5 percentage points higher on an annual basis for firms with a positive MD&A tone than that for firms with a negative tone. An interquartile change in the MD&A tone implies a difference in annual return on assets of 1 percentage point. These effects are found after controlling for current earnings, stock returns, accruals, and other factors that may affect future performance. An examination of the information content of MD&As over time shows that, despite the SEC's continuous efforts and the passage of the Sarbanes-Oxley Act, there is no systematic change in the informativeness of MD&As over time.

[1] The dictionary approach relies on a "mapping" algorithm and assigns each word (or phrase) in a document into different categories based on some predefined dictionaries.

The tone of MD&As is also related to the cross-sectional association of accruals with future stock returns. While accruals are negatively associated with future returns (Sloan [1996]), lower (higher) accruals are not associated with higher (lower) future returns if managers "warn" about the future outlook in the MD&As (i.e., the accruals are negative (positive) and the MD&A tone is positive (negative)). This result suggests that MD&A disclosures mitigate the mispricing of accruals.

Finally, empirical results based on the dictionary-approach tone measures do not lend support to the hypothesis that MD&As contain information content about future performance. Specifically, when the MD&A tone is measured using dictionaries such as Diction, General Inquirer (GI), or Linguistic Inquiry and Word Count, it does not relate positively to future performance. The correlations between the MD&A tone measured using the dictionaries and current earnings and MD&A Fog index are also much smaller compared with those between the Bayesian tone measure and current earnings and MD&A Fog. These results suggest that the dictionaries do not work well for the corporate financial statement domain.

This paper contributes to the literature in several ways. This study is the first to use a statistical learning methodology to analyze financial disclosures. 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, which is widely used in other research areas (e.g., Mitchell [2006]), can be successfully applied to financial statement settings and thus could be useful for future research on disclosure.

This paper is also among the first of several large-sample studies on FLS in 10-Q and 10-K filings.[2] This study extends the literature on management disclosures of forward-looking information (Patell [1976], Penman [1980], Pownell, Wasley, and Waymire [1993], Skinner [1994], Dietrich et al. [1997], Miller and Piotroski [2000], Hutton, Miller, and Skinner [2003], and Lang and Lundholm [2003]).

Several prior small-sample studies based on content analysis by human coders also find that MD&As provide information content about future firm performance (Bryan [1997], Barron, Kile, and O'Keefe [1999], and Callahan and Smith [2004]). The innovations of this paper over these studies are numerous. First, none of the prior studies examine the determinants of MD&A tones. In this paper, I test hypotheses about economic factors that may explain MD&A tone variations. Second, this study provides evidence on the time-series change in the MD&A information content and thus sheds light on the effectiveness of the SEC regulations to strengthen the requirements on MD&A disclosures. Third, accruals and MD&As are two important

[2] A contemporaneous study (Muslu et al. [2008]) also examines the information content of FLS in MD&As, but the focus of that paper is on the intensity of the forward-looking information, rather than the tone. Feldman et al. [2009] examine the MD&A tone and post–earnings announcement drift using the GI.

communication devices for managers. This paper contributes to the literature by being the first to examine whether information from the MD&As impacts investors' response to accruals.

The rest of the paper proceeds as follows. Section 2 discusses the nature of MD&A 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. Section 6 concludes.

2. *Literature Review and Research Questions*

2.1 MD&A AND FORWARD-LOOKING STATEMENTS

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

The SEC also encourages forward-looking information in circumstances such as known material trends, events, commitments, and uncertainties. However, these requirements are not set in objective terms, as they allow for management discretion. Whether the disclosure of forward-looking information is required depends cumulatively on management's assessment of (1) whether a circumstance is "presently known," (2) 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. Therefore, management's discretion is given considerable leeway.

In order to encourage companies to provide FLS in their MD&As, any predictive information is explicitly covered by the safe-harbor rule for projections (Regulation S-K Item 303 (a) Instr. 7). However, companies may be reluctant because of their uncertainty regarding judicial interpretation of the safe harbor provisions and because of fears regarding state court litigation where no such safe harbor is available (Grundfest and Perino [1997]). Also, it is not mandatory for MD&A content to be audited. At best, auditors have a professional responsibility to check MD&A information for material inconsistencies against the respective financial statements.

2.2 LITERATURE REVIEW

There is extensive research on the economic implications of voluntary corporate disclosure in the accounting literature.[3] Conceptually, there are at least three disclosure characteristics that are interesting to researchers: the level (how much you say), the tone (what do you mean), and the transparency (how you say it). The majority of prior studies examine the level of disclosure as illustrated in appendix A1. Most of these studies rely on manual coding of the disclosure level, perhaps due to the difficulty in creating a large-sample measure of disclosure quality. A few papers examine disclosure transparency by examining the readability (Li [2008], Miller [2010]) and vocal tone of disclosures (Mayew and Venkatachalam [2008]) or the magnitude of the change in disclosures (Brown and Tucker [2010]).

This paper studies the tone in corporate disclosures. Most papers in this area carry out a 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. While manual coding may be more precise, it has two disadvantages (Core [2001]): small sample size due to cost considerations and difficulty with replication due to subjectivity in the coding process. In particular, the small sample size may limit the scope of the empirical tests. For instance, to study any potential change in the information content of MD&A disclosures over time, researchers need a relatively large panel of data. One solution is to rely on a dictionary-based content analysis. The difference between the dictionary approach and the approach used in this paper is discussed in greater detail in sections 3 and 5.

The empirical results regarding the information content of MD&As are mixed. Bryan [1997] finds that discussions of future operations and capital expenditures are associated with future short-term performance. Likewise, Callahan and Smith [2004] find that their disclosure index based on a content analysis provides incremental explanatory power in predicting future firm performance and market valuation. A contemporaneous study by Muslu et al. [2008] examines the intensity of the forward-looking information in MD&As and finds that a more intense discussion of forward-looking information makes stock returns incorporate future earnings in a more timely fashion and reduce analysts' forecast errors. However, Pava and Epstein [1993] show that, while most companies accurately describe historical events, very few provide useful and accurate forecasts. They also find a strong bias in favor of correctly projecting positive trends, while negative trends tend to be either ignored or not fully reported.

[3] 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 forecasts (e.g., Waymire [1984]), and the consequences of mandatory disclosures (e.g., Leuz and Verrecchia [2000]), which are discussed in detail in Healy and Palepu [2001] and Core [2001].

In addition to examining the tone of corporate disclosures, this paper also furthers our understanding of the content analysis of textual documents using computer algorithms. Within this literature, there is extensive research on the information content of corporate earnings releases (Davis, Piger, and Sedor [2005], Rogers, Buskirk, and Zechman [2009], Henry and Leone [2010]), accounting policy disclosures (Levine and Smith [2006]), audit opinions (Butler, Leone, and Willenborg [2004]), financial news (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, Li, and Short [2009]), presidential election campaigns (Pennebaker and Stone [2001]), and art history (Martindale [1990]). Appendix A2 lists a sample of textual analysis papers in accounting and finance and tabulates them according to their methodology. The majority of the papers use the dictionary-based approach, and hence have a firm-level measure based on the percentage of specific categories of words.

2.3 HYPOTHESES

2.3.1. Determinants of MD&A Tone. I explore several factors to explain the cross-sectional variations in the tone of MD&A forward-looking statements (FLS):

Current firm performance. While there is substantial theoretical and empirical literature on the amount of disclosure, little work exists on the tone. Many arguments made in the disclosure level literature can apply to the tone as well. For example, litigation concern may encourage firms with good current performance to be more cautious in discussing future events in their MD&A. Momentum in firm performance also suggests that the MD&A FLS may be more positive for firms with good current performance. However, earnings are mean-reverting, which implies a more negative tone in FLS for firms with better current performance. Therefore, ex ante the relation between tone and current performance is unclear.

Accruals. It is well documented that accruals are negatively associated with future firm performance and investors underreact 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 understand the implications of the accruals for future performance. Alternatively, accruals may simply proxy for a firm's economic conditions (e.g., distress) and managers are likely to understand (at least partially) the implications of accruals for future earnings. In either case, a negative relation between accruals and the MD&A tone is expected. However, if managers are overconfident or fixated on current earnings, they may not understand the implications of accruals for future performance and hence no relation between accruals and MD&A tone is expected. Also, if managers understand the implications of accruals but have incentives to mislead investors, a positive relation between accruals and MD&A tone might result.

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]). Larger firms may have more cautious FLS because of the higher political and legal cost due to their visibility.

Market-to-book ratio. High MTB firms are different from low MTB 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 the MTB ratio and MD&A tone 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. These firms are also more likely to have severe information asymmetry between managers and investors. Finally, performance variability may be related to MD&A tone because of its effect on a firm's vulnerability to legal action. These factors all posit a negative relation between volatility and MD&A tone.

Complexity of operations. More complex operations are more likely to lead to complex disclosures. I use the numbers of segments and the number of nonmissing financial items in Compustat to capture the complexity of operations and examine how they are associated with MD&A tone.

Firm age. Young firms face more uncertainties. Managers of young companies are likely to be more cautious when discussing future outlook. A negative relation between MTB ratio and MD&A tone is expected.

Firm event. I include two firm-event dummies as potential determinants of the MD&A tone—seasoned equity offering and merger and acquisition activities. Firms that make seasoned equity offering or acquisitions might have incentives to discuss future outlook more positively. Also, firms with more positive outlook are more likely to issue new equity or make acquisitions. Therefore, a positive association between a firm's equity offering and merger transactions and its MD&A tone is expected.

Special items. Firms with a significant amount of special items are more likely to experience unusual performance. I expect firms with more negative special items to have more negative MD&A tone.

Incorporation state. Daines [2001] shows that Delaware firms, due to their different corporate laws and investor protections, are more likely to receive takeover bids and be acquired. Such firms are also valued higher than similar firms incorporated elsewhere. I include a dummy variable for Delaware-incorporated firms to examine whether such firms tend to have different MD&A tone.

Disclosure transparency. Li [2008] shows that firms' annual report readability captures management's obfuscation behavior. When performance is poor, annual reports tend to be less readable and have a high Fog index. I include the Fog index of the MD&A section as a determinant of the MD&A tone and expect Fog to be negatively correlated with the MD&A tone.

Reporting quarter. Prior research has shown that the accounting information behaves differently across different reporting quarters (Das and Shroff [2002]). To examine the potential implications of reporting quarter on the MD&A tone, I include three quarterly dummies (*Q*2, *Q*3, and *Q*4) for the 2nd, 3rd, and 4th quarter, respectively.

2.3.2. Information Content of MD&As. I examine whether the FLS in corporate MD&A disclosures contain information about future profitability and liquidity. The empirical tests are a joint test of (1) the machine learning algorithm's ability to capture tone, (2) whether managers follow the SEC instructions and provide information about the future in their MD&A disclosures, and (3) whether managers are truthful in their disclosures. If the Bayesian algorithm is able to capture management tone, then the evidence that the MD&A tone predicts future performance is consistent with the hypotheses that managers are truthfully disclosing information in MD&A and that corporate filings have information content. However, if the MD&A tone based on the Bayesian algorithm is not associated with future performance, then I cannot reject the hypothesis that MD&A has no information content because the result could be due to the low power of the algorithm.

I also examine whether there is a systematic change in the information content of MD&As over time with a particular focus on a pre- and post-2003 comparison. The SEC issued new guidelines on MD&A disclosures in 2003 and encouraged a significant increase in information content and a reduction in boilerplate language (SEC [2003]). The Sarbanes-Oxley Act also enhanced the MD&A disclosure requirements (Bainbridge [2007]). First, the MD&A section is chosen 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 present the financial conditions and results of operations of the issuer. Consequently, a test of the change in the information content of MD&As will shed light on the effectiveness of the SEC regulations.

2.3.3. MD&A Tone and the Accrual Anomaly. Prior literature has shown that accruals contain information content about future performance and can predict future stock returns (Sloan [1996]). Following Sloan [1996], a large literature has emerged examining the possible explanations for the accrual anomaly (Richardson, Tuna, and Wysocki [2009] provide a comprehensive review of this literature). To shed light on this issue, many prior papers empirically examine whether the accrual anomaly is a function of transaction costs and firm characteristics (such as growth). However, none of these papers link the accrual anomaly to the other disclosures made by managers.[4]

Both accruals and MD&A tone can be considered signals from managers about future firm performance. While accruals are subject to the financial reporting rules, managers tend to have a higher degree of freedom with the MD&A disclosures (Bainbridge [2007]). To the extent that the MD&A tone provides a more direct prediction about future outlook than do accruals, the information in the MD&A tone about future performance is more salient. Given the limited attention of investors, salient information could affect investors' interpretation of other information (Hirshleifer and Teoh [2003]). I therefore examine the accrual anomaly as a function of the MD&A tone. This sheds light on how investors price accruals conditional on the information communicated by managers in their MD&As.

Sloan [1996] suggests that investors fixate on reported earnings and, as a result, accruals are negatively related to future stock returns. If more positive (negative) management discussions about the future outlook accompany negative (positive) accruals, investors may be less likely to fixate on accruals to the extent that the discussions made by managers provide more salient information about future performance and help investors understand the implications of accruals (Hirshleifer and Teoh [2003]). In this scenario, information in the MD&A helps mitigate the accrual anomaly and one should expect the accrual anomaly to be stronger for firms with managers who do not "warn" investors about the future performance implications of the accruals. Alternatively, investors may ignore the MD&A tone because there is a substantial amount of boilerplate disclosure in MD&As and the information processing cost could be high. In this scenario, the degree of mispricing of the accruals would be unrelated to the MD&A tone. I conduct empirical tests to distinguish between the two scenarios to shed light on how investors respond to the information in accruals depending on other information signaled by managers.

3. Naïve Bayesian Algorithm and Text Classification

There are two general approaches for conducting content analysis: a rule-based approach (i.e., dictionary approach) and a statistical approach. The first approach uses a "mapping" algorithm in which a computer program reads the text and classifies words (or phrases) into different categories based on predefined rules (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.

[4] One exception is a recent study by Huddart and Louis [2009], who examine the relation between insider trading and the pricing of "managed" accruals. In another study, Barth and Hutton [2004] find that the accrual anomaly is a function of analyst forecast revisions.

The second approach, which was pioneered by computer scientists and mathematicians, relies on statistical techniques to infer the content of text and classify documents based on statistical inference (e.g., Manning and Schütze [1999] and Mitchell [2006]). 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 use the statistical approach that 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, consider 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 (http://www.webuse.umd.edu:9090/), the sentence has two 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 dictionary-based approach does not take into consideration the context of a sentence. For instance, if a sentence is about revenue, then "increase" should be treated as a positive word; however, it is likely to be of a negative tone if the topic is "cost." Third, the rule-based approach generally ignores prior knowledge that researchers may have about the text. For example, if most of the sentences that appear in MD&A reports are neutral, then, unless there is strong evidence that a sentence is of negative tone, it is more efficient to classify a random sentence as being of a neutral tone. This point is especially salient for managerial disclosure, because managers have incentives to disclose strategically. Finally, the statistical approach typically provides a natural way to validate classification efficiency using the training data. The training data are human coded and thus 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. Under this method, a given sentence is first reduced to a list of words (*words*) with each word weighted in some fashion (e.g., by frequency in the sentence). The goal is to classify the sentence into a specific category (*cat*) from a set of all possible categories (*cats*). In this paper, there are four possible tone categories: positive, negative, neutral, and uncertain. The Naïve Bayesian algorithm chooses the best category by solving the following problem:

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

Since $P(words)$ does not change over the range of categories, it can be eliminated. The problem thus becomes:

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

Finally, if $w_1, w_2, \ldots, w_n$ are the words in the document and their probability of appearing in a sentence is assumed to be independent, then this expression is equivalent to:

$$cat^* = \underset{cat \in cats}{argmax}\, P(w_1 \mid cat) * P(w_2 \mid cat) * \cdots * P(w_n \mid cat) * P(cat),$$

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

The last step is the only nonrigorous one in the derivation and thus 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. The independence assumption simplifies the computation and avoids the "curse of dimensionality" problem (Bellman [1961]). Independence is assumed even though it is not true. For example, in the financial statement setting, the words "adverse effect" are more likely to appear together with the word "material." However, empirical results from other fields suggest that making this assumption even if it is not true may have little effect on the results.[5]

4. *Empirical Implementation*

4.1 DATA PREPARATION

To apply the Naïve Bayesian algorithm, I first download all 10-Ks and 10-Qs filed between 1994 and 2007 from the SEC Edgar Web site and remove any HTML tag. I then extract the MD&A section of these filings.[6] Next, I split the MD&A text into sentences using the Lingua::EN::Sentence module in Perl, which takes acronyms into consideration in the splitting process.[7] The FLS from the 10-Qs and 10-Ks are then extracted and the details of this process are presented in appendix B.

To construct the training data, I manually classify 30,000 randomly selected FLS along the tone and content dimensions. First, every sentence is classified into one of four tones: positive, neutral, negative, and uncertain.[8] The "uncertain" category is added because managers tend to

[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% and 90% for 10-Ks. The details of the MD&A extraction 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.

[8] "Positive" has a richer meaning than "optimistic"—while "optimistic" means "tending, or conforming, to the opinion that all events are ordered for the best," "positive" can have other meanings (e.g., "fully assured," "confident," "certain," "sometimes overconfident," "dogmatic," "overbearing"). Some prior studies and computational linguistic dictionaries use the term "positive/negative," others use "optimism/pessimism." Kothari, Li, and Short [2009] calculate positive/negative tone of disclosures using the General Inquirer. The 2007 version of the LIWC dictionary classifies "positive emotion" words and "negative emotion" words (which include three subcategories: "Anxiety," "Anger," and "Sadness") as part of the "Affective processes" and do not have an "Optimism" category. Davis, Piger, and Sedor [2005] define optimistic (pessimistic) tone as the sum of words from the praise, satisfaction, and inspiration

TABLE 1
Percentage Distributions of MD&A FLS Tone and Content

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*

This table shows the percentage distributions of the 30,000 sentences (i.e., the training data) that are manually coded into different tone and content categories. The 30,000 forward-looking sentences are extracted randomly from the MD&As. Fifteen research assistants manually categorize them along two dimensions: tone and content. The details of the procedures are presented in appendices C and D.

convey negative information by using words like "risk" and "uncertainty" (Li [2010]). A typical uncertain tone sentence is 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 content of the FLS into 12 categories, the details of which are shown in appendix C. The details of the training data preparation are shown in appendix D.

Table 1 reports the descriptive statistics of the training data set. Of the 30,000 FLS that are manually classified, 19.59% are coded as being of positive tone, 39.97% neutral, 17.82% negative, and 22.55% uncertain. Somewhat inconsistent with the findings in Pava and Epstein [1993], there is a fair amount of negative tone in the MD&As with a percentage close to that of the positive tone group. Prior research shows that managers tend to express their negative views using words like "risk" and "uncertain" (Li [2010]). Therefore, the 22.55% sentences with uncertain tone are also possibly of negative tone. In the empirical analysis, I combine the uncertain tone group with the negative tone group, but the results are robust to excluding sentences in the uncertain category.

Table 1 also shows the percentages of the training data for each content category. The sum of the percentages across the 12 categories is greater

---

categories (blame, hardship, and denial categories) using Diction, but they also use the term "positive/negative tone" in the paper. Rogers, Buskirk, and Zechman [2009] calculate their "net optimism" measure as the sum of three positive components (praise, satisfaction, and inspiration) minus the sum of three negative components (blame, hardship, and denial) from Diction. In this paper, I standardize the terms to "positive/negative," rather than "optimistic/pessimistic."

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 classified as both category 4 (operations) and category 5 (liquidity).

The first four categories, which account for 62.81% of the sample, relate to revenue, cost, profitability, and operations. Category 5 (liquidity) is discussed in 11.57% of the sentences, Category 6 (investment) in 10.79%, and Category 7 (financing) in 16.45%. These last three categories are combined in later analyses. The remaining five categories, combining to less than 14%, represent a very small share of the MD&A discussions. The fact that most of the FLS are about profitability, liquidity, and capital resources, as evidenced by the combined 101.62% in categories 1 to 7, is consistent with the SEC's intention that the MD&A section provide investors with information primarily about capital resources and liquidity.

4.2 COMPUTATION

I use the Algorithm::NaiveBayes module in Perl to conduct the computation. I first convert the vector of words for each sentence after the stemming and stopwording processes into a hash variable in Perl. (See appendix B for a complete description of the processes.) I then feed the hash variables from the 30,000 manually coded sentences into the Bayesian classifier and run the training process. After this step, the algorithm predicts the tone and category of all the FLS, which total to about 13 million sentences.

4.3 VALIDATION OF THE ALGORITHM

There are three common methods for evaluating the effectiveness of a text classification algorithm: training error, train and test, and N-fold cross-validation.[9] To evaluate the effectiveness of the Naïve Bayesian algorithm, I first calculate the training errors. Untabulated results show that training errors are, in general, less than 10%. This means that if the algorithm is used to predict the sentences in the training data set (after learning from

[9] In the case of training error calculation, the classifier is trained on a training data set and evaluated on the same data set. Although this method is obviously biased toward the training data, it can detect underfitting of the data. In the "train and test" method, the data are divided into two parts: training and testing. The split is usually 90% for training and 10% for testing. This is an unbiased evaluation and can detect underfitting as well as overfitting. The theory of cross-validation was developed by Geisser [1975]. It is useful in guarding against testing hypotheses suggested by the data (Type III error), especially where further samples are hazardous, costly, or impossible. In the N-fold cross-validation test, 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 the remaining $N - 1$ parts are used for training. At the end, the results over the N experiments are averaged. This unbiased testing gives more statistical significance than the train-and-test method, but it is not applicable for examining the training data during classifier construction.

TABLE 2
N-fold Cross-Validation Tests

	Tone		Content	
N	4 Categories	3 Categories	12 Categories	3 Categories
Bayesian learning				
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				
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

This table reports the N-fold cross-validation test results for the machine learning algorithm and for an "informed guessing" strategy. In the "Bayesian learning" section of the table, for each N, the 30,000 training sentences 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 using the Bayesian learning algorithm. The average percentage of correct classifications is reported for each N. For the tone, the algorithm classifies each sentence into four possible categories (positive, neutral, negative, and uncertain). The N-fold tests for the four categories are reported under "Tone-4 categories" and "Tone-3 categories," with the negative and uncertain categories combined. For the content, the algorithm classifies each sentence into 12 possible categories (details in appendix C) and the N-fold tests for the 12 categories are reported under "Content-12 categories" and "Content-3 categories," with Categories 1 to 4 combined as a "profitability" category, Categories 5 to 7 combined as a "liquidity" category, and Categories 8 to 12 combined as an "other" category. In the "Informed guessing" section of the table, for each N-fold cross-validation tests, classification success rates are calculated using an "informed guessing" strategy. In this strategy, the probability for each tone/content category is calculated using the learning data (i.e., $(N - 1)/N$ of the 30,000 sentences) and the sentences in the predicting set (i.e., $1/N$ of the 30,000 sentences) are classified following these probabilities.

the same data set), it will correctly classify the categories and tone more than 90% of the time. This suggests that the chances of underfitting the data are small.

As the "train and test" method can be seen as a special case of the N-fold cross validation test, I next report the empirical validation of the Naïve Bayesian learning algorithm using the N-fold cross-validation method, with N varying from 3 to 50. For instance, when $N = 3$, I randomly divide the 30,000 sentences of training data into three 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 threefold cross-validation result.

Table 2 reports the average correct classification rate for the different levels of N. As shown in row (1) of table 2, in the threefold cross-validation test for the four-category tone (positive, neutral, negative, and uncertain), the learning algorithm classifies the testing sentences correctly 59.15% of the time. This is a higher rate than that obtained with an informed guessing approach, where the percentage of each of the four tone categories is calculated using the learning data and is applied naïvely to the testing data. With this method, the correct classification rate is only 32.44%. If we combine the negative and uncertain tones into one category, the Bayesian algorithm success rate increases to 66.95% in the threefold cross-validation test and the informed guessing rate becomes 40.47%. These results are stable when N increases from 3 to 50. The success rate of the four-category tone by the Bayesian algorithm is about 59% and that of the three-category tone is about 67%. Both are much higher than the rate obtained from the informed guessing strategy of about 32% and 40% for the four-category and three-category tests, respectively.

Classifying every FLS into a content category out of 12 possible categories as defined in appendix C, 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 three groups—profit (categories 1 to 4), liquidity (categories 5 to 7), and other (categories 8 to 12), the Bayesian algorithm achieves a success rate of more than 82%, while the informed guessing rate is about 44%. Overall, the N-fold cross-validation tests show that the Bayesian learning algorithm achieves a good classification rate compared with an informed guessing strategy.

Even though the Bayesian algorithm performs well in the cross-validation tests, a caveat is in order. The algorithm is used at a sentence-level analysis in this paper but it may not work well for analyzing passages. My current application of the Bayesian algorithm at the sentence level assumes that the tone of a paragraph or a document is the simple average of the tone of all sentences. This assumption could be invalid; for instance, some words in a subset of sentences may change the meaning of the entire document dramatically. Whether the Bayesian algorithm works for analyzing a passage (rather than a sentence) in the financial statement setting is an empirical question for future research.

5. *Information Content of MD&A FLS*

5.1 DESCRIPTIVE STATISTICS FOR MD&A TONE AND CONTENT

I start with all 10-K and 10-Q filings between 1994 and 2007 from the SEC Edgar Web site. To be included in the final data set, 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 Center for Research in Security Prices (CRSP), (2) quarterly earnings (item 69 in the Quarterly file) and cash flows from operations (item 108) from Compustat[10], (3) stock returns in CRSP, and (4) at least five sentences of FLS. Although the requirement of at least five sentences of FLS is arbitrary, its purpose is to ensure

[10] For cash flow statement items, Compustat reports data for the cumulative interim period year to date. Therefore, for the second, third, and fourth quarters, the differences between the data item 108 in quarter t and quarter $t-1$ are computed to arrive at the correct amount of cash flow from operations for the three months ended in the current quarter.

that the empirical measures derived from the filing are not due to 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 (i.e., the predicted probability of the sentence being positive is higher than the probability of being any of the other three categories); $TONE_k = 0$ if the prediction is neutral; and $TONE_k = -1$ if the prediction is negative or uncertain. For every firm i in quarter j, I define its MD&A tone as the average tone of all the forward-looking sentences in that filing as predicted by the learning algorithm:

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

where K is the total number of FLS. By construction, $TONE_{ij}$ is a variable that is between -1 and 1, with 1 being completely positive and -1 being completely negative. The more positive $TONE_{ij}$ is, the more positive the tone of the FLS made by firm i in quarter j's 10-Q or 10-K filing.

Table 3 shows the descriptive statistics for MD&A tone for the final sample of 145,479 firm-quarters. On average, the FLS in MD&A disclosures are negative, as indicated by a mean (median) of $TONE$ of -0.23 (-0.21).

TABLE 3

Descriptive Statistics

Variable	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
EARNVOL	0.06	-	0.00	0.01	0.02	0.06	0.19	0.24
FOG	18.31	-	15.07	16.89	18.24	19.64	21.76	2.08
NITEMS	223.15	-	194	216	226	233	247	16.95
NBSEG	0.87	-	0.00	0.00	0.69	1.39	2.48	0.89
NGSEG	0.81	-	0.00	0.00	0.00	1.39	2.56	0.94
FIRMAGE	11.34	-	1	4	8	16	32	9.68
MA	0.07	-	–	–	–	–	–	0.26
SEO	0.02	-	–	–	–	–	–	0.14
SI	−0.01	-	−0.03	0.00	0.00	0.00	0.01	0.07
DLW	0.61	-	–	–	–	–	–	0.49

(*Continued*)

TABLE 3—*Continued*

	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

This table shows the descriptive statistics for 145,479 sample firm-quarters. The column labeled "Pr(=0)" is the p-value of the t-test testing whether a variable equals 0. In the second part of panel A, the column "p-value" shows the p-value of the t-test testing whether the mean of two variables are the same. The variables are defined as follows: *TONE* is the average tone of the FLS 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 FLS 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 appendix C). *LIQUIDITY_TONE* is the average tone of the FLS 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 appendix C). *OTHER_TONE* is the average tone of the FLS of a firm-quarter that are about other topics (i.e., the statements that are classified as Categories 8 to 12 as defined in appendix C). *PROFIT_PCT* is the percentage of the FLS 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 appendix C). *LIQUIDITY_PCT* is the percentage of the FLS 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 appendix C). *OTHER_PCT* is the percentage of the FLS of a firm-quarter that are about other topics (i.e., the statements that are classified as Categories 8 to 12 as defined in appendix C). *EARN* is the quarterly earnings (Compustat Quarterly file data item 69) scaled by the book value of assets (Compustat Quarterly file data item 44), winsorized at −3 and 3. *RET* is the contemporaneous stock returns in the fiscal quarter calculated using CRSP monthly return data. *CFRATIO* is the quarterly cash flows from operations (Compustat data item 108) scaled by book value of current liability (Compustat Quarterly file data item 49); for the second, third, and fourth quarters, the differences between the data item 108 in quarter t and quarter $t-1$ are computed to arrive at the correct amount of cash flow from operations for the three months ended in the current quarter. *ACC* is the accruals (earnings subtract cash flow from operations) scaled by the book value of assets (Compustat Quarterly file data item 44). *SIZE* is the logarithm of the market value of equity at the end of the quarter (Compustat Quarterly file data item 14 times item 61). *MTB* is the market value of equity (Compustat Quarterly file data item 14 times item 61) plus the book value of total liabilities (Compustat Quarterly file data item 54) scaled by the book value of total assets (Compustat Quarterly file data item 44). *RETVOL* is the stock return volatility calculated using 12 months of monthly return data before the fiscal quarter ending date. *EARNVOL* is the standard deviation of earnings (Compustat item 69, scaled by book value of assets, Compustat item 44) calculated using data from the last five years. *FOG* is the Fog index of the MD&A. *NITEMS* is the number of nonmissing items in Compustat. *NBSEG* is the logarithm of 1 plus the number of business segment. *NGSEG* is the logarithm of 1 plus the number of geographic segment. *FIRMAGE* is the number of years since a firm appears in CRSP monthly file. *MA* is a dummy variable that equals 1 if a firm makes a merger or acquisition in a given fiscal quarter and 0 otherwise, calculated using data from SDC platinum. *SEO* is a dummy that equals 1 if a firm has seasoned equity offering in a fiscal quarter and 0 otherwise, calculated using data from SDC platinum. *SI* is the amount of special items reported for the quarter (Compustat item 32) scaled by the book value of assets (Compustat item 44). *DLW* is a dummy variable that equals 1 if a firm is incorporated in Delaware and 0 otherwise.

The mean of *TONE* is significantly different from 0 with a p-value of 0.000 in a t-test. I decompose *TONE* into three different components: *PROFIT_TONE* (the average tone of sentences in categories 1 to 4 as described in appendix C), *LIQUIDITY_TONE* (the average tone of sentences in categories 5 to 7), and *OTHER_TONE* (the average tone of sentences in categories 8 to 12). The negative tone of the MD&A FLS is mainly from 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 3 also presents the descriptive statistics for the content of the MD&A FLS. *PROFIT_PCT* is the percentage of content devoted to profitability and operations (categories 1 to 4) as predicted by the learning algorithm and *LIQUIDITY_PCT* is the percentage devoted to liquidity and capital resources. The means (medians) of *PROFIT_PCT* and *LIQUIDITY_PCT* are 53.65% (55.32%) and 32.99% (29.63%), respectively.

In table 4, the Pearson correlations for tone and content are reported. There is a significant negative correlation between $TONE$ and *PROFIT_PCT* (−0.47). This indicates that when there is more discussion of future profitability, the average tone is more negative. This is consistent with the negative mean of *PROFIT_TONE* documented in table 3. The positive correlation between $TONE$ and *LIQUIDITY_PCT* (0.52) shows that when there is more discussion of liquidity and capital resources, the average tone is more positive. Current earnings ($EARN$) are positively correlated with tone (with a Pearson correlation coefficient of 0.142). Firms with more positive accruals tend to have more negative MD&A statements with the correlation between ACC and $TONE$ being −0.082.

Figure 1 plots the means for tone for firms sorted into quintiles. In quarter 0, all firms are sorted into five quintiles based on that quarter's tone, with Quintile 1 firms having the most negative tone and quintile 5 the most positive. 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.

5.2 DETERMINANTS OF MD&A TONE AND CONTENT

Table 5 reports the OLS regression results for $TONE$ when it is regressed 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), $EARNVOL$ (earnings volatility), FOG (the Fog index of the MD&A), $NITEMS$ (the number of nonmissing items in Compustat), $NBSEG$ (the logarithm of one plus the number of business segment), $NGSEG$ (the logarithm of one plus the number of geographic segment), $FIRMAGE$ (firm age), MA (a dummy variable that equals one if a firm makes a merger or acquisition in a given fiscal quarter and zero otherwise), SEO (a dummy that equals one if a firm has seasoned equity offering in a fiscal quarter and zero otherwise), SI (the amount of special items reported for the quarter), DLW (a dummy variable that equals one if a firm is incorporated in Delaware and zero otherwise), and three reporting quarter dummies, $Q2$, $Q3$, and $Q4$. Year-fixed effects are also included in the regression. As there are likely to be both cross-sectional correlations and autocorrelations, the standard errors are calculated with two-way clustering by quarter and firm. All empirical inferences remain unchanged if the following specifications are used: (1) year- and industry-fixed effects are included with standard errors calculated by clustering by year, firm, or industry; (2) Fama-MacBeth

TABLE 4
Correlation Matrix (p-values in Parentheses)

Variables	TONE	PROFIT_ TONE	LIQUIDITY_ TONE	OTHER_ TONE	PROFIT_ PCT	LIQUIDITY_ PCT	EARN	RET	ACC	SIZE	MTB	RETVOL	FOG	FIRMAGE
TONE	1.00													
PROFIT_TONE	0.78	1.00												
	(0.00)													
LIQUIDITY_TONE	0.54	0.29	1.00											
	(0.00)	(0.00)												
OTHER_TONE	0.49	0.26	0.23	1.00										
	(0.00)	(0.00)	(0.00)											
PROFIT_CAT	−0.47	−0.21	0.03	−0.14	1.00									
	(0.00)	(0.00)	(0.00)	(0.00)										
LIQUIDITY_CAT	0.53	0.23	0.01	0.22	−0.83	1.00								
	(0.00)	(0.00)	(0.05)	(0.00)	(0.00)									
OTHER_CAT	−0.05	−0.02	−0.06	−0.12	−0.36	−0.21	1.00							
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)								
EARN	0.15	0.11	0.16	0.07	−0.03	0.02	0.02	1.00						
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)							
RET	−0.01	−0.01	0.00	0.00	−0.01	−0.00	0.02	0.08	1.00					
	(0.06)	(0.00)	(0.58)	(0.68)	(0.00)	(0.87)	(0.00)	(0.00)						
ACC	0.05	0.03	0.06	0.01	−0.01	0.02	−0.00	0.75	0.05	1.00				
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.16)	(0.00)	(0.00)					
SIZE	0.05	0.07	−0.01	−0.01	−0.05	−0.02	0.12	0.22	0.10	0.09	1.00			
	(0.00)	(0.00)	(0.00)	(0.03)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)				
MTB	−0.17	−0.13	−0.12	−0.08	0.10	−0.09	−0.02	−0.21	0.22	−0.06	0.15	1.00		
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)			
RETVOL	−0.26	−0.20	−0.17	−0.09	0.16	−0.11	−0.10	−0.25	0.17	−0.11	−0.33	0.20	1.00	
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)		

(*Continued*)

TABLE 4—*Continued*

Variables	TONE	PROFIT_ TONE	LIQUIDITY_ TONE	OTHER_ TONE	PROFIT_ PCT	LIQUIDITY_ PCT	EARN	RET	ACC	SIZE	MTB	RETVOL	FOG	FIRMAGE	
FOG	−0.25	−0.18	−0.18	−0.16	0.09	−0.10	0.02	−0.15	−0.00	−0.03	−0.14	0.14	0.14	1.00	
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.44)	(0.00)	(0.00)	(0.00)	(0.00)		
FIRMAGE	0.19	0.17	0.16	0.07	−0.06	0.00	0.10	0.12	0.01	0.05	0.17	−0.10	−0.25	−0.08	1.00
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.36)	(0.00)	(0.00)	(0.03)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	

This table shows the pair-wise Pearson correlation coefficients of selected variables with the p-values testing whether the correlation coefficients are significantly different from 0 in the parentheses. The variables are defined as follows: *TONE* is the average tone of the FLS 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 FLS 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 appendix C). *LIQUIDITY_TONE* is the average tone of the FLS 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 appendix C). *OTHER_TONE* is the average tone of the FLS of a firm-quarter that are about other topics (i.e., the statements that are classified as Categories 8 to 12 as defined in appendix C). *PROFIT_PCT* is the percentage of the FLS 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 appendix C). *LIQUIDITY_PCT* is the percentage of the FLS 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 appendix C). *OTHER_PCT* is the percentage of the FLS of a firm-quarter that are about other topics (i.e., the statements that are classified as Categories 8 to 12 as defined in appendix C). *EARN* is the quarterly earnings (Compustat Quarterly file data item 69) scaled by the book value of assets (Compustat Quarterly file data item 44), winsorized at −3 and 3. *RET* is the contemporaneous stock returns in the fiscal quarter calculated using CRSP monthly return data. *CFRATIO* is the quarterly cash flows from operations (Compustat data item 108) scaled by book value of current liability (Compustat Quarterly file data item 49); for the second, third, and fourth quarters, the differences between the data item 108 in quarter t and quarter $t-1$ are computed to arrive at the correct amount of cash flow from operations for the three months ended in the current quarter. *ACC* is the accruals (earnings subtract cash flow from operations) scaled by the book value of assets (Compustat Quarterly file data item 44). *SIZE* is the logarithm of the market value of equity at the end of the quarter (Compustat Quarterly file data item 14 times item 61). *MTB* is the market value of equity (Compustat Quarterly file data item 14 times item 61) plus the book value of total liabilities (Compustat Quarterly file data item 54) scaled by the book value of total assets (Compustat Quarterly file data item 44). *RETVOL* is the stock return volatility calculated using 12 months of monthly return data before the fiscal quarter ending date. *FOG* is the Fog index of the MD&A. *FIRMAGE* is the number of years since a firm appears in CRSP monthly file.

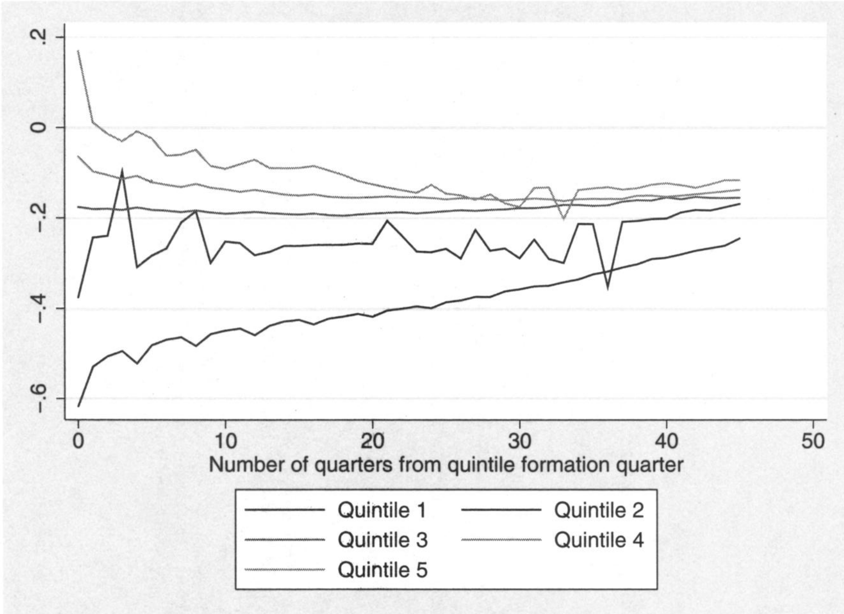

FIG. 1.—MD&A Tone over Time by Quintile. This figure shows the average tone of the 10-K and 10-Q MD&As for the five quintile portfolios sorted on *TONE* in the 48 quarters after the portfolios are formed. Every quarter, five portfolios are formed by sorting *TONE*, the average tone of all the FLS in a firm's MD&A (a forward-looking sentence's tone has a value of 1 if the Bayesian learning algorithm classifies the sentence as positive, 0 if neutral, and −1 if negative or uncertain). In quarter 0, firms are sorted into 5 quintiles with firms in quintile 1 having the most negative MD&A tone and quintile 5 firms having the most positive tone. The average tone of the five-quintile firms is plotted over time for the next 48 quarters.

approach is used; or (3) standard errors are estimated using bootstrapping with clustering at quarter or firm level.

Column (1) of table 5 includes earnings, returns, accruals, size, MTB, return volatility, and earnings volatility in the analysis. The results indicate that tone is positively related to current performance (the coefficient on *EARN* is 0.293 with a *t*-statistic of 9.02 and that on *RET* is 0.047 with a *t*-statistic of 6.01) and confirm the univariate correlations in table 4. This suggests that when a firm is performing well, managers discuss its future outlook in a more positive tone. Accruals are significantly negatively related to *TONE* (the coefficient on *ACC* being −0.208 with a *t*-statistic of −8.21). This suggests that when current accruals are very positive, management's discussion of the firm's future outlook is more negative. Given that accruals are negatively related to future performance, this suggests that managers understand the implications of accruals for future performance. Furthermore, bigger firms are more likely to use a negative tone in their MD&As as indicated by the negative coefficient on *SIZE* of −0.003 with a *t*-statistic of −1.74. This is consistent with the hypothesis that large firms

TABLE 5
Determinants of MD&A Tone

COEFFICIENT	(1) TONE	(2) TONE	(3) TONE
EARN	0.293***	0.295***	0.212***
	(9.02)	(7.86)	(6.18)
RET	0.047***	0.042***	0.038***
	(6.01)	(5.81)	(5.76)
ACC	−0.208***	−0.188***	−0.139***
	(−8.21)	(−7.77)	(−6.25)
SIZE	−0.003*	−0.005***	−0.008***
	(−1.74)	(−2.86)	(−4.19)
MT	−0.015***	−0.013***	−0.010***
	(−10.99)	(−10.02)	(−8.34)
RETVOL	−0.542***	−0.453***	−0.418***
	(−13.71)	(−12.68)	(−12.49)
EARNVOL	−0.029**	−0.017	−0.006
	(−2.11)	(−1.48)	(−0.58)
NITEMS		−0.001***	−0.002***
		(−7.33)	(−8.62)
NBSEG		0.017***	0.018***
		(4.54)	(4.96)
NGSEG		−0.010***	−0.015***
		(−2.69)	(−3.96)
FIRMAGE		0.004***	0.004***
		(11.65)	(12.04)
MA		−0.002	0.000
		(−0.45)	(0.11)
SEO		0.034***	0.026***
		(3.80)	(2.78)
SI		−0.068**	−0.038
		(−2.03)	(−1.30)
DLW		−0.043***	−0.037***
		(−7.10)	(−6.33)
Q2	−0.005	−0.006	−0.022***
	(−1.03)	(−1.17)	(−4.46)
Q3	−0.011*	−0.013**	−0.024***
	(−1.85)	(−2.21)	(−4.42)

(*Continued*)

are more cautious in their disclosures due to political and legal concerns. Additionally, firms with high MTB ratios have a less positive tone in their MD&As, consistent with the hypothesis that growth firms have more uncertain information environments and are thus more conservative in discussing future events. Firms with more volatile earnings and returns tend to have a less positive tone when discussing their future outlook (the coefficient on $RETVOL$ is −0.542 with a t-value of −10.99).[11] This suggests that firms with more volatile business environments may be more cautious in

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

TABLE 5—*Continued*

COEFFICIENT	(1) TONE	(2) TONE	(3) TONE
Q4	−0.016*	−0.019**	−0.030***
	(−1.67)	(−1.98)	(−3.32)
FOG			−0.029***
			(−17.87)
Observations	115,949	106,050	105,846
R^2	0.14	0.17	0.21

This table shows the regression results of MD&A tone on its potential determinants. The dependent variable is *TONE*, which is the average tone of the FLS of a firm-quarter. A forward-looking sentence's tone has a value of 1 if the Bayesian learning algorithm classifies the sentence as positive, 0 if neutral, and −1 if negative or uncertain. The independent variables include *EARN*, *RET*, *ACC*, *SIZE*, *MTB*, *RETVOL*, *EARNVOL*, *FOG*, *NITEMS*, *NBSEG*, *NGSEG*, *FIRMAGE*, *MA*, *SEO*, *SI*, *DLW*, *Q2*, *Q3*, and *Q4*. *EARN* is the quarterly earnings (Compustat Quarterly file data item 69) scaled by the book value of assets (Compustat Quarterly file data item 44), winsorized at −3 and 3. *RET* is the contemporaneous stock returns in the fiscal quarter calculated using CRSP monthly return data. *CFRATIO* is the quarterly cash flows from operations (Compustat data item 108) scaled by book value of current liability (Compustat Quarterly file data item 49); for the second, third, and fourth quarters, the differences between the data item 108 in quarter t and quarter $t-1$ are computed to arrive at the correct amount of cash flow from operations for the three months ended in the current quarter. *ACC* is the accruals (earnings minus cash flow from operations) scaled by the book value of assets (Compustat Quarterly file data item 44). *SIZE* is the logarithm of the market value of equity at the end of the quarter (Compustat Quarterly file data item 14 times item 61). *MTB* is the market value of equity (Compustat Quarterly file data item 14 times item 61) plus the book value of total liabilities (Compustat Quarterly file data item 54) scaled by the book value of total assets (Compustat Quarterly file data item 44). *RETVOL* is the stock return volatility calculated using 12 months of monthly return data before the fiscal quarter ending date. *EARNVOL* is the standard deviation of earnings (Compustat item 69, scaled by book value of assets, Compustat item 44) calculated using data from the last five years. *FOG* is the Fog index of the MD&A. *NITEMS* is the number of non-missing items in Compustat. *NBSEG* is the logarithm of 1 plus the number of business segment. *NGSEG* is the logarithm of 1 plus the number of geographic segment. *FIRMAGE* is the number of years since a firm appears in CRSP's monthly file. *MA* is a dummy variable that equals 1 if a firm makes a merger or acquisition in a given fiscal quarter and 0 otherwise, calculated using data from SDC platinum. *SEO* is a dummy that equals 1 if a firm has seasoned equity offering in a fiscal quarter and 0 otherwise, calculated using data from SDC platinum. *SI* is the amount of special items reported for the quarter (Compustat item 32) scaled by the book value of assets (Compustat item 44). *DLW* is a dummy variable that equals 1 if a firm is incorporated in Delaware and 0 otherwise. *Q2* (*Q3* or *Q4*) is a dummy variable that is set to 1 if the current reporting quarter is the second (third or fourth) fiscal quarter. Year fixed effects are included in the regressions, but are not reported. *T*-statistics based on two-way clustering at both year-quarter level and firm level are reported in parentheses.

*** indicates $p < 0.01$; ** $p < 0.05$; * $p < 0.1$.

forward-looking disclosures because of either information uncertainty or potential legal concerns.

In column (2), additional determinants are included in the regression. The measures of operation complexity have different associations with tone. While firms with more business segments tend to have more positive tone, firms with more nonmissing financial items in Compustat and more geographic segments have more negative tone. More mature firms and firms that are involved in seasoned equity offerings also have more positive discussions about future outlook in MD&As. The coefficient on *FIRMAGE* is 0.004 with a *t*-statistic of 11.65 and that on *SEO* is 0.034 with a *t*-statistic of 3.80. Finally, firms incorporated in Delaware have more negative tone compared with firms not incorporated in Delaware as indicated by the negative coefficient on *DLW* of −0.043 with a *t*-statistic of −7.10. Column (3) of table 5 includes the Fog index of the MD&A in addition to

TABLE 6
Future Earnings and MD&A Tone

COEFFICIENT	(1) EARN($t+1$)	(2) EARN($t+2$)	(3) EARN($t+3$)	(4) EARN($t+4$)
TONE	0.006***	0.005***	0.004**	0.003
	(4.63)	(3.25)	(2.20)	(1.34)
EARN	0.679***	0.616***	0.618***	0.626***
	(24.27)	(21.11)	(21.52)	(22.56)
RET	0.011***	0.008***	0.006***	0.005**
	(6.59)	(4.05)	(4.64)	(2.54)
ACC	−0.240***	−0.243***	−0.257***	−0.272***
	(−10.43)	(−10.88)	(−12.32)	(−9.89)
SIZE	0.002***	0.002***	0.002***	0.002***
	(7.40)	(7.09)	(6.74)	(5.96)
MTB	−0.002***	−0.003***	−0.003***	−0.005***
	(−3.95)	(−4.55)	(−5.95)	(−6.03)
RETVOL	−0.048***	−0.055***	−0.053***	−0.046***
	(−8.22)	(−6.36)	(−5.47)	(−5.52)
EARNVOL	−0.016***	−0.017***	−0.016**	−0.006*
	(−3.06)	(−2.94)	(−2.55)	(−1.83)
FOG	−0.001***	−0.001***	−0.001***	−0.001***
	(−7.92)	(−6.15)	(−6.85)	(−5.27)
NITEMS	0.000	0.000	0.000	0.000
	(0.01)	(0.24)	(0.15)	(0.75)
NBSEG	0.000	0.000	0.000	−0.000
	(0.91)	(0.15)	(0.17)	(−0.69)
NGSEG	0.002***	0.002***	0.003***	0.003***
	(3.90)	(4.78)	(4.58)	(4.86)
FIRMAGE	0.000***	0.000***	0.000***	0.000***
	(5.14)	(4.48)	(4.56)	(3.36)

(*Continued*)

the variables examined in columns (1) and (2). The results indicate that firms with high Fog MD&A have more negative MD&A tone as indicated by the coefficient on *FOG* of −0.029 with a *t*-statistic of −17.87.

5.3 FLS TONE AND FUTURE PERFORMANCE

In this section, I examine the implications of the FLS tone generated by the Naïve Bayesian algorithm for a firm's future performance beyond that contained in the numeric financial information. To do so, I first check the link between *TONE* and future profitability.[12] Table 6 shows the

[12] Li [2008] examines the relation between annual report Fog index and earnings persistence because the obfuscation hypothesis argues that managers have incentives to disclose less transparently if current good performance is less persistent. In this paper, I do not hypothesize any asymmetric implications of *TONE*. As a result, I focus on future earnings rather than earnings persistence. Unreported results also indicate that *TONE* is not significantly associated with earnings persistence.

TABLE 6—*Continued*

COEFFICIENT	(1) EARN($t+1$)	(2) EARN($t+2$)	(3) EARN($t+3$)	(4) EARN($t+4$)
MA	0.001	−0.000	−0.001	−0.004***
	(1.31)	(−0.12)	(−0.87)	(−2.62)
SEO	0.002*	0.000	0.002	0.003*
	(1.84)	(0.12)	(1.49)	(1.68)
SI	−0.422***	−0.388***	−0.380***	−0.353***
	(−10.35)	(−11.25)	(−10.71)	(−6.94)
DLW	−0.002***	−0.002***	−0.003***	−0.003***
	(−3.00)	(−3.32)	(−3.21)	(−2.69)
Q2	−0.004***	−0.011***	0.005***	−0.002**
	(−5.84)	(−7.97)	(3.74)	(−2.34)
Q3	−0.012***	−0.004***	0.005***	−0.004***
	(−9.55)	(−3.88)	(3.38)	(−4.48)
Q4	−0.005***	−0.004***	0.003**	−0.011***
	(−4.66)	(−3.30)	(2.14)	(−7.99)
Observations	95,325	91,350	87,348	83,316
R^2	0.36	0.29	0.25	0.24

This table shows the regression results of future earnings on MD&A tone and other control variables. The dependent variables are the earnings in the next four quarters (Compustat Quarterly file data item 69) scaled by the book value of assets at the end of the current quarter (Compustat Quarterly file data item 44). Independent variable *TONE* is the average tone of the FLS of a firm-quarter. A forward-looking sentence's tone has a value of 1 if the Bayesian learning algorithm classifies the sentence as positive, 0 if neutral, and −1 if negative or uncertain. Other independent variables include *EARN*, *RET*, *ACC*, *SIZE*, *MTB*, *RETVOL*, *EARNVOL*, *FOG*, *NITEMS*, *NBSEG*, *NGSEG*, *FIRMAGE*, *MA*, *SEO*, *SI*, *DLW*, *Q2*, *Q3*, and *Q4*. *EARN* is the quarterly earnings (Compustat Quarterly file data item 69) scaled by the book value of assets (Compustat Quarterly file data item 44), winsorized at −3 and 3. *RET* is the contemporaneous stock returns in the fiscal quarter calculated using CRSP monthly return data. *ACC* is the accruals (earnings minus cash flow from operations) scaled by the book value of assets (Compustat Quarterly file data item 44). *SIZE* is the logarithm of the market value of equity at the end of the quarter (Compustat Quarterly file data item 14 times item 61). *MTB* is the market value of equity (Compustat Quarterly file data item 14 times item 61) plus the book value of total liabilities (Compustat Quarterly file data item 54) scaled by the book value of total assets (Compustat Quarterly file data item 44). *RETVOL* is the stock return volatility calculated using 12 months of monthly return data before the fiscal quarter ending date. *EARNVOL* is the standard deviation of earnings (Compustat item 69, scaled by book value of assets, Compustat item 44) calculated using data from the last five years. *FOG* is the Fog index of the MD&A. *NITEMS* is the number of nonmissing items in Compustat. *NBSEG* is the logarithm of 1 plus the number of business segment. *NGSEG* is the logarithm of 1 plus the number of geographic segment. *FIRMAGE* is the number of years since a firm appears in CRSP monthly file. *MA* is a dummy variable that equals 1 if a firm makes a merger or acquisition in a given fiscal quarter and 0 otherwise, calculated using data from SDC platinum. *SEO* is a dummy that equals 1 if a firm has seasoned equity offering in a fiscal quarter and 0 otherwise, calculated using data from SDC platinum. *SI* is the amount of special items reported for the quarter (Compustat item 32) scaled by the book value of assets (Compustat item 44). *DLW* is a dummy variable that equals 1 if a firm is incorporated in Delaware and 0 otherwise. *Q2* (*Q3* or *Q4*) is a dummy variable that is set to 1 if the current reporting quarter is the second (third or fourth) fiscal quarter. Year fixed effects are included in the regressions, but are not reported. *T*-statistics based on two-way clustering at both year-quarter level and firm level are reported in parentheses.

***indicates $p < 0.01$; **$p < 0.05$; *$p < 0.1$.

regression results for earnings in the next four quarters scaled by the book value of assets at the end of the current quarter ($EARN(t+1)$, $EARN(t+2)$, $EARN(t+3)$, and $EARN(t+4)$) when they are regressed on *TONE* and control variables including all the variables examined as the determinants of *TONE* as well as year-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 become even more significant statistically.[13]

In column (1), the coefficient on $TONE$ is 0.006 with a t-statistic of 4.63, suggesting that when managers are more positive in discussing a firm's future outlook in the MD&A, the earnings in the next quarter are significantly higher. The economic magnitude of this effect is substantial. The next quarter's earnings scaled by the book value of assets of firms with all positive FLS (i.e., $TONE = 1$) is higher than that of firms with extremely negative FLS (i.e., $TONE = -1$) by 1.2 percentage points (0.006×2). This translates into an annual difference in ROA of 5 percentage points. Examining this effect from a more realistic angle, we see that the interquartile range of $TONE$ validates this finding. From table 3, the 25th percentile and 75th percentile of $TONE$ are -0.41 and -0.03, respectively. Thus, an interquartile change in $TONE$ implies a difference in annual ROA of 1 percentage point even after controlling for other variables. 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 0.005 ($t = 3.25$), 0.004 ($t = 2.20$), and 0.003 ($t = 1.34$), respectively, showing that $TONE$ has diminishing predictive power for future earnings for at least three quarters after the current quarter. Not surprisingly, the coefficients on current earnings and stock returns both positively predict earnings in the next four quarters. However, current quarter accruals have a significantly negative relation with future performance consistent with prior findings (Sloan [1996]).

In table 7, the dependent variable is the change in earnings in the next four quarters. 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 7 is 0.006 ($t = 4.74$), the same as that in table 6. Unreported results also show that $TONE$ is positively and significantly associated with future liquidity as measured by operating cash flows divided by the current liabilities. It is interesting to test whether positive and negative tones have asymmetric implications for future performance. To examine this issue, in untabulated results, $TONE$ is interacted with a dummy variable $PTONE$, which is set to one if $TONE \geq 0$ and zero otherwise. The results indicate that there is no nonlinear relation between $TONE$ and future earnings as the coefficient on $TONE \times PTONE$ is insignificant. This suggests that positive and negative MD&As have similar implications for a firm's future earnings.

5.4 INFORMATION CONTENT OF MD&A OVER TIME

To assess whether the information content of MD&As changes over time, table 8 shows the regression of future earnings on $TONE$ and its

[13] The coefficients remain the same because all four equations contain the same explanatory variables.

TABLE 7
Future Earnings Changes and MD&A Tone

COEFFICIENT	(1) DEARN($t+1$)	(2) DEARN($t+2$)	(3) DEARN($t+3$)	(4) DEARN($t+4$)
TONE	0.006***	0.004***	0.003**	0.002
	(4.74)	(2.93)	(1.98)	(1.25)
EARN	−0.324***	−0.370***	−0.357***	−0.362***
	(−12.37)	(−12.29)	(−14.71)	(−12.14)
RET	0.011***	0.008***	0.006***	0.005**
	(6.69)	(4.10)	(4.51)	(2.51)
ACC	−0.245***	−0.244***	−0.259***	−0.273***
	(−11.41)	(−10.63)	(−11.69)	(−9.91)
SIZE	0.002***	0.002***	0.002***	0.002***
	(7.89)	(7.01)	(6.75)	(6.02)
MTB	−0.002***	−0.003***	−0.003***	−0.005***
	(−4.06)	(−4.23)	(−6.22)	(−6.02)
RETVOL	−0.049***	−0.052***	−0.050***	−0.045***
	(−8.28)	(−5.71)	(−5.53)	(−5.30)
EARNVOL	−0.015***	−0.018***	−0.016**	−0.006*
	(−3.06)	(−2.86)	(−2.48)	(−1.80)
FOG	−0.001***	−0.001***	−0.001***	−0.001***
	(−8.16)	(−5.49)	(−6.45)	(−5.13)
NITEMS	−0.000	0.000	0.000	0.000
	(−0.00)	(0.28)	(0.16)	(0.74)
NBSEG	0.000	0.000	0.000	−0.000
	(0.97)	(0.27)	(0.30)	(−0.68)
NGSEG	0.002***	0.002***	0.002***	0.003***
	(3.90)	(4.72)	(4.64)	(4.92)
FIRMAGE	0.000***	0.000***	0.000***	0.000***
	(5.40)	(3.91)	(4.34)	(3.24)
MA	0.001	−0.000	−0.001	−0.004***
	(1.14)	(−0.20)	(−0.99)	(−2.61)
SEO	0.002*	0.000	0.002	0.003*
	(1.88)	(0.15)	(1.58)	(1.71)
SI	−0.429***	−0.406***	−0.404***	−0.363***
	(−12.05)	(−10.26)	(−10.29)	(−6.99)
DLW	−0.002***	−0.002***	−0.002***	−0.003**
	(−2.98)	(−3.33)	(−2.99)	(−2.58)
Q2	−0.004***	−0.011***	0.005***	−0.002**
	(−5.92)	(−7.93)	(3.58)	(−2.30)

(*Continued*)

interaction with a time dummy, *POST*2003. *POST*2003 is equal to one if the report is filed in or after 2003 and zero otherwise. This test is designed to capture any systematic change in the information content of MD&As after the new SEC guidelines and the passage of the Sarbanes-Oxley Act, both of which significantly enhanced MD&A disclosure. A significant positive (negative) interaction term indicates that MD&As have become more (less) informative over time. In this test, three of the four interaction terms of *POST*2003 with *TONE* are statistically insignificant. The other

TABLE 7—*Continued*

COEFFICIENT	(1) DEARN($t+1$)	(2) DEARN($t+2$)	(3) DEARN($t+3$)	(4) DEARN($t+4$)
Q3	−0.012***	−0.004***	0.005***	−0.004***
	(−9.63)	(−3.93)	(3.34)	(−4.56)
Q4	−0.005***	−0.004***	0.003**	−0.012***
	(−4.90)	(−3.36)	(2.17)	(−7.97)
Observations	95,325	91,350	87,348	83,316
R^2	0.45	0.43	0.37	0.36

This table shows the regression results of future earnings changes on MD&A tone and other control variables. The dependent variables are the earnings in the next four quarters (Compustat Quarterly file data item 69) minus the earnings in the current quarter (Compustat Quarterly file data item 69) scaled by the book value of assets at the end of the current quarter (Compustat Quarterly file data item 44). The independent variable *TONE* is the average tone of the FLS of a firm-quarter. A forward-looking sentence's tone has a value of 1 if the Bayesian learning algorithm classifies the sentence as positive, 0 if neutral, and −1 if negative or uncertain. Other independent variables include *EARN*, *RET*, *ACC*, *SIZE*, *MTB*, *RETVOL*, *EARNVOL*, *FOG*, *NITEMS*, *NBSEG*, *NGSEG*, *FIRMAGE*, *MA*, *SEO*, *SI*, *DLW*, *Q*2, *Q*3, and *Q*4. *EARN* is the quarterly earnings (Compustat Quarterly file data item 69) scaled by the book value of assets (Compustat Quarterly file data item 44), winsorized at −3 and 3. *RET* is the contemporaneous stock returns in the fiscal quarter calculated using CRSP monthly return data. *ACC* is the accruals (earnings minus cash flow from operations) scaled by the book value of assets (Compustat Quarterly file data item 44). *SIZE* is the logarithm of the market value of equity at the end of the quarter (Compustat Quarterly file data item 14 times item 61). *MTB* is the market value of equity (Compustat Quarterly file data item 14 times item 61) plus the book value of total liabilities (Compustat Quarterly file data item 54) scaled by the book value of total assets (Compustat Quarterly file data item 44). *RETVOL* is the stock return volatility calculated using 12 months of monthly return data before the fiscal quarter ending date. *EARNVOL* is the standard deviation of earnings (Compustat item 69, scaled by book value of assets, Compustat item 44) calculated using data from the last five years. *FOG* is the Fog index of the MD&A. *NITEMS* is the number of nonmissing items in Compustat. *NBSEG* is the logarithm of 1 plus the number of business segment. *NGSEG* is the logarithm of 1 plus the number of geographic segment. *FIRMAGE* is the number of years since a firm appears in CRSP's monthly file. *MA* is a dummy variable that equals 1 if a firm makes a merger and acquisition in a given fiscal quarter and 0 otherwise, calculated using data from SDC platinum. *SEO* is a dummy that equals 1 if a firm has seasoned equity offering in a fiscal quarter and 0 otherwise, calculated using data from SDC platinum. *SI* is the amount of special items reported for the quarter (Compustat item 32) scaled by the book value of assets (Compustat item 44). *DLW* is a dummy variable that equals 1 if a firm is incorporated in Delaware and 0 otherwise. *Q*2 (*Q*3 or *Q*4) is a dummy variable that is set to 1 if the current reporting quarter is the second (third or fourth) fiscal quarter. Year fixed effects are included in the regressions, but are not reported. *T*-statistics based on two-way clustering at both year-quarter level and firm level are reported in parentheses.

*** indicates $p < 0.01$; ** $p < 0.05$; * $p < 0.1$.

interaction term is negative and significant. This result suggests that, despite the continuous effort by the SEC to strengthen the disclosure requirements for MD&As, there is no systematic change in their information content.[14]

5.5 AGGREGATE TONE

In this section, I investigate the aggregate tone for all U.S. public filers and examine whether there are any systematic changes in tone over time. Specifically, I calculate the tone of each MD&A in the 10-Ks and 10-Qs filed by firms with Compustat and CRSP coverage for each month and then plot the average tone over time.

[14] It is possible that other information sources become more important over time compared with MD&A.

TABLE 8
Information Content of MD&A Before and After 2003

COEFFICIENT	(1) EARN($t+1$)	(2) EARN($t+2$)	(3) EARN($t+3$)	(4) EARN($t+4$)
TONE	0.007***	0.006***	0.005***	0.002
	(4.57)	(4.15)	(2.88)	(1.10)
POST2003	−0.003*	−0.003	−0.006***	−0.001
	(−1.71)	(−1.61)	(−3.13)	(−0.61)
TONE× post2003	−0.003	−0.004**	−0.004	0.001
	(−1.30)	(−1.99)	(−1.26)	(0.29)
Observations	95,325	91,350	87,348	83,316
R^2	0.36	0.29	0.25	0.24

This table shows the regression results of future earnings and liquidity measures on MD&A tone, a dummy for post-2003, the interaction of MD&A tone and the post-2003 dummy, and other control variables. The dependent variables are the earnings in the next four quarters (Compustat Quarterly file data item 69) scaled by the book value of assets at the end of the current quarter (Compustat Quarterly file data item 44); for the second, third, and fourth quarters, the differences between the data item 108 in quarter t and quarter $t-1$ are computed to arrive at the correct amount of cash flow from operations for the three months ended in the current quarter. *TONE* is the average tone of the FLS of a firm-quarter. A forward-looking sentence's tone has a value of 1 if the Bayesian learning algorithm classifies the sentence as positive, 0 if neutral, and −1 if negative or uncertain. *POST*2003 is a dummy variable that equals 1 if the 10-K or 10-Q report is filed in or after 2003 and 0 otherwise. Other control variables (coefficients unreported) are the same as those in table 6. *T*-statistics based on two-way clustering at both year-quarter level and firm level are reported in parentheses.

***indicates $p < 0.01$; **$p < 0.05$; *$p < 0.1$.

Figure 2 plots the equal-weighted aggregate tone of MD&As over time. This figure shows substantial variation in the aggregate tone with three vertical lines indicating three possible dates of interest. The first vertical line indicates March 2000, when the NASDAQ index peaked at 4,572.83. Here, the figure shows that, contrary to the dramatic increase in the index in the late 1990s, the average tone in MD&As actually became more negative. The second vertical line shows that after the September 11, 2001 terrorist attack, there is a downward trend in management tone. It is not clear whether this reflects a revision in management expectation or is 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 tone.

Several possible reasons may explain the puzzling finding that the tone in MD&As dropped significantly in the 1990s when the stock market was going up. First, because of anxiety over Y2000 issues, managers may have devoted more discussion to these issues in an uncertain tone. Second, managers might have been anticipating the equity market downturn and thus have become more cautious. One possible explanation is the "market expectation hypothesis": when investors have high expectations for a firm's future performance, managers tend to become more cautious in their MD&A discussions. The cross-sectional test results in table 5 show that, consistent with this hypothesis, firms with a high MTB ratio tend to have more negative MD&As.

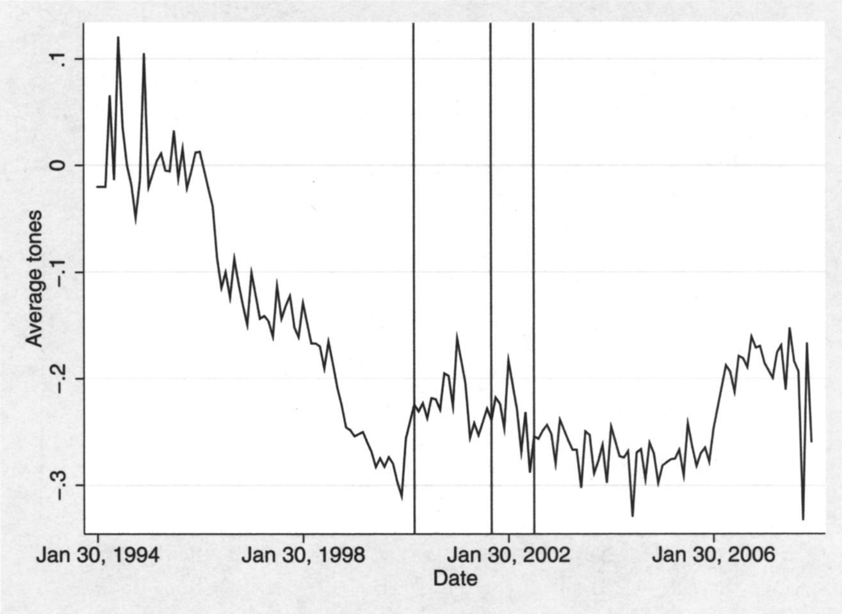

FIG. 2.—Aggregate Tone of MD&As over Time (All Firms). This figure shows the average *TONE* of all MD&As from the 10-Ks and 10-Qs filed in a given month between January 1994 and December 2007. *TONE* is the average tone of all the FLS in a firm's MD&A. A forward-looking sentence's tone has a value of 1 if the Bayesian learning algorithm classifies the sentence as positive, 0 if neutral, and −1 if negative or uncertain. The first vertical line indicates the month of March 2000 (when the NASDAQ index peaked), the second indicates the month of September 2001 (the terrorist attack event), and the third indicates the month of July 2002 (passage of the Sarbanes-Oxley Act).

This hypothesis can also be tested in a time-series setting at the macro level. The hypothesis is that, as the equity market index increases (which is the case between 1994 and 1999), managers face more pressure from the market and begin to show more caution in their MD&As and hence a significant drop in the aggregate MD&A tone. To test this hypothesis, I divide my sample into five quintiles based on their MTB ratios and then plot the time-series trend of the MD&A tone for each portfolio. Consistent with the hypothesis, figure 3 shows that firms in the top *MTB* quintile experience the biggest drop in the MD&A tone. During the mid-1990s, the aggregate tone of this portfolio is around 0. By the end of the 1990s it drops to about −0.4. In contrast, the bottom *MTB* quintile portfolio firms experience a drop from 0 in the mid-1990s to about −0.2 at the end of the 1990s as shown in figure 4 . Unreported results show that the differences in the drop in the MD&A tone between portfolios with different MTB ratios are statistically significant. Overall, the aggregate evidence suggests that market expectation is negatively correlated with manager's tone in the MD&A section.

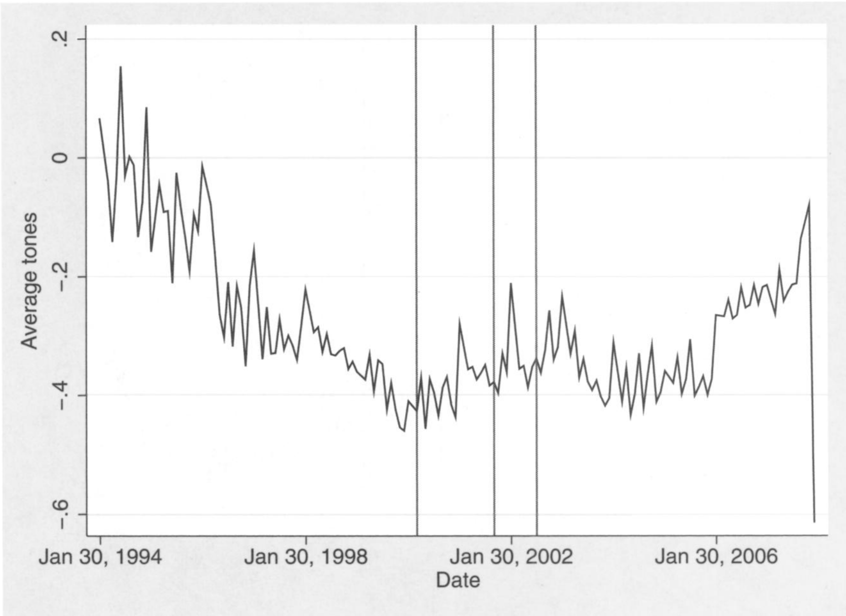

FIG. 3.—Aggregate Tone of MD&As over Time (High MTB Firms). This figure shows the average *TONE* of the MD&As from the 10-Ks and 10-Qs filed by high MTB ratio firms in a given month between January 1994 and December 2007. High MTB firms are those firms in the top quintile MTB portfolios for a month. *TONE* is the average tone of all the FLS in a firm's MD&A. A forward-looking sentence's tone has a value of 1 if the Bayesian learning algorithm classifies the sentence as positive, 0 if neutral, and −1 if negative or uncertain. The first vertical line indicates the month of March 2000 (when the NASDAQ index peaked), the second indicates the month of September 2001 (the terrorist attack event), and the third indicates the month of July 2002 (passage of the Sarbanes-Oxley Act).

5.6 MD&A TONE AND THE ACCRUAL ANOMALY

This section investigates the implications of the MD&A tone for the accrual anomaly. I calculate quarterly accruals using the Cash Flow Statement data because the evidence in Collins and Hribar [2002] suggests that accruals should be measured directly from the Statement of Cash Flows instead of the Balance Sheet accounts. Prior studies either use annual accruals or aggregate the most recent four quarters of accruals into pseudo-annual accruals to examine the accrual anomaly because of the seasonality concerns (e.g.,Green, Hand, and Soliman [2009]). I therefore construct and focus on the pseudo-annual accruals for quarter t by summing up the quarterly accruals from t, $t-1$, $t-2$, and $t-3$.

At the beginning of each month, I sort firms into decile portfolios based on the sum of the four most recently available accruals. Accruals are considered available if the related 10-K/Q filing date occurs prior to the beginning of the month. The accrual portfolios are then divided into two

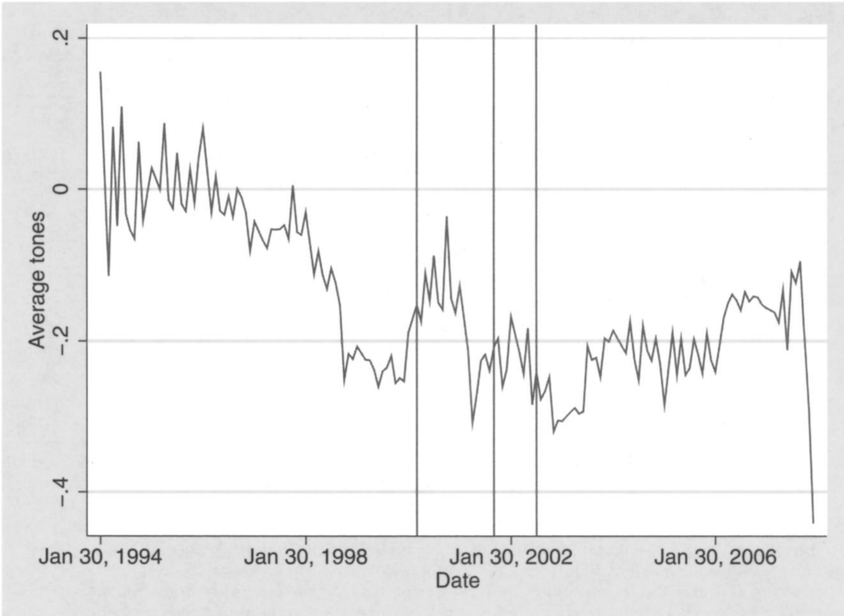

FIG. 4.—Aggregate Tone of MD&As over Time (Low MTB Firms). This figure shows the average *TONE* of the MD&As from the 10-Ks and 10-Qs filed by low MTB ratio firms in a given month between January 1994 and December 2007. Low MTB firms are those firms in the bottom quintile MTB portfolios for a month. *TONE* is the average tone of all the FLS in a firm's MD&A. A forward-looking sentence's tone has a value of 1 if the Bayesian learning algorithm classifies the sentence as positive, 0 if neutral, and −1 if negative or uncertain. The first vertical line indicates the month of March 2000 (when the NASDAQ index peaked); the second indicates the month of September 2001 (the terrorist attack event); and the third indicates the month of July 2002 (passage of the Sarbanes-Oxley Act).

subsamples based on the value of *WARN*, a dummy variable defined as follows:

$$WARN = \begin{cases} 1, & \text{if a firm has accruals below the median of all firms} \\ & \quad \text{and an MD\&A tone above the median;} \\ 1, & \text{if a firm has accruals above the median of all firms} \\ & \quad \text{and an MD\&A tone below the median;} \\ 0, & \text{if a firm has accruals above the median of all firms} \\ & \quad \text{and an MD\&A tone above the median;} \\ 0, & \text{if a firm has accruals below the median of all firms} \\ & \quad \text{and an MD\&A tone below the median.} \end{cases} \tag{2}$$

Like accruals, the variable *WARN* is updated on the first day of each month based on the most recent data. *WARN* is considered available if the related 10-K/Q filing date occurs prior to the beginning of the month. I then compute the value-weighted monthly return for each portfolio. Much

TABLE 9
Future Excess Returns of Value-Weight Portfolios Sorted on Accruals

	All Years								
	All Firms			$WARN = 0$ Firms			$WARN = 1$ Firms		
	D1	D10	D1–D10	D1	D10	D1–D10	D1	D10	D1–D10
Excess returns	0.93	−0.64	1.58	1.42	0.19	1.22	0.10	−0.57	0.67
T-statistic	1.74	−1.81	2.48	2.79	0.51	2.25	0.26	−1.01	0.47
	1994–2000								
Excess returns	1.86	−0.70	2.57	3.60	0.62	2.98	0.34	−0.70	1.04
T-statistic	2.48	−0.99	2.87	3.39	0.74	2.44	0.45	−0.60	0.71
	2001–2007								
Excess returns	0.08	−0.51	0.59	−0.21	−0.47	0.26	−0.42	0.13	−0.55
T-statistic	0.11	−1.31	0.69	−0.48	−0.77	0.33	−0.62	0.37	−0.68

This table shows the monthly excess value-weighted returns (in percentage) of the extreme accruals decile portfolios and hedge portfolios as a function of the MD&A tone. At the beginning of each month, I sort firms into deciles based on the sum of the four most recently available quarterly accruals (*ACC*). Accruals are considered available if the related 10-K/Q filing date occurs prior to the beginning of the month. I then compute the value-weighted monthly return for each portfolio. Quarterly accruals are calculated as earnings (Compustat Quarterly file data item 69) minus cash flow from operations (Compustat Quarterly file data item 108); for the second, third, and fourth quarters, the differences between the data item 108 in quarter t and quarter $t-1$ are computed to arrive at the correct amount of cash flow from operations for the three months ended in the current quarter. Accruals are scaled by the book value of assets (Compustat Quarterly file data item 44). Portfolio 1 (D1) has the lowest *ACC* and portfolio 10 (D10) has the highest *ACC*. The hedge portfolio (D1–D10) is a zero-investment portfolio long in D1 and short in D10. The excess returns of the portfolios are returns controlling for the market returns, size, book-to-market, and momentum factors, which are calculated as the intercepts from the time-series regressions of portfolio returns on $R_m - R_f$ (market returns minus risk-free returns), *SMB* (the Fama-French small-minus-big mimicking portfolio returns), *HML* (the Fama-French low-minus-high mimicking portfolio returns), and *MOM* (the Fama-French momentum mimicking portfolio returns). The table reports results for three samples. The "All firms" sample includes all firms; the full sample is also divided into two sub-samples: firms with $WARN = 1$ and those with $WARN = 0$. *WARN* is a dummy variable that equals 0 if *ACC* and *TONE* (both variables measured using data from the most recent four quarters) are both either above the median or below the median based on data from the most recent quarter, and 1 otherwise (i.e., when one is above median and the other is below median). Like accruals, *WARN* is updated on the first day of each month based on data from the most recent four quarters.

of the accrual anomaly literature focuses on the value-weighted portfolio returns, as Green, Hand, and Soliman [2009, section 2.1] argue that "equal-weighting tilts the portfolio heavily toward small cap stocks where the total costs of moving into and out of positions are much higher than they are for large cap stocks. As such, we expect that the value-weighted returns series are closer to those that would be obtained after accounting for such costs than are equally weighted returns." In addition, the potential biases in computed returns due to bid-ask spread bounces arising from the monthly rebalancing of the portfolios can be minimized by the value-weighted procedure (Blume and Stambaugh [1983]). Therefore, I present the value-weighted portfolio results. The equal-weighted results lead to the same empirical inferences and are available upon request.

Table 9 presents the future excess returns of the extreme accrual portfolios and a hedge portfolio. The time-series of the monthly portfolio returns are regressed on the Fama-French three factors returns and the momentum factor return. The intercepts from the regressions are presented for decile

1 portfolio (D1, i.e., firms in the lowest accruals decile), decile 10 portfolio (D10, i.e., firms in the highest accruals decile), and the zero-investment hedge portfolios (D1 minus D10). I first examine the accrual anomaly using the full sample. The results show that the monthly hedge portfolio return (i.e., buying firms with the lowest decile accruals and shorting the highest decile firms) is 1.58% ($t = 2.48$). This result is comparable to and consistent with the evidence documented in recent studies that examine the accrual anomaly using a similar time period (e.g., Green, Hand, and Soliman [2009]).

The sample is then divided into two sub-samples based on the value of $WARN$: $WARN = 0$ firms and $WARN = 1$ firm. The results in table 9 indicate that the accrual anomaly is statistically significant only when managers do not "warn" about the implications of accruals. The excess return of the hedge portfolio is 1.22% ($t = 2.25$) for the $WARN = 0$ sample but is only 0.67% and statistically insignificant ($t = 0.47$) for the $WARN = 1$ sample. Untabulated results indicate that the difference is statistically significant.

Richardson, Tuna, and Wysocki [2009] examine the accrual anomaly results by decades and find that there is a trend of attenuation over time that they interpret as "adaptive market efficiency." Green, Hand, and Soliman [2009] also document that the accrual anomaly seems to become weaker over time especially in the last 10 years. I therefore split my sample period (1994 to 2007) into two sub-periods (1994–2000 period and 2001–2007 period) and reconduct the empirical tests. The evidence in table 9 shows that, consistent with prior studies, the accrual anomaly disappears in the post-2001 years. In the 1994–2000 period, the accrual hedge portfolio generates positive excess returns (2.57% with a t-statistic of 2.87). During this period, the accruals hedge portfolio based on $WARN = 0$ firms generates significant excess returns (2.98% with a t-statistic of 2.44), while those based on firms with $WARN = 1$ do not have significant excess returns (1.04% with a t-statistic of 0.71).[15] Untabulated results also indicate that this difference is statistically significant. During the 2001–2007 period, the accrual anomaly disappears for the full sample (excess hedge return being 0.59% with a t-statistic of 0.69), the $WARN = 1$ sample (excess hedge return being -0.55% with a t-statistic of -0.68), and the $WARN = 0$ sample (excess hedge return being 0.26% with a t-statistic of 0.33).

Overall, the evidence is consistent with my hypothesis that management discussion about the implications of accruals for future performance mitigates the mispricing of accruals although over time the market seems to become more efficient and the mispricing of accruals disappears.[16]

[15] The monthly abnormal return of 2.98% implies an annualized return of about 36%, a substantial amount. Whether similar returns can be obtained in the pre-1994 years is an interesting question for future research.

[16] One concern of the empirical results in table 9 is that $WARN$ might be correlated with factors that determine the arbitrage costs for the accrual anomaly (e.g., Mashruwala, Rajgopal, and Shevlin [2006] show that the accrual anomaly is more significant for firms with higher

5.7 ADDITIONAL ANALYSIS

To examine whether conditioning the tone of MD&A on the content (e.g., profitability versus liquidity) makes a difference in forecasting future performance, I include profitability-related tone (*PROFIT_TONE*), liquidity-related tone (*LIQUIDITY_TONE*), and other tone (*OTHER_TONE*) separately in the statistical analysis. Unreported results show that both profitability-related FLS and liquidity-related FLS have implications for future earnings while other types of FLS do not. Surprisingly, in predicting future earnings, the effect of liquidity-related FLS is much bigger than that of profitability-related FLS. Next, I compare the information content of 10-K MD&As with that of 10-Qs by separating the sample into 10-Qs and 10-Ks. Untabulated results show that, while both have substantial information content about future earnings and liquidity, 10-Q MD&As tend to have more information content than 10-K MD&As. One explanation is that in annual reports management may be more likely to discuss longer-term events. Therefore, the implications for short-term earnings and liquidity are not as strong.

One potential concern is that research assistants may code a sentence as "uncertain" if they are uncertain about whether it is positive or negative, thus introducing noise into the training data. However, an unreported empirical analysis excluding the "uncertain" tone in the training data yields similar results.[17] This is consistent with the argument that the sentences in the "uncertain" category coded by the research assistants are neither random noise nor a reflection of their uncertainty about the categorization, since they appear to improve the information content of the tone measure generated by the Bayesian learning algorithm.

Another potential concern is that different companies have a different number of FLS in their MD&As. The different number of forward-looking sentences can lead to different levels of precision in the MD&A tone estimates. For instance, a firm that has five FLS may have high variance in the MD&A tone measure. To mitigate any potential concerns about this issue, I re-run the main tests by weighting the observations in the regressions using the number of forward-looking sentences in the MD&A. The intuition is that lower weighting for observations with fewer sentences in the MD&A section may mitigate any concerns about their influences on the empirical results. Unreported results indicate that greater weighting for MD&As with

idiosyncratic volatility, lower price, and smaller trading volume.) In unreported tests, I find that *WARN* is not significantly associated with firm size, the percentage of institutional ownership, the number of institutional investors, idiosyncratic stock return volatility, and the level of stock price, and is negatively related to trading volume. Hence, the empirical results are unlikely to be driven by *WARN* being correlated with these arbitrage factors.

[17] Dropping the "uncertain" category reduces the training data from 30,000 sentences to about 23,000 sentences.

more sentences leads to qualitatively similar results and does not change the inferences of the paper.

I also examine the implications of within-firm variations of tone for future profitability by including firm-fixed effects (instead of industry-fixed effects) in the regressions. Unreported results indicate that the coefficients on *TONE* are slightly smaller compared with those in the main analysis in table 6. Nonetheless, the MD&A tone is still positively and significantly associated with future earnings, even after controlling for firm and year fixed effects and other control variables, and the economic magnitudes are still substantial.

Finally, unreported results indicate that sell-side financial analysts' consensus earnings forecasts are higher when MD&A FLS are more positive, indicating that analysts' forecasts reflect the information in MD&As. However, they do not appear to fully utilize the information in the MD&A tone. Even after controlling for the latest analyst forecasts before the announcement of future earnings, the tone of the MD&A FLS still has significant predictive power for future earnings and liquidity.

5.8 FLS TONE BASED ON THE DICTIONARY APPROACH

In this section, I examine the information content of the MD&A tone based on the dictionary approach, which has been used widely in prior literature (e.g., Kothari, Li, and Short [2009] and Davis, Piger, and Sedor [2005]). Specifically, I focus on the FLS tone calculated using Diction, the General Inquirer, and the Linguistic Inquiry and Word Count (LIWC). First, I feed the 30,000 training sentences into the dictionaries and evaluate the success rate of the dictionary-based classification using human coder as the benchmark. Table 10 shows the success rates of the three dictionaries for classifying 1,000, 2,000, 5,000, 10,000, 15,000, and 30,000 randomly selected sentences of the training data. The dictionaries analyze each FLS and calculate the percentages of positive and negative words in the sentence. Unlike the learning algorithm, the dictionaries do not directly classify each sentence into a positive or negative tone category.[18] Hence, to classify each FLS sentence using the dictionary approach, I implement the following rule: a sentence is classified as being of positive tone if the percentage of positive tone words is greater than the percentage of negative

[18] General Inquirer has positive and negative words categories. LIWC has a positive emotions category that combines "positive feelings" and "optimism and energy" subcategories and a negative emptions category that combines "anxiety or fear," "anger," and "sadness or depression" subcategories. Diction does not have a positive or negative words category, but defines an optimism score as the difference between the percentage of words from the "praise," "satisfaction," and "inspiration" categories and the percentage of words from the "blame," "hardship," and "denial" categories. I therefore classify "praise," "satisfaction," and "inspiration" words as those with positive tone and the "blame," "hardship," and "denial" words as those with negative tone when using Diction.

TABLE 10

Percentage Success Rate of the Dictionaries in Classifying the Training Data

N	Band=0	Band=1	Band=3	Band=5	Band=10	Band=20	Band=30	Band=50
Diction								
1,000	49.14	46.83	46.36	49.34	48.10	45.04	41.72	41.49
2,000	48.48	48.44	47.18	49.52	47.53	46.64	44.06	40.82
5,000	48.65	47.83	48.05	48.60	48.45	46.08	43.49	41.89
10,000	48.13	48.38	49.24	48.66	48.22	45.84	43.69	41.11
15,000	48.10	48.77	48.55	47.98	48.56	46.51	43.72	40.64
30,000	48.48	48.46	48.50	48.52	48.53	45.93	43.94	40.90
General Inquirer								
1,000	33.62	32.74	35.32	37.71	38.64	38.50	39.46	37.04
2,000	33.18	33.27	33.12	37.51	39.45	40.70	40.35	40.86
5,000	32.52	32.88	33.18	36.65	40.21	39.68	39.65	40.20
10,000	31.93	32.69	34.09	36.73	39.36	39.72	40.36	39.80
15,000	33.09	33.21	33.98	36.47	39.38	39.85	40.08	39.75
30,000	32.56	32.61	33.62	36.39	39.27	39.96	39.98	39.99
LIWC								
1,000	37.69	41.33	39.60	39.37	38.47	41.31	38.41	40.26
2,000	41.12	39.91	40.57	41.04	38.91	41.91	38.17	39.32
5,000	38.99	39.82	39.42	39.20	40.86	39.86	39.48	40.33
10,000	39.95	39.80	39.76	39.63	39.21	40.58	39.93	40.02
15,000	40.24	40.35	40.31	40.06	39.77	39.90	40.38	40.45
30,000	40.26	40.28	40.29	40.10	39.95	39.98	39.99	39.99

This table shows the success rate of using the Diction 6.0, General Inquirer (GI), and Linguistic Inquiry and Word Count (LIWC) softwares to classify the 30,000 training sentences into different tones. A sentence is considered to be classified as positive tone by Diction, 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*. For Diction, the percentage of positive words is the sum of the percentages of *praise* words, *satisfaction* words, and *inspiration* words as defined by Diction and the percentage of negative words is the sum of the percentages of *blame* words, *hardship* words, and *denial* words as defined by Diction. For GI, the percentage of positive words is the percentage of positive words classified by the General Inquirer default dictionary and the percentage of negative words is the percentage of negative words classified by the General Inquirer default dictionary. For LIWC, the percentage of positive words is the percentage of positive emotion words classified by the LIWC software and the percentage of negative words is the percentage of negative emotion words classified by the LIWC software. The classification by the Diction, 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 1,000, 2,500, 5,000, 10,000, and 15,000 randomly selected sentences from the 30,000 sentences.

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 different classification results when varying *BAND* between 0 and 50. These classifications are then compared with those made by the human coder to determine the success rate.

The results in table 10 indicate that when $BAND = 0$, the General Inquirer has a success rate of 32.56% when classifying all 30,000 sentences.

This is below the result based on the informed guessing strategy reported in table 2. The LIWC has a success rate of 40.26%, which is comparable to that obtained through informed guessing. Of all three dictionaries, Diction has the highest success rate of 48.48%. As *BAND* increases (i.e., more sentences are now classified as neutral), the General Inquirer success rate improves. 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, the General Inquirer and the LIWC yield a classification rate comparable to that of the informed guessing strategy (around 40%). This is likely due to the fact that they both use very general dictionaries and thus may not work well for corporate filings. Diction has the best performance among the three dictionaries with a success rate close to 50%. This performance is likely due to the fact that it has a "corporate financial reports" norm. However, all three dictionaries have a success rate much lower than the 67% (table 2) of the Bayesian algorithm.

While interesting, this result needs to be interpreted with caution because the dictionary approach is typically used to analyze documents rather than sentences. Suppose a document consists of S sentences with each sentence i ($i = 1, \ldots, S$) having N_i words. The total number of words of the article is then $N = \sum_{i=1}^{N} N_i$. Assume that the dictionary approach finds that, of the N words, NP are positive and NN negative. The document-level tone based on the dictionary approach is then calculated based on the metric $TONE(document) = (NP - NN)/N$. Now apply the dictionary approach at the sentence level and assume that sentence i has NP_i positive words and NN_i negative words, respectively. Then the tone measure for sentence i using the dictionary approach is $(NP_i - NN_i)/N_i$.

Since $NP = \sum_{i=1}^{N} NP_i$ and $NN = \sum_{i=1}^{N} NN_i$, $TONE(document)$ can be viewed as the average sentence-level tone weighted by the length of each sentence:

$$\frac{1}{N}(NP - NN) = \frac{1}{N}\sum_{i=1}^{N}[(NP_i - NN_i)/N_i] \times N_i. \tag{3}$$

If N_i is the same across sentences, then applying the dictionary approach at the document level yields a measure that is the simple average of the tone measured at the sentence level. However, if the sentences vary dramatically in length, the two approaches could lead to different results. Therefore, a dictionary-based tone measure could perform poorly in sentence-level analysis but perform well at the document level if a document is heterogeneous in sentence length. Similarly, even though the Bayesian algorithm performs well in the sentence-level cross-validation tests, it may not work well for analyzing documents. Thus, while the superiority of the Bayesian algorithm in classifying sentences suggests that the algorithm will be superior in classifying MD&A FLS, this superiority is not clear. This concern is mitigated in

TABLE 11
Descriptive Statistics of Tone Measures Based on Dictionaries

Variable	Mean	P5	P25	Median	P75	P95	Std.
DICTION_POS	5.64	1.30	3.25	5.03	7.24	9.84	3.80
DICTION_NEG	10.05	2.50	6.14	9.36	13.07	19.70	5.68
DICTION_TONE	−4.41	−15.59	−8.16	−4.04	−0.36	5.72	7.08
GI_POSITIV	5.81	3.64	4.97	5.80	6.61	8.04	1.36
GI_NEGATIV	2.87	1.05	2.06	2.79	3.63	4.90	1.18
GI_TONE	2.94	0.00	1.74	2.89	4.07	6.04	1.85
LIWC_POSEMO	2.00	0.74	1.48	1.95	2.44	3.42	0.83
LIWC_NEGEMO	0.79	0.00	0.37	0.70	1.14	1.82	0.58
LIWC_TONE	1.21	−0.36	0.56	1.16	1.80	2.93	1.03
TONE	−0.23	−0.75	−0.41	−0.21	−0.03	0.23	0.29

This table shows the descriptive statistics for the MD&A tone based on Diction, General Inquirer (GI), and Linguistic Inquiry and Word Count (LIWC). The variables are defined as follows: *DICTION_POS* is the percentage of positive words in the MD&A FLS classified by Diction. $DICTION\_POS = praise + satisfaction + inspiration$, where *praise* is the percentage of words in the praise word list of Diction, *satisfaction* is the percentage of words in the satisfaction word list of Diction, and *inspiration* is the percentage of words in the inspiration word list of Diction. *DICTION_NEG* is the percentage of negative words in the MD&A FLS classified by Diction. $DICTION\_NEG = blame + hardship + denial$, where *blame* is the percentage of words in the blame word list of Diction, *hardship* is the percentage of words in the hardship word list of Diction, and *denial* is the percentage of words in the denial word list of Diction. *DICTION_TONE* is calculated as ($DICTION\_POS - DICTION\_NEG$). *GI_POSITIV* is the percentage of positive words of the MD&A FLS classified by the General Inquirer. *GI_NEGATIV* is the percentage of negative words of the MD&A forward-looking statements classified by the General Inquirer. *GI_TONE* is calculated as ($GI\_POSITIV - GI\_NEGATIV$). *LIWC_POSEMO* is the percentage of positive emotion words in the MD&A forward-looking statements classified by the LIWC software. *LIWC_NEGEMO* is the percentage of negative emotion words in the MD&A forward-looking statements classified by the LIWC software. *LIWC_TONE* is calculated as ($LIWC\_POSEMO - LIWC\_NEGEMO$). *TONE* is the tone of the MD&A forward-looking statements classified by the Bayesian learning algorithm.

my approach, as my analysis is based on the forward-looking sentences contained in the MD&A not the full MD&A. Because my analysis pulls these sentences out of context, context becomes less important. In an analysis of a random sample of 25 MD&As, I extract an average of 49 sentences per MD&A but only 11 of these are contiguous. This relatively small number of contiguous sentences suggests my sentence-level approach is well specified.

Next, I calculate the FLS tone for each MD&A using the dictionaries. Table 11 shows the summary statistics. Following Davis, Piger, and Sedor [2005] and Rogers, Buskirk, and Zechman [2009], I use the Diction 6.0 "corporate financial reports" norm to estimate the percentage of positive words for each MD&A as the sum of the percentages of praise, satisfaction, and inspiration words: $DICTION\_POS = praise + satisfaction + inspiration$ and the percentage of negative words as the sum of the percentages of blame, hardship, and denial words: $DICTION\_NEG = blame + hardship + denial$. The positive and negative tone based on the General Inquirer are *GI_POSITIV* and *GI_NEGATIV*, respectively, which are the percentages of positive and negative words in each MD&A as classified by the GI. Similarly, *LIWC_POSEMO* and *LIWC_NEGEMO* are the percentages of positive emotion and negative emotion words respectively based on the LIWC package. I also calculate a summary measure of MD&A tone based on each dictionary by taking the difference between the positive tone and negative tone percentages: $DICTION\_TONE = DICTION\_POS - DICTION\_NEG$, $GI\_TONE = GI\_POSITIV - GI\_NEGATIV$, and $LIWC\_TONE = LIWC\_POSEMO - LIWC\_NEGEMO$.

Table 11 shows that Diction classifies 5.64% of the words in FLS as positive and 10.05% negative. In contrast, both GI and LIWC report more positive words (5.81% and 2.00% respectively) than negative words (2.87% and 0.79% respectively).[19] As a result, the mean of *DICTION_TONE* is negative (−4.41) and those of *GI_TONE* and *LIWC_TONE* are positive (2.94 and 1.21, respectively).

Table 12 presents the Pearson correlations between *TONE* and the dictionary-based tone measures. Not surprisingly, the percentages of positive and negative words classified by the three dictionaries are positively correlated. The Pearson correlation coefficient between *DICTION_POS* and *GI_POSITIV* is 0.246, that between *DICTION_POS* and *LIWC_POSEMO* is 0.240, and that between *GI_POSITIV* and *LIWC_POSEMO* is 0.483. The Pearson correlation coefficient between *DICTION_NEG* and *GI_NEGATIV* is 0.339, that between *DICTION_NEG* and *LIWC_NEGEMO* is 0.440, and that between *GI_NEGATIV* and *LIWC_NEGEMO* is 0.627. Of the three dictionaries, GI and LIWC appear to generate tone measures that are closer to each other. The correlation between *GI_TONE* and *LIWC_TONE* is 0.565 compared to the correlation coefficient between *DICTION_TONE* and *GI_TONE* of 0.290 and that between *DICTION_TONE* and *LIWC_TONE* of 0.324. The results in table 12 also show that the machine learning algorithm and the dictionaries capture similar components of the MD&A tone. The Pearson correlation between *TONE* and *DICTION_TONE* (*GI_TONE*, *LIWC_TONE*) is 0.247 (0.299, 0.254).

Table 13 examines the information content of the different dictionary-based measures of MD&A tone for future performance by regressing next quarter's earnings on the tone measures and control variables. The control variables are the same as those in tables 6 and 7 but their coefficients are not reported. In columns (1) to (6), the tone measures based on Diction, GI, and LIWC are included separately in the regressions. The results in columns (1) and (2) show that the Diction-based tone measures are not significantly associated with future performance. The positive word measures of both GI and LIWC are negatively associated with future earnings and the percentage of negative words classified by LIWC (*LIWC_NEGEMO*) is actually positively related to future earnings. As a result, columns (4) and

[19] The percentages of positive and negative words based on the GI are comparable to those reported in Kothari, Li, and Short [2009].

TABLE 12

Correlation Matrix of the Different Tone Measures (p-Values in Parentheses)

Variables	TONE	DICTION_POS	DICTION_NEG	DICTION_TONE	GI_POSITIV	GI_NEGATIV	GI_TONE	LIWC_POSEMO	LIWC_NEGEMO	LIWC_TONE
TONE	1.000									
DICTION_POS	0.025	1.000								
	(0.000)									
DICTION_NEG	−0.292	−0.080	1.000							
	(0.000)	(0.000)								
DICTION_TONE	0.247	0.600	−0.845	1.000						
	(0.000)	(0.000)	(0.000)							
GI_POSITIV	−0.025	0.246	−0.011	0.141	1.000					
	(0.000)	(0.000)	(0.000)	(0.000)						
GI_NEGATIV	−0.500	−0.042	0.339	−0.295	−0.064	1.000				
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)					
GI_TONE	0.299	0.207	−0.223	0.290	0.774	−0.681	1.000			
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)				
LIWC_POSEMO	−0.016	0.240	−0.017	0.142	0.483	−0.053	0.388	1.000		
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)			
LIWC_NEGEMO	−0.476	−0.038	0.440	−0.373	−0.073	0.627	−0.451	−0.042	1.000	
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)		
LIWC_TONE	0.254	0.214	−0.261	0.324	0.429	−0.395	0.565	0.828	−0.596	1.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	

This table shows the Pearson correlation coefficients between *TONE* and MD&A tone based on Diction, General Inquirer (GI), and Linguistic Inquiry and Word Count (LIWC). The variables are defined as follows. *DICTION_POS* is the percentage of positive words in the MD&A forward-looking statements classified by Diction. *DICTION_POS* = *praise* + *satisfaction* + *inspiration*, where *praise* is the percentage of words in the praise word list of Diction, *satisfaction* is the percentage of words in the satisfaction word list of Diction, and *inspiration* is the percentage of words in the inspiration word list of Diction. *DICTION_NEG* is the percentage of negative words in the MD&A forward-looking statements classified by Diction. *DICTION_NEG* = *blame* + *hardship* + *denial*, where *blame* is the percentage of words in the blame word list of Diction, *hardship* is the percentage of words in the hardship word list of Diction, and *denial* is the percentage of words in the denial word list of Diction. *DICTION_TONE* is calculated as (*DICTION_POS* − *DICTION_NEG*). *GI_POSITIV* is the percentage of positive words of the MD&A forward-looking statements classified by the General Inquirer. *GI_NEGATIV* is the percentage of negative words of the MD&A forward-looking statements classified by the General Inquirer. *GI_TONE* is calculated as (*GI_POSITIV* − *GI_NEGATIV*). *LIWC_POSEMO* is the percentage of positive emotion words in the MD&A forward-looking statements classified by the LIWC software. *LIWC_NEGEMO* is the percentage of negative emotion words in the MD&A forward-looking statements classified by the LIWC software. *LIWC_TONE* is calculated as (*LIWC_POSEMO* − *LIWC_NEGEMO*). *TONE* is the tone of the MD&A forward-looking statements classified by the Bayesian learning algorithm.

TABLE 13

Information Content of Dictionary Tone Measures

COEFFICIENT	(1) EARN($t+1$)	(2) EARN($t+1$)	(3) EARN($t+1$)	(4) EARN($t+1$)	(5) EARN($t+1$)	(6) EARN($t+1$)	(7) EARN($t+1$)	(8) EARN($t+1$)
DICTION_POS	−0.000						0.000	
	(−0.01)						(0.98)	
DICTION_NEG	0.000						0.000	
	(0.71)						(0.49)	
DICTION_TONE		−0.000						−0.000
		(−0.58)						(−0.08)
GI_POSITIV			−0.001***				−0.001***	
			(−5.24)				(−5.07)	
GI_NEGATIV			−0.000				−0.000	
			(−0.30)				(−0.14)	
GI_TONE				−0.001***				−0.001***
				(−3.23)				(−3.95)
LIWC_POSEMO					−0.001**		0.000	
					(−2.17)		(0.07)	
LIWC_NEGEMO					0.001**		0.003***	
					(2.14)		(5.35)	
LIWC_TONE						−0.001***		−0.001**
						(−2.80)		(−1.99)

(*Continued*)

TABLE 13—*Continued*

COEFFICIENT	(1) EARN($t+1$)	(2) EARN($t+1$)	(3) EARN($t+1$)	(4) EARN($t+1$)	(5) EARN($t+1$)	(6) EARN($t+1$)	(7) EARN($t+1$)	(8) EARN($t+1$)
TONE							0.008***	0.007***
							(5.43)	(5.19)
Observations	85,023	85,023	85,023	85,023	85,023	85,023	85,023	85,023
R^2	0.36	0.36	0.36	0.36	0.36	0.36	0.36	0.36

This table shows the regression results of future earnings on the tone measures based on Diction, General Inquirer (GI), and Linguistic Inquiry and Word Count (LIWC) and other control variables. The dependent variables are the earnings in the next quarter (Compustat Quarterly file data item 69) scaled by the book value of assets (Compustat Quarterly file data item 44) at the end of this quarter, winsorized at −3 and 3. The independent variables are defined as follows: *DICTION_POS* is the percentage of positive words in the MD&A forward-looking statements classified by Diction. *DICTION_POS* = *praise* + *satisfaction* + *inspiration*, where *praise* is the percentage of words in the praise word list of Diction, *satisfaction* is the percentage of words in the satisfaction word list of Diction, and *inspiration* is the percentage of words in the inspiration word list of Diction. *DICTION_NEG* is the percentage of negative words in the MD&A forward-looking statements classified by Diction. *DICTION_NEG* = *blame* + *hardship* + *denial*, where *blame* is the percentage of words in the blame word list of Diction, *hardship* is the percentage of words in the hardship word list of Diction, and *denial* is the percentage of words in the denial word list of Diction. *DICTION_TONE* is calculated as (*DICTION_POS* − *DICTION_NEG*). *GI_POSITIV* is the percentage of positive words of the MD&A forward-looking statements classified by the General Inquirer. *GI_NEGATIV* is the percentage of negative words of the MD&A forward-looking statements classified by the General Inquirer. *GI_TONE* is calculated as (*GI_POSITIV* − *GI_NEGATIV*). *LIWC_POSEMO* is the percentage of positive emotion words in the MD&A forward-looking statements classified by the LIWC software. *LIWC_NEGEMO* is the percentage of negative emotion words in the MD&A forward-looking statements classified by the LIWC software. *LIWC_TONE* is calculated as (*LIWC_POSEMO* − *LIWC_NEGEMO*). *TONE* is the tone of the MD&A forward-looking statements classified by the Bayesian learning algorithm. Year- and two-digit SIC industry-fixed effects are included in the regressions but are not reported. The following control variables are also included in the regressions but the coefficients are not reported: *EARN*, *RET*, *ACC*, *SIZE*, *MTB*, *RETVOL*, *EARNVOL*, *FOG*, *NITEMS*, *NBSEG*, *NGSEG*, *FIRMAGE*, *MA*, *SEO*, *SI*, *DLW*, *Q*, *Q3*, and *Q4*, all of which are as defined in the notes to table 3A. Year-fixed effects are included in the regressions, but are not reported. *T*-statistics based on two-way clustering at both year-quarter level and firm level are reported in parentheses.

*** indicates $p < 0.01$; ** $p < 0.05$; * $p < 0.1$.

TABLE 14

Correlations Between Dictionary Tone Measures and Current Earnings and Fog

	TONE	DICTION_TONE	GI_TONE	LIWC_TONE
EARN	0.15	0.03	−0.03	−0.02
	(0.00)	(0.00)	(0.00)	(0.00)
FOG	−0.25	−0.07	0.05	0.01
	(0.00)	(0.00)	(0.00)	(0.00)

This table shows the Pearson correlation coefficients between the MD&A tone measures and current earnings (*EARN*) and MD&A Fog index (*FOG*). The variables are defined as follows: *TONE* is the tone of the MD&A forward-looking statements classified by the Bayesian learning algorithm. *DICTION_TONE* is calculated as (*DICTION_POS* − *DICTION_NEG*). *DICTION_POS* is the percentage of positive words in the MD&A forward-looking statements classified by Diction. *DICTION_POS* = *praise* + *satisfaction* + *inspiration*, where *praise* is the percentage of words in the praise word list of Diction, *satisfaction* is the percentage of words in the satisfaction word list of Diction, and *inspiration* is the percentage of words in the inspiration word list of Diction. *DICTION_NEG* is the percentage of negative words in the MD&A forward-looking statements classified by Diction. *DICTION_NEG* = *blame* + *hardship* + *denial*, where *blame* is the percentage of words in the blame word list of Diction, *hardship* is the percentage of words in the hardship word list of Diction, and *denial* is the percentage of words in the denial word list of Diction. *GI_TONE* is calculated as (*GI_POSITIV* − *GI_NEGATIV*). *GI_POSITIV* is the percentage of positive words of the MD&A forward-looking statements classified by the General Inquirer. *GI_NEGATIV* is the percentage of negative words of the MD&A forward-looking statements classified by the General Inquirer. *LIWC_TONE* is calculated as (*LIWC_POSEMO* − *LIWC_NEGEMO*). *LIWC_POSEMO* is the percentage of positive emotion words in the MD&A forward-looking statements classified by the LIWC software. *LIWC_NEGEMO* is the percentage of negative emotion words in the MD&A forward-looking statements classified by the LIWC software. *EARN* is the quarterly earnings (Compustat Quarterly file data item 69) scaled by the book value of assets (Compustat Quarterly file data item 44), winsorized at −3 and 3. *FOG* is the Fog index of the MD&A. *p*-values are in parentheses.

(6) show that *GI_TONE* and *LIWC_TONE* are both negatively associated with future performance.

Column (7) includes the positive and negative words classified by all three dictionaries, together with *TONE* based on the Bayesian algorithm, in one regression to explain future performance. Confirming the results in columns (1) to (6), the evidence in column (7) indicates that *GI_POSITIV* is significantly and negatively related to future earnings. However, the Bayesian measure *TONE* is positively and significantly associated with future earnings. In column (8), the dictionary-based measures (*DICTION_TONE*, *GI_TONE*, and *LIWC_TONE*) are included together with *TONE* to predict future earnings. Again, the results based on the dictionary-based tone measures do not lend support to the hypothesis that when the tone of corporate MD&A FLS is more positive, future performance is better. The coefficients on *DICTION_TONE*, *GI_TONE*, and *LIWC_TONE* are −0.000 ($t = -0.08$), −0.001 ($t = -3.95$), and −0.001 ($t = -1.99$), respectively. However, *TONE* is positively and significantly associated with future performance (coefficient 0.007 with a *t*-statistic of 5.19).

The empirical tests in table 13 are joint tests of the economic hypothesis and the methodology. The following factors suggest that *TONE* captures MD&A information content better than the dictionary-based measures. First, table 14 shows the Pearson correlation coefficients between the tone measures and current earnings. The correlation coefficient between *TONE* and current earnings is 0.15, and those between the three dictionary-based tone measures (*DICTION_TONE*, *GI_TONE*, and *LIWC_TONE*) and current earnings are 0.03, −0.03, and −0.02, respectively. To the extent that MD&A tone should positively relate to current performance, this evidence suggests that *TONE* is more likely to capture management tone. Second, table 14 also shows that the correlation coefficient between *TONE* and the MD&A Fog index is −0.25; the correlation coefficients between Fog and *DICTION_TONE*, *GI_TONE*, and *LIWC_TONE* are −0.07, 0.05, and 0.01, respectively. To the extent that MD&As with high Fog tend to be negative, this evidence also suggests that the Bayesian algorithm captures management tone better. Finally, unreported results indicate that the tone measures based on the dictionaries are not related to the accrual anomaly. Therefore, the accrual anomaly results in table 9 are also consistent with the validity of the Bayesian algorithm in capturing management tone.

Overall, the evidence based on the MD&A tone measured using the dictionary approach does not support the hypothesis that when MD&A is more positive, future performance is better. This evidence suggests that researchers be cautious when using dictionary-based measures in empirical tests. In particular, a lack of results based on the dictionary approach could be due to its low power.

6. *Conclusions*

This paper examines the implications of the FLS in the MD&A section of 10-Q and 10-K filings for future performance. I use a Naïve Bayesian machine learning algorithm to categorize the tone and content of FLS from more than 140,000 10-Q and 10-K filings between 1994 and 2007. I find that the tone of the FLS is a function of current performance, accruals, firm size, MTB ratio, return volatility, MD&A Fog, and firm age. The tone of the FLS is positively correlated with future performance and has explanatory power incremental to other variables. However, the informativeness of MD&As has not changed systematically over time despite continuous efforts from the SEC to strengthen MD&A disclosures. Furthermore, when managers warn in the MD&A about the future performance implications of accruals, accruals are less likely to be mispriced by investors. Finally, MD&A tone measures based on dictionary approaches do not associate positively with future performance; however, the Bayesian tone measure remains positively and significant associated with future earnings even when the dictionary-based tone measures are controlled for. This evidence suggests the dictionary approach might not work well for analyzing the tone of corporate filings.

APPENDIX A1

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., Tone)	(3) "How you say it" (e.g., Transparency/ Truthfulness)
Cost of capital	Botosan [1997] Botosan and Plumlee [2002]*	Kothari, Li, and Short [2009]	
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] Feldman et al. [2009]	Miller [2010]

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

APPENDIX A2

Sample Textual Analysis Papers

Paper	Text Analyzed	Name of Method	Method	Firm-Level Measure
This paper	Forward-looking MD&A	Naïve Bayes	Classify sentences as positive/negative/neutral/uncertain	Average of sentence tones
Davis, Piger, and Sedor [2005]	Press release	DICTION	Classify words as optimistic/pessimistic	% of words
Kothari, Li, and Short [2009]	MD&A, analyst reports, etc.	General Inquirer	Classify words as optimistic/pessimistic	% of words
Li [2008]	10-Ks and different sections	Fog/word count and LIWC	Classify at word level and averaged across sentences	% of words and average length of sentence
Henry [2008]	Earnings release	Customized dictionary	Classify words as positive/negative	% of words
Matsumoto, Pronk, and Roelofsen [2008]	Conference calls	Customized dictionary and LIWC	Classify words into forward-looking	% of words
Tetlock, Saar-Tsechansky, and Macskassy [2007]	News articles	General Inquirer	Classify words as optimistic/pessimistic	% of words
Feldman et al. [2009]	MD&A	Customized dictionary	Classify words as positive/negative	% of words
Rogers, Buskirk, and Zechman [2009]	Earnings announcements	DICTION	Classify words as optimistic/pessimistic	% of words
Henry and Leone [2010]	Earnings releases	Customized dictionary	Classify words as positive/negative	% of words

APPENDIX B
Data Preparation

I define 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 associated with legal language and boilerplate disclosures. For instance, I randomly select 5% (or 601 filings) of all 10-Ks filed in 1998 by all public issuers in the United States with a file size greater than 10K bytes. These 601 filings contain 68,878 sentences with the word "shall." 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 FLS search results. I then randomly checked 100 of the 68,878 "shall" sentences and conclude that almost all can be classified as boilerplate disclosures. Hence, I shall not include "shall" in the search for FLS.

I also exclude sentences that consist of all capital letters and sentences containing words such as "undersigned," "herein," "hereinafter," "hereof," "hereon," "hereto," "theretofore," "therein," "thereof," and "thereon," because they are almost certain to be legal boilerplate. Furthermore, I exclude all sentences that contain "expected," "anticipated," "forecasted," "projected," or "believed" when such words follow "was," "were," "had," and "had been." These typically indicate a sentence that is 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. 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. However, some sentences may be flagged as forward-looking that may not be.

I carry out the content and tone analysis at the sentence level. The disadvantage of sentence level (rather than paragraph or document level) analysis is that, in some instances, the tone and content of a sentence depend on its context. However, there are several advantages to a 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.

Before doing the Naïve Bayesian classification, I implement "stemming" and "stopwording" processes. Stemming is a 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 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 do not add 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 conduct a stopwords cleanup, I use the Lingua::EN::Stopwords modules from Perl. The stopwords 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." The reason for this modification is that one of the goals of this paper is to categorize the tone of statements, and words like "cannot" impact tone. Empirically, including or excluding these words does not affect any of the results in this paper.[20]

APPENDIX C
Classification Categories

- 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
- 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

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

APPENDIX D
Training Data Preparation

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 Ross School of Business. To make sure that the training data set is of high quality, I impose the following criteria for research assistants: (1) Native English speakers are given higher priority; (2) If a person is not a native English speaker, then she has to rank within the top 10% on the TOEFL test or GMAT/GRE verbal test; and (3) The student should have received an A– or above (or equivalent) in an intermediate financial accounting course or hold 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. This can be positive news if the reader has been expecting a material effect or neutral if the immateriality is fully anticipated. Without a more specific context, it is almost impossible to determine 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 classification judgments. In the environment liability example above, the sentence would be classified as positive because I assume that the reader has no prior information about the potential environmental liability. In later empirical analysis, the content and tone of the FLS are examined conditional on other observable variables such as contemporaneous stock returns. To the extent that any expectations of the investors are reflected in these observable variables, my "uninformative prior" research design in the training process should be effective.

REFERENCES

ANTWEILER, W., AND M. Z. FRANK. "Is All That Talk Just Noise? The Information Content of Internet Stock Message Boards." *Journal of Finance* 59 (2004): 1259–94.

BAINBRIDGE, S. M. *A Complete Guide to Sarbanes-Oxley: Understanding How Sarbanes-Oxley Affects Your Business.* Cincinnati, OH: Adams Media, 2007.

BARRON, O. E.; C. O. KILE; AND T. B. O'KEEFE. "MD&A Quality as Measured by the SEC and Analysts' Earning Forecasts." *Contemporary Accounting Research* 16 (1999): 75–109.

BARTH, M. E., AND A. P. HUTTON. "Analyst Earnings Forecast Revisions and the Pricing of Accruals." *Review of Accounting Studies* 9 (2004): 59–96.

BELLMAN, R. E. *Adaptive Control Processes.* Princeton, NJ: Princeton University Press, 1961.

BLUME, M., AND R. F. STAMBAUGH. "Biases in Computed Returns: An Application to the Size Effect." *Journal of Financial Economics* 12 (1983): 387–404.

BOTOSAN, C. A. "Disclosure Level and the Cost of Equity Capital." *The Accounting Review* 72 (1997): 323–49.

BOTOSAN, C. A., AND M. S. HARRIS. "Motivations for a Change in Disclosure Frequency and Its Consequences: An Examination of Voluntary Quarterly Segment Disclosures." *Journal of Accounting Research* 38 (2000): 329–53.

BOTOSAN, C. A., AND M. A. PLUMLEE. "A Re-examination of Disclosure Level and the Expected Cost of Equity Capital." *Journal of Accounting Research* 40 (2002): 21–40.

BROWN, S., AND J. TUCKER. "Large-sample Evidence on Firms' Year-over-Year MD&A Modifications." Working paper, University of Florida, 2010.

BRYAN, S. H. "Incremental Information Content of Required Disclosures Contained in Management Discussion and Analysis." *The Accounting Review* 72 (1997): 285–301.

BUTLER, M.; A. J. LEONE; AND M. WILLENBORG. "An Empirical Analysis of Auditor Reporting and Its Association with Abnormal Accruals." *Journal of Accounting and Economics* 37 (2004): 139–65.

CALLAHAN, C. M., AND R. E. SMITH. "Firm Performance and Management's Discussion and Analysis Disclosures: An Industry Approach." Working paper, University of Arkansas-Fayetteville, 2004.

COLLINS, D., AND P. HRIBAR. "Errors in Estimating Accruals: Implications for Empirical Research." *Journal of Accounting Research* 40 (2002): 105–34.

CORE, J. E. "A Review of the Empirical Disclosure Literature: Discussion." *Journal of Accounting and Economics* 31 (2001): 441–56.

CORE, J. E.; W. GUAY; AND D. LARCKER. "The Power of the Pen and Executive Compensation." *Journal of Financial Economics* 88 (2008): 1–25.

DAINES, R. "Does Delaware Law Improve Firm Value?" *Journal of Financial Economics* 62 (2001): 525–58.

DAS, S., AND P. K. SHROFF. "Fourth Quarter Reversals in Earnings Changes and Earnings Management." Working paper, University of Minnesota, 2002.

DAVIS, A. K.; J. M. PIGER; AND L. M. SEDOR. "Beyond the Numbers: An Analysis of Optimistic and Pessimistic Language in Earnings Press Releases." Working paper, Washington University, 2005.

DICHEV, I., AND W. TANG. "Earnings Volatility and Earnings Predictability." *Journal of Accounting and Economics* 1–2 (2009): 160–81.

DIETRICH, J. R.; S. J. KACHELMEIER; D. N. KLEINMUNTZ; AND T. J. LINSMEIER. "An Experimental Investigation of Forward-looking Non-financial Performance Disclosures." Working paper, University of Illinois at Urbana-Champaign and University of Texas at Austin, 1997.

FELDMAN, R.; S. GOVINDARAJ; J. LIVNAT; AND B. SEGAL. "Management's Tone Change, Post Earnings Announcement Drift and Accruals." Working paper, New York University, 2009.

GEISSER, S. "The Predictive Sample Reuse Method with Applications." *Journal of The American Statistical Association* 70 (1975): 320–28.

GREEN, J.; J. HAND; AND M. SOLIMAN. "Going, Going, Gone? The Demise of the Accruals Anomaly." Working paper, University of North Carolina, Chapel Hill, 2009.

GRUNDFEST, J. A., AND M. A. PERINO. "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." 1997. Web site, http://securities.stanford.edu/research/articles/19970723sen1.html.

HEALY, P. M., AND K. G. PALEPU. "Information Asymmetry, Corporate Disclosure, and the Capital Markets: A Review of the Empirical Disclosure Literature." *Journal of Accounting and Economics* 31 (2001): 405–40.

HENRY, E. "Are Investors Influenced by How Earnings Press Releases Are Written?" *Journal of Business Communication* 45 (2008): 363–407.

HENRY, E., AND A. J. LEONE. "Measuring Qualitative Information in Capital Markets Research." Working paper, University of Miami, 2010.

HIRSHLEIFER, D., AND S. TEOH. "Limited Attention, Financial Reporting and Disclosure." *Journal of Accounting and Economics* (2003): 337–86.

HUDDART, S., AND H. LOUIS. "Insider Selling, Earnings Management, and the 1990s Stock Market Bubble." Working paper, Pennsylvania State University, 2009.

HÜFNER, B. "The SEC's MD&A: Does It Meet the Informational Demands of Investors? A Conceptual Evaluation." *Schmalenbach Business Review* 59 (2007): 58–84.

HUTTON, A.; G. MILLER; AND D. SKINNER. "The Role of Supplementary Statements with Management Earnings Forecasts." *Journal of Accounting Research* 41 (2003): 867–90.

KNUTSON, P. H. *Financial Reporting in the 1990s and BEYOND*. Charlottesville, VA: Association for Investment Management and Research, 1993.

KOTHARI, S. P.; X. LI; AND J. E. SHORT. "The Effect of Disclosures by Management, Analysts, and Financial Press on the Equity Cost of Capital: A Study Using Content Analysis." *Accounting Review* 84 (2009): 1639–70.

LANG, M., AND R. LUNDHOLM. "Cross-sectional Determinants of Analyst Ratings of Corporate Disclosures." *Journal of Accounting Research* Autumn (1993): 246–71.

LANG, M., AND R. LUNDHOLM. "Corporate Disclosure Policy and Analyst Behavior." *Accounting Review* 71 (1996): 467–92.

LANG, M., AND R. LUNDHOLM. "Voluntary Disclosure and Equity Offerings: Reducing Information Asymmetry or Hyping the Stock?" *Contemporary Accounting Research* 17 (2003): 623–62.

LEUZ, C., AND R. E. VERRECCHIA. "The Economic Consequences of Increased Disclosure." *Journal of Accounting Research* 38 (2000): 91–124.

LEVINE, C. B., AND M. J. SMITH. "Critical Accounting Policy Disclosures." Working paper, Carnegie Mellon University and Boston University, 2006.

LI, F. "Annual Report Readability, Current Earnings, and Earnings Persistence." *Journal of Accounting and Economics* 45 (2008): 221–47.

LI, F. "Do Stock Market Investors Understand the Risk Sentiment of Corporate Annual Reports?" Working paper, University of Michigan, 2010.

MANNING, C., AND H. SCHÜTZE. *Foundations of Statistical Natural Language Processing*. Cambridge, MA: The MIT Press, 1999.

MARTINDALE, C. *The Clockwork Muse: The Predictability of Artistic Change*. New York: Basic Books, 1990.

MASHRUWALA, C.; S. RAJGOPAL; AND T. SHEVLIN. "Why Is the Accrual Anomaly Not Arbitraged Away? The Role of Idiosyncratic Risk and Transaction Costs." *Journal of Accounting and Economics* 42 (2006): 3–33.

MATSUMOTO, D.; M. PRONK; AND E. ROELOFSEN. "What Makes Conference Calls Useful? The Role of Analyst Participation in Conference Calls." Working paper, University of Washington, 2008.

MAYEW, W. J., AND M. VENKATACHALAM. "The Power of Voice: Managerial Affective States and Future Firm Performance." Working paper, Duke University, 2008.

MILLER, B. P. "The Effects of Reporting Complexity on Small and Large Investor Trading." *Accounting Review* (2010).

MILLER, G. S. "Earnings Performance and Discretionary Disclosure." *Journal of Accounting Research* 40 (2002): 173–204.

MILLER, G. S., AND J. D. PIOTROSKI. "Forward-Looking Earnings Statements: Determinants and Market Response." Working paper, Harvard University and University of Chicago, 2000.

MITCHELL, T. M. "The Discipline of Machine Learning." Working paper, Carnegie Mellon University CMU-ML-06-108, 2006.

MUSLU, V.; S. RADHAKRISHNAN; K. R. SUBRAMANYAM; AND D. LIM. "Causes and Consequences of Forward Looking Disclosures in the Management Discussion and Analysis (MD&A) Sections: An Exploratory Study." Working paper, University of Texas at Dallas and University of Southern California, 2008.

PATELL, J. "Corporate Forecasts of Earnings per Share and Stock Price Behavior: Empirical Tests." *Journal of Accounting Research* (1976): 246–76.

PAVA, M. L., AND M. J. EPSTEIN. "How Good Is MD&A as an Investment Tool?" *Journal of Accountancy* 175 (1993): 51–3.

PENMAN, S. "An Empirical Investigation of the Voluntary Disclosure of Corporate Earnings Forecasts." *Journal of Accounting Research* (1980): 132–60.

PENNEBAKER, J. W., AND L. D. STONE. "Contribution to a CATA Workshop Presented at the 51st Annual Meeting of the International Communication Association." (2001).

PORTER, M. F. "An Algorithm for Suffix Stripping, Program." *Program* 14 (1980): 130–37.

POWNELL, G.; C. WASLEY; AND G. WAYMIRE. "The Stock Price Effect of Alternative Types of Management Earnings Forecasts." *Accounting Review* 68 (1993): 869–912.

RICHARDSON, S.; I. TUNA; AND P. WYSOCKI. "Accounting Anomalies and Fundamental Analysis: A Review of Recent Research Advances." Working paper, London Business School, 2009.

ROGERS, J. L.; A. VAN BUSKIRK; AND S. L. C. ZECHMAN. "Disclosure Tone and Shareholder Litigation." Working paper, University of Chicago, 2009.

ROGERS, R. K., AND J. GRANT. "Content Analysis of Information Cited in Reports of Sell-Side Financial Analysts." *Journal of Financial Statement Analysis* 3 (1997): 17–30.

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

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

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

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

SKINNER, D. J. "Why Firms Voluntarily Disclose Bad News." *Journal of Accounting Research* 32 (1994): 38–60.

SLOAN, R. "Do Stock Prices Fully Reflect Information in Accruals and Cash Flows about Future Earnings?" *Accounting Review* 71 (1996): 289–315.

TAVCAR, L. R. "Make the MD&A More Readable." *CPA Journal Online* (1998). http://nysscpa.org/cpajournal/1998/0198/newsViews/0198NV7.htm (1.12.2005).

TETLOCK, P.; M. SAAR-TSECHANSKY; AND S. MACSKASSY. "More Than Words: Quantifying Language to Measure Firm's Fundamentals." *Journal of Finance* 63 (2007): 1437–67.

VERRECCHIA, R. E. "Discretionary Disclosure." *Journal of Accounting and Economics* 5 (1983): 179–94.

WATTS, R. L., AND J. L. ZIMMERMAN. *Positive Accounting Theory*. Englewood Cliffs, NJ: Prentice-Hall, 1986.

WAYMIRE, G. "Additional Evidence on the Information Content of Management Earnings Forecasts." *Journal of Accounting Research* 22 (1984): 703–19.