ESSAYS ON TEXTUAL ANALYSIS OF CORPORATE DISCLOSURES

by

PRATIK GOEL

A DISSERTATION

Presented to the Accounting Area
of the Indian School of Business
in partial fulfillment of the requirements
for the
Fellow Programme in Management

June 2016

ProQuest LLC.
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106 – 1346

Indian School of Business

Confirmation of Approval and Acceptance of Dissertation prepared by:

Pratik Goel

Title:

"Essays on Textual Analysis of Corporate Disclosures"

This dissertation has been accepted and approved in partial fulfillment of the requirements for the Fellow Programme in Management in the Accounting area by:

Sanjay Kallapur, Chairperson, Accounting
Hariom Manchiraju, Member, Accounting
Shivaram Rajgopal, Outside Member, Accounting

June 17, 2016

Original approval signatures are on file with the FPM office of the Indian School of Business.

An Abstract of the Dissertation of

Pratik Goel for the Fellow Programme in Management

in the Accounting Area to be taken June 2016

Title: ESSAYS ON TEXTUAL ANALYSIS OF CORPORATE DISCLOSURES

Approved: ……...
Sanjay Kallapur

This dissertation consists of two studies. In the first study, I examine the direction and incidence of tone management at different stages over the CEO's tenure and differentially for high-ability CEOs. From a textual analysis of the 10-K filings of US non-financial firms during 1993-2010, I compute abnormal tone as the residual term from a regression of the tone of words on its determinants, and then use it as a measure of strategic tone management – a linguistic tool often employed to influence the perceptions of investors and analysts.

I find that, on average, CEOs manage the tone downwards (more pessimistic) in the early years of their tenure than in the later years, and this relation is more pronounced for firms in high-litigation industries. Furthermore, the proportion of forward-looking sentences to total sentences in the 10-K filing is significantly lower in the early years. Since optimistic disclosures increase the likelihood of shareholder litigation (Rogers et al. 2011), these results suggest that new CEOs are concerned about litigation more in their early years, when the market is still assessing their ability. Interestingly, more able CEOs engage in upward tone management (more optimistic),

more so in their early years than the later years. Finally, while departing CEOs also manage tone downwards (on average), I find that it is only the CEOs of balance sheet-constrained firms (proxy for the limits to accruals management) who adopt a more optimistic tone in their final year. Overall, my findings suggest that it is important to study tone in addition to other tools (such as, earnings manipulation) to better understand CEOs' impression management strategy at different stages of their tenure.

In the second study, I along with my collaborators construct a weighted word-count based measure to capture the quantity of a firm's 10-K narrative R&D-related disclosures, and document a persistent & significant (statistical and economic) negative association with subsequent firm profitability (ROA) during 1993-2006. These results stand in contrast to prior literature on 10-K narrative disclosures, across disciplines, where such disclosures have been found to convey meaningful information (via positive association) about current and future firm fundamentals. Also, the observed negative bias remains significant even after accounting for alternate explanations in this context. We argue that the unique characteristics of the R&D disclosure environment make it difficult for managers to develop skilled intuitive judgments about the outcome of their firms' R&D investments, which in turn could adversely affect the accuracy and credibility of these disclosures. The assertion that the worst R&D performers are also the most biased then seems to explain the negative association. The disclosure-type and features of the environment are thus important considerations in this area.

ACKNOWLEDGMENTS

I would like to express my deep gratitude to my committee chair, Sanjay Kallapur, for his patience, support and guidance throughout the dissertation process. In addition, I would like to thank my committee members, Shiva Rajgopal, and Hariom Manchiraju, for their assistance in the completion of this dissertation. For their helpful comments, I would like to thank workshop participants at the Indian School of Business and fellow doctoral student, Ankit Jain. Finally, I express my heartfelt gratitude to my family, particularly my parents and sister, for their love and constant support.

TABLE OF CONTENTS

LIST OF TABLES

Table Page

CHAPTER 1:

CHAPTER 2:

LIST OF FIGURES

Figure Page

CHAPTER 2:

CHAPTER 1

CEO TENURE AND TONE MANAGEMENT

I. INTRODUCTION

CEOs have strong incentives to influence outsiders' perceptions at different stages of their tenure. In their first year or two, CEOs face the critical challenge of managing the stakeholders' expectations and building confidence among all parties by reaching their initial performance targets (Vancil 1987). Also, the market is likely to be more uncertain about CEOs' ability in the early years of their tenure (Gibbons and Murphy 1992; Ali and Zhang 2015),[1] and the way the market perceives their ability is valuable to CEOs since in the long-term it guarantees several benefits to them, such as higher compensation, reappointments etc. (Fama 1980; Hermalin and Weisbach 1998). Previous studies in accounting and finance have focused on CEOs' attempts to attain these goals through quantitative management using discretionary accounting choices (such as, accruals and write-offs). For instance, Elliott and Shaw (1988) find that new executives make large discretionary write-offs in their first year and seek to blame prior managers for the resulting poor performance. Further including other years of the CEOs' tenure, Ali and Zhang (2015) show that CEOs attempt to favorably influence the market's perception of their ability through

[1] It is rare for CEOs to leave a firm and join another (Ali and Zhang 2015): CEO exits for taking a CEO position in another firm are 2.2% (Gibbons and Murphy 1992) and 3.2% (Brickley et al. 1999). Thus, most CEOs are unlikely to have any past records of their performance as CEO. Also, even if CEOs were promoted from within their firms, the market is still likely to be uncertain about their ability, since the portfolio of skills required to be successful in that role is different from that required for a lower level position (Gibbons and Murphy 1992; Ali and Zhang 2015).

greater earnings overstatement in the early years than in the later years of their tenure.

Another tool that CEOs may employ in this context is the linguistic *tone* of firm narrative disclosures (such as particular sections in the 10-K and 10-Q filings, earnings press releases, and conference call transcripts), as the effect of tone has been found to be incremental to that of the quantitative tools, such as accruals, in managing perceptions (Huang et al. 2014). The relevance of studying the tone of narrative disclosures[2] in this regard stems from the following two reasons. First, tone has been found to significantly influence the perception of investors and analysts (Lang and Lundholm 2000; Henry 2008; Loughran and McDonald 2011; Huang et al. 2014). Second, unlike financial statements, narrative disclosures are marked by the absence of any concrete regulation regarding their exact format or content,[3] thereby making them amenable to CEO (or manager) discretion with regards to both the extent of detail provided as well as the language (or rhetoric) used (Davis and Tama-Sweet 2012). When CEOs use this discretion to employ a disclosure tone that is unequal to their firms' underlying quantitative fundamentals, it is referred to as *tone management*.[4] Using the qualitative text in earnings announcements, Huang et al. (2014) show that managers employ tone management – an abnormal positive or

[2] Tone of a disclosure is typically calculated as the difference in the frequency of positive and negative words occurring in that particular (narrative) disclosure.

[3] Narrative disclosures constitute an "unregulated window" because there are minimal explicit rules regarding the content and format of these disclosures, as words are much more elastic than numbers and thereby more difficult to regulate (Lang and Lundholm 1996; Merkl-Davies and Brennan 2007; Henry 2008; Rogers et al. 2011; Huang et al. 2014).

[4] Huang et al. (2014) decompose tone into a normal component to reflect a neutral description of current performance, and an abnormal (or discretionary) component that proxies for the strategic choice of tone either to inform or misinform investors. Tone management refers to this abnormal tone component.

overoptimistic tone - to mislead investors and other financial statement users about future performance, especially in cases where they have strong incentives to bias investor perceptions upwards (or downwards). But, optimistic language is significantly associated with greater litigation risk, and such language has been the subject of shareholder lawsuits in the past (Rogers et al. (2011)).[5] Therefore, it is possible for CEOs to either portray themselves in a favorable light by overhyping firm performance through a greater use of optimistic words in the firm narrative disclosures, or to lower the outsiders' expectations by depressing firm performance using more pessimistic words.

In this study, I examine the direction and incidence of tone management at different stages over the CEO's tenure and differentially for high-ability CEOs. I argue that the CEOs' incentives to manage tone (and thus influence perceptions and expectations) are different in the early years (the first three years) and the final year of their tenure as compared to the other years. From a textual analysis of the 10-K filings of all US non-financial firms during the sample period 1993-2010, I compute an appropriate term-weighted measure of the tone of words and then regress the same on its determinants to obtain abnormal tone (the residual term), which reflects the CEOs' tone management strategy. My focus on 10-K filings is motivated from prior research suggesting that these filings are an important source of information (Previts et al. 1994; Leder 2003; Brown and

[5] For example, in *Hack, et al. v. Metris Companies, Inc., et al.* (class action complaint, 2002, p. 43), the plaintiffs alleged that Metris's CEO "offered a hopelessly positive spin on his Company's capital plan." In the *VoiceFlash Networks, Inc. Securities Litigation* (2005), it was alleged that "The news that VoiceFlash would liquidate came on the heels of numerous highly optimistic announcements by the defendants touting the success of the Company's business condition and financial performance." (Rogers et al. 2011)

Tucker 2010; Lehavy et al. 2011; Merkley 2014), and that the narrative disclosures in them contain information incremental to that in other firm disclosures (Davis and Tama-Sweet 2012).[6]

Regarding CEOs' tone management strategy in the early years of their tenure, I find evidence of greater downward tone management (negative abnormal tone or pessimistic tone) as compared to the later years. That is, CEOs attempt to lower outsiders' expectations through greater pessimistic words in their early years. This can partly be explained by the significant association of optimistic language with greater litigation risk, which leads to CEOs' heightened litigation concerns as the market is still assessing their ability in the early years. CEOs become more concerned about the potentially enormous losses of shareholders' wealth and personal credibility that could possibly result from a litigation filing during this time (Graham et al. 2005, Bhattacharya et al. 2007; Gande and Lewis 2009 etc.).[7] It is important to note that although tone is not the sole determinant of litigation risk, it is both associated with litigation risk and under the discretion of management (CEO), thus offering CEOs a straightforward means to reduce their litigation risk exposure (Rogers et al. 2011; Li 2010).

[6] While other types of firm narratives (such as, conference call transcripts and earnings press releases) could also have been chosen, I believe that the narrative disclosures contained in the 10-K filing comprehensively cover other important firm matters besides performance, such as those pertaining to R&D, competition, liquidity and debt obligations etc. The choice of the tone and content of these specific disclosures also forms part of the CEOs' overall disclosure strategy, and the 10-K tone would incorporate this. Earnings announcement-related conference calls and earnings press releases, on the other hand, focus only on the firms' financial performance (earnings, sales, gross profit etc.).

[7] I assume that the litigation risk in the early years is the same as in the later years, but CEOs are more concerned about it in their early years.

I test for the above explanation by examining the tone management strategy of CEOs in high-litigation industries, and find that the result of greater downward tone management in CEOs' early years is more pronounced in this case. I also find that CEOs disclose a significantly lower proportion of forward-looking information in the 10-K in their early years, which seeks to corroborate the litigation argument because a threat of litigation reduces managers' incentives to supply forward-looking disclosures (Weetman and Collins 1996). Furthermore, I show that the level of tone management is highest in the CEOs' first year and then declines (almost linearly) over time as their tenure progresses, which implies that CEOs are concerned about litigation to a much greater extent in their early years as compared to the later years.

Next, to account for a possible variation in the perceived severity of the litigation concerns depending on the ability of the CEOs, I examine the tone management strategy of high-ability CEOs. I use a novel measure of managerial ability from Demerjian et al. (2012), based on managers' efficiency in generating revenues, which reflects managerial talent that is distinct from the firm and outperforms alternate proxies of CEO ability. I find that as their level of ability goes up, CEOs engage in upward tone management (adopt an overoptimistic disclosure tone), and this positive association is stronger in the CEOs' early years. In other words, more able CEOs engage in greater upward tone management in the early years than in the later years of their tenure. This contrary finding can be explained by the high-ability CEOs' efforts in ensuring the subsequent good performance of

their firms, making any upward tone management in the current period much harder to detect and thus substantially mitigating their litigation concerns.

Finally, in the case of CEOs' final year, I again find evidence of greater downward tone management as compared to the previous years. To explain this result, I first point out that the most important reason for CEO departure is retirement,[8] as CEO exits for taking a CEO position in another firm are very rare (Gibbons and Murphy 1992; Brickley et al. 1999). Commensurate with retiring CEOs' incentives to safeguard their professional reputation in the hope of securing post-retirement board positions within their existing firm or at other firms (Brickley et al. 1999), these CEOs may be reluctant to adopt an optimistic disclosure tone as it could harm their reputation if the firm subsequently underperforms. Interestingly, however, I find that departing CEOs of balance-sheet constrained firms (a proxy for the limits to accruals management) instead engage in upward tone management in their final year.

These results are robust to the use of alternate definitions for the CEO variables, alternate control proxies, regression specifications with only firm (or industry) and year fixed effects, and against outliers. With the intention of expanding the scope of this study to cover more dimensions of qualitative management, I additionally investigate whether CEOs also simultaneously manage the readability (or complexity) of the 10-K over their tenure. Using three different proxies of text readability from prior literature (Li 2008; Loughran and

[8] Across my sample, 2,490 out of 4,797 (51.91 per cent) observations for which the reason of CEO departure is non-missing in *ExecuComp* correspond to these retiring CEOs.

McDonald 2014), however, I obtain mixed evidence for the same. Specifically, when the *Fog* index (10-K file size) is used, I find that CEOs decrease (increase) the complexity of the 10-K document in the early years of their tenure as compared to the later years.

My study makes the following contributions. First, it contributes to the literature on CEO behavior by highlighting tone management as a strategy employed by CEOs to manage perceptions at critical stages of their tenure, and thus offers new insights. Prior studies have examined CEO impression management around turnovers or at different stages of the tenure by concentrating only on CEOs' discretionary accounting choices (DeAngelo 1988; Elliott and Shaw 1988; Pourciau 1993; Wells 2002; Ali and Zhang 2015), abnormal stock returns (Brickley et al. 1999), or the presentational format of graphs in the financial reports (Godfrey et al. 2003; Mather and Ramsay 2007). The directionally opposite results for tone documented in the current study suggest that CEOs' tone management strategy cannot be inferred from their earnings management strategy adopted in the early years, due to the difference in the underlying determinants. Therefore, to better understand the CEOs' impression management strategy at key stages of their tenure, it is important to study tone in addition to earnings manipulation.

Second, it contributes to the literature on tone management (and, more broadly, impression management) by showing that contextual factors, such as ability, have a significant influence on the CEOs' tone management strategy at key stages of their tenure. The limited prior evidence on the role of ability for CEO

impression management (via earnings management) in the early years of tenure suggests that CEOs' ability does not matter. Specifically, Ali and Zhang (2015) show that all CEOs, regardless of ability, manipulate earnings upwards in their early years. In contrast, I find that while all CEOs, on average, bias tone downwards, high-ability CEOs instead bias it upwards in the early years.

Finally, it makes a contribution to the extant literature on narrative disclosures in the context of litigation risk (Nelson and Pritchard 2007; Rogers et al. 2011). While the evidence in Rogers et al. (2011) suggests that managers can reduce their litigation risk exposure by dampening the disclosure tone, the current study is the first to document this empirically by showing that CEOs belonging to firms in high-litigation industries indeed dampen the 10-K disclosure tone. Li's comprehensive review paper (2010) on textual disclosures mentions that "linking textual characteristics of disclosures to litigation or litigation risk a company faces could shed light on how firms behave in a litigious environment". I offer new insights on the same in the CEO tenure setting, where I show that greater litigation concerns prompt CEOs to not only adopt an over-pessimistic tone but also disclose a much lower proportion of forward-looking information in the 10-K in the early years of their tenure.

II. SAMPLE AND DESCRIPTIVE STATISTICS

Sample and Data

For the years 1993-2010, I obtain the CEO sample and related data from ExecuComp, accounting and segments data from Compustat, stock return data from CRSP, and analyst data from IBES. Firms in the finance, insurance and real estate sectors (SIC codes between 6000 and 6999) were removed.

To compute the tone variables, I begin with a sample comprising of 46,756 firm-year observations and consisting of firms in the intersection of the Compustat and SEC EDGAR databases, matched using the Central Index Key (CIK). Any unmatched observations were then matched using the IRS tax identification number (Nini et al. 2012). All 10-Ks were downloaded from the SEC EDGAR database. Following Loughran and McDonald (2011), I removed 10-Ks that contain less than 2,000 words, and only included one filing per firm per year by removing the filings that were filed within 180 days from a prior filing. In case there were multiple 10-Ks filed within a year, I considered only the first filing. The algorithm to parse the 10-K documents is outlined in Appendix B. Upon merging with the CEO data from ExecuComp, my final sample comprises of 18,576 firm-year observations, representing 4,547 CEOs and 2,557 unique firms.

Variable Measurements

Tone of the 10-K Document

I define tone *(10K TONE)* as the difference in the frequency of positive minus negative words occurring in the firms' 10-K filing divided by the total number of words in that document.

First, following the bulk of prior studies, I employ the *bag of words* approach to represent the 10-K text numerically. Under this approach, each document is represented by the words it contains, ignoring any punctuation and ordering. Every word is identified and counted the number of times it appears in the document. Next, Python scripts were used to search for positive and negative words in the entire document for each 10-K filing.[9] To identify positive and negative words, I refer to the financial sentiment dictionary developed by Loughran and McDonald (2011). This domain-specific dictionary, which comprises of 354 positive and 2,329 negative words, was created for analyzing financial communications and is now widely used by business researchers to assess the linguistic tone of a document (Bodnaruk et al. 2013; Huang et al. 2014; Kearney and Liu 2014; Law and Mills 2015; among others). Also, as a further refinement, a Python algorithm was used to reduce each word to its 'stem', so that different forms of the same word are considered as a single word

[9] Some prior studies restrict attention to the tone of only the MD&A section in the 10-K document. However, Loughran and McDonald (2011) show empirically that the MD&A section does not contain richer tonal content. Therefore, I compute tone of the entire 10-K document and not for any one specific section. Other studies that examine 10-K disclosure tone include Li (2008), Lehavy et al. (2011), Loughran and McDonald (2011), Bodnaruk et al. (2013), and Campbell et al. (2014).

[for example, the (positive) words *"improve"*, *"improved"*, *"improving"*, and *"improvement"* were stemmed to *"improve"*].

Finally, and importantly, I adopt a term-weighting scheme from prior literature to appropriately weight my *10K TONE* measure. Since a weighting scheme can help attenuate the impact of high frequency words thereby allowing less frequently used words to have a greater impact (Jurafsky and Martin 2009; Loughran and McDonald 2011), prior studies (such as, Loughran and McDonald (2011)) have argued for and shown empirically that a weighted word-count-based measure of tone is superior to an unweighted one. The following term-weighting scheme was used:[10]

$$w_{i,j,t} = \begin{cases} \frac{(1 + \log(tf_{i,j,t}))}{(1 + \log(a_{j,t}))} * \log(\frac{N_t}{df_{i,t}}), & \text{if } tf_{i,j} \geq 1 \\ 0, & \text{Otherwise} \end{cases}$$

, where $a_{j,t}$ denotes the average word count of documents in year t, $tf_{i,j,t}$ is the raw count of the i^{th} word in the j^{th} document in year t, $df_{i,t}$ represents the number of documents containing at least one occurrence of the i^{th} word in year t, and finally N_t is the total number of 10-K documents in year t.

Abnormal Tone (The Expected Tone Model)

The mere occurrence of positive (or negative) words in firm narratives doesn't necessarily indicate tone management, since their presence has been found to

[10] The weighting scheme was modified to additionally adjust for document length [similar to Loughran and McDonald (2011)] and account for the variation in length over time, since the 10-K has become significantly lengthier over time and it is more likely for a word appearing in 1993 to have a different impact than a word appearing in 2010 (my sample period is from 1993-2010).

correlate positively with firm performance (Henry 2008). Only when firms adopt a tone in their narrative disclosures that is incongruent with the underlying quantitative fundamentals does it constitute as tone management. Thus, tone has a normal component (reflecting a neutral tone commensurate with current available information about firm fundamentals), and an abnormal component (discretionary component capturing managers' strategic choice of tone). In research, it is vital to distinguish between these abnormal and normal components of tone to avoid making any erroneous conclusions (Rogers et al. 2011).

To compute abnormal tone, I first run annual cross-sectional regressions of tone (*10K TONE*) on its determinants as suggested in Li (2010a) and first applied in the tone management context by Huang et al. (2014). The determinants include measures for current firm fundamentals (*EARN*, *RET*, and SIZE), growth opportunities (*BTM*), firm's business and operating risk environment (*RET VOL* and *EARN VOL*), operating complexity (*BUS SEG* and *GEOG SEG*), and life-cycle stage of the firm (*AGE*). All variable definitions are outlined in Appendix A. The expected tone model is thus stated as follows:[11]

$$10K\ TONE_{j,t} = \alpha_0 + \alpha_1 EARN_{j,t} + \alpha_2 RET_{j,t} + \alpha_3 SIZE_{j,t} + \alpha_4 BTM_{j,t} + \alpha_5 RET\ VOL_{j,t} + \alpha_6 EARN\ VOL_{j,t} + \alpha_7 ln\ BUS\ SEG_{j,t} + \alpha_8 ln\ GEOG\ SEG_{j,t} + \alpha_9 AGE_{j,t} + \varepsilon_{j,t} \quad (1)$$

[11] Although Huang et al. (2014) additionally include performance benchmarks and future performance expectation variables as tone determinants based on a referee suggestion, they mention that the analyses in the paper is robust to their exclusion. I do not include these variables in model (1) especially since performance benchmarks and expectation variables seem more appropriate in the case of narrative disclosure that are focused on the financial performance of the firm, which is true for earnings press releases as used in Huang et al. (2014). The focus of the 10-K, on the other hand, is much broader and encompasses several other important firm matters besides performance. Thus, inclusion of current earnings performance suffices in my context.

The 10-K abnormal tone (*AB_TONE*) is the residual term from the estimation of model (1). Following Huang et al. (2014), I exclude variables related to managerial discretionary behavior (such as, seasoned equity offering, special items, and mergers and acquisition) from the above model, so that the residual term as a measure of abnormal tone can reflect these strategic incentives.

Early Years and Final Year of the CEOs' Tenure

To measure the CEOs' early years and final year of tenure, I create two separate indicator variables, following Ali and Zhang (2015). Specifically, *CEO Early Years* is an indicator variable equal to one for firm-years corresponding to the first three years of the CEOs' tenure, and zero otherwise. That is, the CEO turnover year (or CEO change year) and the two years following it constitute early years. Although the three years cutoff seems arbitrary, I later justify it by estimating tone management for each of the first five years of the CEOs' tenure[12] (see Table 4). *CEO Final Year* is an indicator variable equal to one for the year immediately preceding the CEO turnover year, and zero otherwise.

Summary Statistics

Each year, all continuous variables are winsorized at the 1 per cent and 99 per cent levels, except firm age which is winsorized at 37 years. As observed in Panel A of Table 1, the mean (median) *10K TONE* is -37.11 (-33.46), indicating

[12] As a further robustness check, I also use two years and four years as cut-offs to measure the CEOs' early years of tenure and obtain similar results (see columns 1 and 2 of Table 9).

that disclosure tone in the 10-K is generally pessimistic.[13] This is consistent with Loughran and McDonald (2011), who report higher mean negative words than positive words in the 10-K filings. Panel B reports the estimation results of model (1). I find that *10K TONE* is more positive when the firm has higher returns, is younger, growing, has less volatile stock returns, and has a greater number of business and geographical segments. *AB_TONE*, which is the residual of model (1), has a mean of zero, as expected (Panel C).[14]

Panel A of Table 2 presents the summary statistics for the rest of the variables (definitions in Appendix A). In my final sample, CEOs have an average tenure of about 7.7 years (median 5.25), which is comparable to statistics reported in prior studies.[15] 31 per cent of the firm-year observations correspond to the early years of CEOs' tenure, while 18 percent of the firm-year observations in the reduced sample[16] correspond to the final year of CEOs' tenure. Also, 33 per cent of the firm-year observations correspond to firms belonging to high-litigation industries. In terms of the textual characteristics, the 10-K filing for an average firm in my

[13] In unreported tests, I find that about 71 per cent of the variation in *10K TONE* is cross-sectional, while 41 per cent is in the time-series. Note that the two components do not sum to 1 for unbalanced data.

[14] To validate my tone measure and ensure that the textual methodology is classifying the words properly, I manually read the ten most over-optimistic (positive abnormal tone) and ten most over-pessimistic (negative abnormal tone) 10-K filings as classified by the *AB_TONE* measure, and noticed significant differences between them. The respective tonal classifications were largely found to be correct. Here is an example each of optimistic and pessimistic sentences from these filings (note, however, that my tone measures are word-count-based):

- *"Looking forward, while the economy remains uncertain, we continue to see numerous market opportunities and we believe Agilent is in a strong position to capitalize on them."* - Agilent Technologies, Inc. (year ended Oct. 31, 2010)
- *"While the reorganization is intended to have long-term benefits for the Company, in the shorter term the Company may experience disruption in its operations and loss of sales and market share as a result of the implementation of the reorganization."* – Sun-Times Media Group, Inc. (year ended Dec. 31, 2006)

[15] For example, the mean of CEO tenure is 7 in Berger et al. (1997), 6.6 in Coles et al. (2008), and 8.31 in Ryan et al. (2009).

[16] The final year of CEOs' tenure was identified using the "LEFTOFC" variable in the *Execucomp* database, which had several missing observations that were dropped, thus giving a reduced sample of 11, 694 firm-year observations (as opposed to 18,576).

final sample consisted of about 1,833 total sentences, of which nearly 217 were characterized as forward-looking. The Spearman correlations are reported in Panel B. Interestingly, *AB_TONE* is (significantly) negatively correlated with both *CEO Early Years* and *CEO Final Year* variables. The implications of this finding are discussed along with the regression results in Sections III and V.

III. EARLY YEARS OF CEOS' TENURE AND TONE MANAGEMENT

Model

To examine the CEOs' tone management strategy in the early years of their tenure, I propose the following model of the determinants of strategic tone management; stated in its basic form as:

$$AB\_TONE_{j,t} = \alpha_0 + \alpha_1 CEO\ Early\ Years_{j,t} + \alpha_2 HighLitigation_{j,t} + \alpha_3 NOA_{j,t} + \alpha_4 CEO\ Optimism_{j,t} + \alpha_5 ACC_{j,t} + \varepsilon_{j,t} \quad (2)$$

All variable definitions are outlined in Appendix A. Model (2) also includes year fixed effects.[17] Standard errors were clustered by firm to control for cross-sectional correlation between the residuals. The motivation for each control variable (i.e., determinants of abnormal tone deduced from prior studies) in this model is discussed below.

High-litigation firms. Li (2010) suggests that managers are reluctant to be optimistic in 10-K filings because of litigation concerns. Furthermore, the 10-K is more subject to "evidentiary use" in litigation filings, since it is not only audited but its form and format are also governed by accounting regulations to a much larger extent (Huang et al. 2014). Thus, it is important to control for the litigation environment of the firm as it could directly affect the CEOs' incentives to

[17] Firm or industry fixed effects cannot be included in the model due to high correlation with the *HighLitigation* dummy. However, I run a regression specification separately with only firm and year fixed effects as controls to show the robustness of my main result to the inclusion of firm effects.

manage tone in the 10-K filing. In fact, given the findings in prior studies (including, Rogers et al. (2011)), I expect a negative association between *HighLitigation* and *AB_TONE*.

Net operating assets. Asset-scaled net operating assets (NOA) at the beginning of the fiscal year are a proxy for the limits to accruals management (Barton and Simko 2002; Hirshleifer et al. 2004; Das et al. 2011). Huang et al. (2014) observe that when managers are constrained in manipulating accruals, they are more likely to resort to (upward) tone management. Thus, I expect a positive association between *NOA* and *AB_TONE*.[18]

Options-based measure of CEO optimism. Prior studies have argued that the tone of corporate disclosures also has a manager-specific component. Upon examining the tone of conference calls, Davis et al. (2015) conclude that tone is significantly influenced by a manager-specific tendency to be optimistic or pessimistic.[19] Thus, in my context, I should also expect a positive correlation between *CEO Optimism* and *AB_TONE*. Furthermore, Campbell et al. (2011) documents the significant effect of CEO optimism on forced turnover, thereby implying that CEO optimism is also likely associated with both the *CEO Early Years* and *CEO Final Year* variables. To avoid biasing my results, therefore, it is especially important to control for the CEOs' level of optimism.

[18] In fact, as observed in Panel B of Table 2, I obtain a correlation coefficient of 0.146 between *AB_TONE* and *NOA.*

[19] Although Davis et al. (2015) also examines the effect of observable manager-specific characteristics (such as, gender, age, early career experiences etc.) associated with optimism on the tone of conference calls, I believe that in my context, the *CEO Optimism* variable suffices. This is because, unlike in 10-K filings, a conference call generally involves a relatively more direct (or personalized) interaction with the company's CEO, thereby rendering such individual attributes important considerations for tone. In preliminary analyses, I ran a regression specification with the CEO's age and gender as additional controls but found them to be statistically insignificant.

Accruals. The negative association between accruals and future firm performance is well-documented. Managers manipulate the accrual component of earnings (Sloan 1996). Alternatively, accruals could proxy for the economic conditions faced by the firm (example, distress). In either of the two cases, Li (2010a) argues that the managers are likely to understand (at least partially) the implications of accruals for future performance. Given this, if managers have incentives to mislead (inform) investors, I expect a positive (negative) relation between *ACC* and *AB_TONE*.

Empirical Results

Table 3 reports the regression results from the estimation of model (2). The standard errors in these regressions as well as all the other regressions in the paper are clustered by firm. As observed from the results stated in column (1) for the basic model (with only firm and year fixed effects), the coefficient on *CEO Early Years* is negative and significant, -1.427 (t-statistic = -3.14), suggesting that abnormal tone is significantly more negative in the early years than in the later years of CEOs' tenure. Even in the full model with controls (column 2), the coefficient on *CEO Early Years* is negative and significant, -2.112 (t-statistic = -3.51). Hence, CEOs engage in greater downward tone management (or, adopt an over-pessimistic tone) in their early years as compared to the later years, and thus attempt to lower the outsiders' expectations.[20]

[20] It could be argued that this negative association is the result of new CEOs making large write-offs or write-downs (which then results in a negative tone). Elliott and Shaw (1988) show that 39 per cent of the firms in their sample

All of the control variables in model (2) are also significant. The coefficient on *HighLitigation* is significantly negative, consistent with high litigation risk environment of the firm discouraging CEOs from adopting an (overly) optimistic tone for narrative disclosures in their firms' 10-K filing (Li 2010; Rogers et al. 2011). The coefficient on *NOA* is positive, consistent with the observation in Huang et al. (2014) that balance sheet constrained firms resort to upward tone management. Also, the significantly positive coefficient on *CEO Optimism* suggests that optimistic CEOs in turn tend to adopt a more optimistic disclosure tone. Finally, the relation between *ACC* and *AB_TONE* is also positive, suggesting that CEOs understand the implications of accruals for future performance but have incentives to mislead (and not inform) investors.

Proposed Explanation: CEOs' Heightened Litigation Concerns in the Early Years

At first glance, the above result seems counterintuitive. Given that the market is likely to be more uncertain about CEOs' ability in the early years of their tenure (Gibbons and Murphy 1992; Ali and Zhang 2015), one would have expected CEOs to bias the tone of the 10-K document upwards (that is, adopt an over-optimistic tone) in their early years, either by overhyping performance or/and masking poor future prospects, as compared to that in the later years.

---

experience management changes (changes in the CEO etc.) during the year of the write-off. Therefore, to check for this possible explanation, I re-compute abnormal tone after including "Special Items" (includes write-downs or write-offs of receivables, intangible assets, etc.) as an additional control in model (1), and then rerun model (2). I again obtain a significant negative association between abnormal tone and the early years variable, suggesting that the evidence of greater over-pessimistic tone in CEOs' early tenure years is not just the result of write-offs and/or write-downs.

However, it should be noted that tone to some extent is associated with both litigation and credibility risk. Specifically, optimistic disclosures increase the likelihood of shareholder litigation (Rogers et al. 2011).[21] Investors may allege that their prior expectations regarding firm value were too high, and that those overoptimistic expectations were based on the firm's disclosures. With regards to credibility, if stakeholders recognize that they've been misled by the firm's disclosures, the firm's (and consequently, the CEO's) image and reputation can suffer severe damages (Rahman 2012).

Although CEOs are likely concerned about their litigation risk exposure throughout their tenures, these concerns may be especially pressing in the early years of their tenure when the market is still assessing their ability. This is because litigation filing announcements could trigger enormous losses of wealth for shareholders of the sued companies, as prior studies have documented (Wier 1983; Bhagat et al. 1998; Bhattacharya et al. 2007; Gande and Lewis 2009).[22] Furthermore, as gleaned from the survey evidence in Graham et al. (2005), these senior executives are also concerned about the press coverage associated with even the potentially frivolous lawsuits, as it may still harm their reputation. Due to these adverse consequences of a potential lawsuit or litigation filing, the CEOs risk being labeled as "low ability" managers in the early years of their tenure and their whole career could suffer as a result. It is, thus,

[21] Shareholder litigation, under Rule 10b-5 of the Securities Exchange Act of 1934, limits managers' opportunistic disclosures. Rule 10b-5 makes it unlawful to ''make any untrue statement of a material fact or to omit to state a material fact necessary in order to make the statements made . . . not misleading." Under this rule, investors can initiate legal action after being harmed by a defendant's misrepresentations (Rogers et al. 2011).

[22] A decline in market valuation occurs on account of significant legal costs that need to be incurred (Romano 1991; Haslem 2005), the diversion of employee attention and time (Johnson et al. 2000; Black et al. 2006), and the risk of the financial liabilities in the event of a loss (Hertzel and Smith 1993).

imperative for CEOs to try and minimize their litigation risk exposure using tools at their disposal, especially in the early years. Prior studies have shown that managers attempt to reduce expected (personal) litigation costs by altering their corporate disclosure choices and accounting decisions (Li 2010; Nelson and Pritchard 2007; Levy et al. 2015).

A straightforward means of reducing litigation risk is the downward adjustment of disclosure tone (Rogers et al. 2011; Li 2010). Although tone is not the sole determinant of litigation risk, but importantly, it is both associated with litigation risk and under the discretion of management (CEO). Based on the main findings in Rogers et al. (2011), I posit that CEOs will mitigate their litigation concerns in the early years by dampening the tone of their disclosures much more than what is justified by current and future firm fundamentals. In fact, using an exogenous shock to CFO personal litigation risk, Levy et al. (2015) find that the negative tone of earnings announcement conference calls significantly increases for CFOs who are most affected by the shock. Specifically in the context of 10-Ks, the pessimistic tone of narrative disclosures has been found to be positively associated with subsequent (positive) earnings surprises, thereby hinting at the possibility of managers attempting to lower expectations by using a higher proportion of negative words (Loughran and McDonald 2011).

Furthermore, the adoption of a negative abnormal tone in the CEOs' early years is consistent with the *information* argument that also stems from their greater litigation concerns around this time. That is, the threat of litigation induces managers to disclose more bad news information [Skinner (1994, 1997); Francis

et al. 1994].[23] Survey evidence in Graham et al. (2005) also supports this contention: when asked detailed questions about the speed of information release, 76.8 per cent of the respondents (senior executives) said that they reveal bad news faster to decrease the possibility of lawsuits resulting from a failure to disclose any unfavorable news to the market in a timely manner.

To empirically test the above explanation, I add the interaction between *CEO Early Years* and *HighLitigation* in model (2).[24] The regression results are presented in column (3) of Table 3. Again, it can be observed that the coefficient on *CEO Early Years* is negative and significant, but also more importantly, the interaction of *CEO Early Years* with *HighLitigation* is negative and significant (-2.180, with t-statistic = -1.72). Thus, the earlier result of greater downward tone management in the CEOs' early years is even stronger for firms in high-litigation industries. These findings support the notion that CEOs have incentive to adopt an overly pessimistic tone in the early years of their tenure, presumably to lower the expectations of the market or to disclose more bad news information - both of which are commensurate with their greater litigation concerns at this time when the market is still assessing their ability. However, as the coefficient on *CEO Early Years* is negatively significant even after adding the interaction term, CEOs' litigation concerns can only serve as a partial (yet significant) explanation for the adoption of an over-pessimistic tone by CEOs in their early years. The

[23] Kothari et al. (2009), on the other hand, find that management, on average, delays the release of bad news to investors. However, they also find that managers' tendency to withhold bad news is attenuated for firms with high litigation risk. Similarly, in the current paper, I argue that the incentives to withhold bad news should be weaker in the early years due to greater CEO litigation concerns around this time.

[24] I also include an interaction between the *CEO Early Years* and *NOA* variables to additionally examine the significance of *NOA* as a determinant of strategic tone management in the CEOs' early years.

existence of other explanations cannot be precluded from this empirical analysis.

Furthermore, to provide justification for using the first three years of the CEOs' tenure as cutoff for defining *CEO Early Years* and to empirically test the notion that CEOs are concerned about litigation to a much greater extent in their early years as compared to the later years of their tenure, I estimate model (2) after replacing *CEO Early Years* with indicator variables for each of the first five years of CEOs' tenure, namely, *CEO First Year*, *CEO Second Year*, *CEO Third Year*, *CEO Fourth Year*, and *CEO Fifth Year*. *CEO First Year* equals one if the observation corresponds to the first year of CEOs' tenure and zero otherwise, and so on. Column (2) of Table 4 presents the regression results for the modified model. The coefficients on *CEO First Year* is -3.277 (t-statistic = -3.51), *CEO Second Year* is -2.452 (t-statistic = -3.12), *CEO Third Year* is -2.383 (t-statistic = -3.02), *CEO Fourth Year* is -1.913 (t-statistic = -2.42), and *CEO Fifth Year* is -2.051 (t-statistic = -2.51). Thus, it can be observed that the incidence of negative abnormal tone (or downward tone management) is highest in the CEOs' first year and then declines linearly over time as their tenure progresses, which is consistent with the above notion.

Corroborating the *Litigation* Explanation: Forward-looking Disclosures in the CEOs' Early Years

To further corroborate the litigation argument, I examine the quantity of forward-looking disclosures contained in the firms' 10-K filing in the CEOs' early years.

Prior research documents that a threat of litigation can potentially reduce managers' incentives to provide forward-looking disclosures (Weetman and Collins 1996). As per the survey evidence in Graham et al. (2005), 46.4 per cent of the executives agreed or strongly agreed when asked whether they would want to avoid possible lawsuits if future results don't match forward-looking disclosures. If this concern is especially binding on the CEOs in the early years of their tenure when the market is still assessing their ability, we should expect fewer forward-looking disclosures during this time. Furthermore, Li et al. (2014) state that absorbing knowledge takes time and individuals newly placed in the CEO position are likely to have less specific knowledge about the firm's operations and future prospects. Thus, it would be highly risky and potentially costly for the CEOs to issue forward-looking statements in their early years as compared to the later years.[25]

To examine this, first, I measure the level of forward-looking disclosures in the 10-K using the forward-looking intensity (*FLI*), which is defined as the number of forward-looking sentences divided by the total number of sentences in the entire 10-K document. To identify forward-looking sentences in 10-K, I employ the methodology developed in Muslu et al. (2015), which includes a comprehensive list of future-oriented keywords and phrases in order to distinguish forward-looking statements from other types of statements (such as those related to past

[25] Although the safe harbor provisions of the Private Securities Litigation Reform Act (PSLRA) of 1995 encourage greater forward-looking information in corporate disclosures, firms may still be reluctant due to their uncertainty regarding the judicial interpretation of these provisions and on account of fears regarding litigation where no such safe harbor is available (Grundfest and Perino 1997).

events, boilerplate sentences etc.). Then, I use the following model (similar to Muslu et al. (2015)) that controls for firm characteristics identified as being determinants of forward-looking disclosures in Li (2010a) and thus ensures focus on the *abnormal* forward-looking intensity:

$$FLI_{j,t} = \alpha_0 + \alpha_1 CEO\ Early\ Years_{j,t} + \alpha_2 ANALYST_{j,t} + \alpha_3 SIZE_{j,t} + \alpha_4 EARN_{j,t} + \alpha_5 LOSS_{j,t} + \alpha_6 RET\ VOL_{j,t} + \alpha_7 EARN\ VOL_{j,t} + \alpha_8 SI_{j,t} + \alpha_9 MA_{j,t} + \alpha_{10} BTM_{j,t} + \alpha_{11} AGE_{j,t} + \alpha_{12} ln\ BUS\ SEG_{j,t} + \alpha_{13} ln\ GEOG\ SEG_{j,t} + \varepsilon_{j,t} \quad (3)$$

All variable definitions are outlined in Appendix A. Model (3) includes both year and firm (or industry) fixed effects. Standard errors were clustered at the firm level.

Table 5 presents the regression results from the estimation of model (3).[26] In both specifications – one with industry and year fixed effects (column 1) and the other with firm and year fixed effects (column 2), the coefficient on *CEO Early Years* is significantly negative, suggesting that the proportion of abnormal forward-looking disclosures contained in the 10-K document is lower in the CEOs' early years as compared to the later years of their tenure. This finding provides yet another piece of corroborating evidence in favor of CEOs' heightened litigation concerns during the early years of their tenure. Also, most of the significant control variables (except *SIZE*) load with the expected sign, similar to Muslu et al. (2015). Firms with more available earnings-related information, poorly performing firms, firms with volatile business, smaller firms, growth firms,

[26] To ensure robustness of results, 211 firm-year observations were deleted out of a total of 13,435 observations using the DFBETA cutoff value of 3/sqrt (n).

and less complex firms make more forward-looking disclosures, as evidenced by the positive and significant coefficients for *ANALYST*, *LOSS*, *RET VOL*, and *EARN VOL* and significantly negative coefficients for *SIZE*, *EARN*, *AGE*, and *ln BUS SEG* in column (1) of the table.

IV. THE CASE OF HIGH-ABILITY CEOs

In this section, I examine the role of ability in the choice of CEOs' tone management strategy. In the context of CEO tenure, it is possible for different CEOs to perceive the severity of the litigation concerns differently in their early years, depending on their respective ability. For the high-ability CEOs, the motivation to manage disclosure tone in their early years of tenure is likely to be substantially different from that of the other CEOs.

To investigate this, I employ a residual-based measure of managerial ability (*ABILITY SCORE*) from Demerjian et al. (2012), which is based on managers' efficiency in generating revenues. Demerjian et al. (2012) create a total firm efficiency measure using data envelopment analysis (DEA), and then "purge" it of key firm-specific characteristics that are expected to aid (like firm size, market share etc.) or hinder (like complex multi-segment and international operations) management's efforts, besides removing industry and time effects (via Tobit regression). The unexplained portion of firm efficiency (residual term) is then attributed to management, and constitutes their measure of managerial ability. Demerjian et al. (2012) run a host of validity tests to corroborate this measure and establish its superiority over alternate CEO ability measures. For instance, they show that their ability measure is strongly associated with manager fixed effects, suggesting it reflects manager characteristics. In my sample, the mean value of *ABILITY SCORE* is 0.01, median value is 0.00, and values range from -0.41 to 0.52 (see Table 2, Panel A).

The regression results from the estimation of model (2) after including *ABILITY SCORE* and the corresponding interaction term are reported in Table 6.[27] As can be observed, the coefficient on both *ABILITY SCORE* (6.546; t-statistic = 2.54) and its interaction with *CEO Early Years* (6.858; t-statistic = 1.955) are positive and significant, suggesting that abnormal tone is positively associated with the level of CEOs' ability, and this association is significantly stronger in the early years than in the later tenure years of more able CEOs.

Hence, it is established that more able CEOs manage tone upwards, and that such tone management is significantly greater in their early years of tenure as compared to the later years. These CEOs adopt an overoptimistic tone for their firms' 10-K narrative disclosures, since litigation concerns are (relatively) less binding for them in the early years of their tenure. To understand this better, recall that market participants would discount the information provided in firms' disclosures over time only when these disclosures do not correspond to future firm performance (Rogers et al. 2011). But high-ability CEOs ensure the subsequent good performance of their firms, thus making any upward tone management in the current period much harder to detect subsequently.

[27] To ensure robustness of results, I delete the top 1 per cent of firm-year observations (415 out of a total of 13,971) after computing the (absolute) DFBETA values of the relevant variables.

V. FINAL YEAR OF CEOS' TENURE AND TONE MANAGEMENT

In this section, I examine CEOs' tone management strategy in the final year of their tenure. To do this, I estimate model (2) after replacing *CEO Early Years* with the *CEO Final Year* variable. Everything else remains the same.

Columns 1-2 in Table 7 report the regression results. As observed from the results stated in column (1) for the basic model (with only firm and year fixed effects), the coefficient on *CEO Final Year* is negative and significant, -2.526 (t-statistic = -4.19), suggesting that abnormal tone is significantly more negative in the final years than in the earlier years of CEOs' tenure. Even in the full model with controls (column 2), the coefficient on *CEO Final Year* is negative and significant, -4.314 (t-statistic = -5.61). Thus, departing CEOs also engage in greater downward tone management in their final year as compared to the earlier years.

I explain the above finding as follows. First, note that the most important reason for CEO departure is retirement,[28] as CEO exits for taking a CEO position in another firm are very rare (Gibbons and Murphy 1992; Brickley et al. 1999). Second, as argued in Brickley et al. (1999), CEOs care about post-retirement opportunities, and many CEOs remain active in their retirement years serving on corporate boards. Specifically, they find that of the CEOs who retire at age 64, 65, or 66 (the most common ages for retirement), nearly 88 per cent hold at least

[28] Across my sample, 2,490 out of 4,797 (51.91 per cent) observations for which the reason of CEO departure is non-missing in *ExecuComp* correspond to these retiring CEOs.

one board seat, 42 per cent hold three or more seats, and just over 28 per cent hold four or more seats. Furthermore, the availability of these opportunities depends heavily on the CEOs' performance during their final year(s). For instance, Brickley et al. (1999) report that retiring CEOs who stay on their own board generate about 10.9 per cent higher annual abnormal stock returns over their final years than CEOs who do not. Given this evidence, I argue that departing CEOs attempt to lower expectations by adopting an overly pessimistic tone in their final year, since the adoption of an overoptimistic disclosure tone instead could potentially damage their professional reputation and image if the firm subsequently underperforms.[29]

Column (3) in the table presents results for the specification with added interaction terms. Again, it can be observed that the coefficient on *CEO Final Year* is negative and significant, but interestingly, the interaction of *CEO Final Year* with *NOA* is positive and significant (12.637, with t-statistic = 5.14). The latter finding implies that when CEOs are constrained in manipulating accruals due to the balance sheet constraint, they resort to upward tone management in the final year of their tenure. To understand this result, note that most prior studies have documented that departing CEOs engage in greater accruals management in their final year (although the empirical findings are mixed with

[29] I assume that CEOs can predict their turnover, and thus have enough time to influence the annual reports. Voluntary retirement and resignations can, of course, be predicted by CEOs. Also, even if CEOs are asked to leave, they are usually given time to prove themselves (Pourciau 1993).

regards to its direction).[30] This further explains their reluctance to manage tone upwards, since simultaneously engaging in both upward tone management and accruals manipulation could potentially aggravate their reputational concerns. However, for departing CEOs in balance sheet constrained firms (a proxy for the limits to accruals management), such reputational concerns are then relatively less serious and so they are less averse to upward tone management in their final year.

Finally, in the context of the CEOs' final year, I also run another specification where I add *CEO Early Years* and the associated interaction terms as additional explanatory variables. Not controlling for tone management in the early years of CEOs' tenure when testing for tone management in the final year could potentially be misleading, since many CEOs leave office within the first few years of starting their jobs and so the difference in tone management between their final year and the other years is likely to be small (given their incentive to manage tone in the early years as well). The choice of this particular specification is motivated from Ali and Zhang (2015), who show that close to 30 per cent of the CEOs left office within the first four years of starting their jobs, during 1992-2010. The regression results are reported in Column (4) of Table 7. The evidence of greater downward tone management in the CEOs' final year and the positive

[30] For instance, while most studies find evidence consistent with greater earnings overstatement in the CEOs' final year (DeAngelo 1988; Ali and Zhang 2015), Pourciau (1993) finds that departing executives record income-decreasing accruals.

interaction between *CEO Final Year* and *NOA* persists even after including *CEO Early Years* as an additional explanatory variable.

VI. ADDITIONAL ANALYSIS

Does 10-K Abnormal Tone Predict Future Financial Performance?

In this subsection, I examine overall whether the abnormal tone of narrative 10-K disclosures predicts future financial performance incremental to the reported accounting numbers (including accruals management). If *AB_TONE* positively predicts future performance, then it reflects the CEO's private information that cannot be conveyed through quantitative disclosures contained in the 10-K filing. On the other hand, if *AB_TONE* negatively predicts future performance, then CEOs likely employ tone management opportunistically to either hype or depress current performance, depending on their specific incentives. Using earnings press releases from 1997-2007, Huang et al. (2014) obtains a negative association between positive abnormal tone and future (one to three years ahead) earnings and cash flows.

I use the following model from Huang et al. (2014):

$$PERFORM_{j,t+n} = \alpha_0 + \alpha_1 AB\_TONE_{j,t} + \alpha_2 DA_{j,t} + \alpha_3 EARN_{j,t} + \alpha_4 SIZE_{j,t} + \alpha_5 BTM_{j,t} + \alpha_6 RET_{j,t} + \alpha_7 RET\,VOL_{j,t} + \alpha_8 EARN\,VOL_{j,t} + \varepsilon_{j,t}\,,$$

$$where\ PERFORM = (EARN\ or\ CFO)\ and\ n = (1, 2, \text{or } 3) \quad \text{- (4)}$$

All variable definitions are outlined in Appendix A. The model contains both industry and year fixed effects, and standard errors are clustered by firm.

Table 8 presents the regression results from the estimation of model (4). For both one- to three-year-ahead *EARN* (columns 1-3) and *CFO* (columns 4-6) regressions, I obtain a significantly negative association between the 10-K abnormal tone and future performance, after controlling for accruals management (*DA*) and other quantitative variables. Hence, 10-K abnormal tone negatively predicts future performance, thereby implying that CEOs manage the tone of their 10-K narrative disclosures for opportunistic reasons and not to convey their private information to market participants.

Robustness of Results

All the results documented in the current paper are robust to DFBETA tests using multiple cutoff values, thus ruling out the effect of any outliers.

Furthermore, to ensure that my results for tone management in the CEOs' early years of tenure are not sensitive to the choice of the cutoff value of three years for defining the *CEO Early Years* variable, I rerun model (2) using alternate cutoff points, namely, two and four years. The results are reported in Columns 1 and 2 of Table 9. In both cases, that is, when *CEO Early Years* is defined based on the two years (Column 1) and four years (Column 2) cutoff points, the earlier result of greater downward tone management in the CEOs' early years of tenure still holds. In unreported results, I also establish the robustness of my results to the specific methodology (mainly with regards to the treatment of missing values)

employed by Ali and Zhang (2015)[31] to create the *CEO Early Years* and *CEO Final Year* variables.

I also rerun model (2) after replacing the options-based measure of CEO optimism with a couple of investment-based proxies of optimism (or overconfidence) from prior literature (Schrand and Zechman 2012; Ahmed and Duellman 2013). Both these alternate proxies, *CEO Optimism_CAPX* and *CEO Optimism_XSINVEST*, are defined in Appendix A. Columns 3 and 4 of Table 9 present the results from using these proxies, respectively. As before, it can be observed that the coefficient on *CEO Early Years* is negative and significant. In unreported results, I find that the earlier result in the case of CEOs' final year also remains unchanged when these proxies are used instead.

[31] Ali and Zhang (2015) treat the starting date of the upcoming CEO as the leaving date of the previous CEO, if the "LEFTOFC" variable is missing in the *Execucomp* database. And if the "BECAMECEO" variable is missing, but "LEFTOFC" is available, they treat the leaving date of the previous CEO as the starting date of the upcoming CEO. In this way, they combine these two variables to calculate tenure, first year and final year.

VII. CEO TENURE AND 10-K READABILITY

In this section, I expand the scope of the discussion so far in the paper to CEO qualitative management at different stages over their tenure, by additionally examining another crucial linguistic attribute, namely, the readability (or complexity) of the 10-K document. Text readability is considered as a proxy for obfuscation, where the presumption is that "preparers manipulate transparency by reducing clarity when they wish to disclose less about their underlying circumstances" (Rutherford 2003). Prior studies have investigated whether managers attempt to manipulate outside perceptions of firm performance by increasing the complexity of corporate narrative disclosures, or whether their intent is simply to leave the readers confused and discourage them from probing further (Merkl-Davies and Brennan 2007). In the context of the CEOs' tenure, it could thus be interesting to examine whether they regard document readability as a complement to tone management by also simultaneously managing the readability of their firm's narratives differentially in the early (and final) years of their tenure, in line with their specific concerns at that time. In the context of CEOs' early years, recall from the discussion in Section III that the litigation concerns are especially pressing for the CEOs during this time when the market is still assessing their ability, which in turn may create incentives to manage readability.

I use three different measures of annual report readability from prior literature for my analyses, namely, *FOG* (Li 2008), *LENGTH* (Li 2008), and *FILESIZE* (Loughran and McDonald 2014). These measures have been defined in

Appendix A. *FOG* captures the syntactical complexity of text, and is based on the number of syllables in a word and the number of words in a sentence. It is also the most popular of the three measures. However, Loughran and McDonald (2014) argue that *FOG* is a poorly specified measure of readability, since the usage of multisyllable words is a common feature in business texts and readily understood by investors and analysts. They further argue that if the objective of the managers is to obscure relevant performance-related information, they are more likely to "bury the results in longer documents" and less likely to employ "complex rhetoric". Loughran and McDonald (2014) propose using the 10-K file size (*FILESIZE*) instead, due to its ease of computation and replicability.

Given the various determinants of annual report readability identified in Li (2008), I employ the following model (where *READABILITY* refers to any of the three measures mentioned earlier):[32] [33]

$$READABILITY_{j,t} = \alpha_0 + \alpha_1 CEO\ Early\ Years_{j,t} + \alpha_2 SIZE_{j,t} + \alpha_3 BTM_{j,t} + \alpha_4 AGE_{j,t} + \alpha_5 SI_{j,t} + \alpha_6 RET\ VOL_{j,t} + \alpha_7 EARN\ VOL_{j,t} + \alpha_8 ln\ BUS\ SEG_{j,t} + \alpha_9 ln\ GEOG\ SEG_{j,t} + \alpha_{10} SEO_{j,t} + \alpha_{11} MA_{j,t} + \alpha_{12} DLW_{j,t} + \varepsilon_{j,t} \quad (5)$$

All variable definitions are in Appendix A. The model includes both industry and year fixed effects, and standard errors are clustered at the firm level.

[32] Although Li (2008) finds that managers increase the complexity of the 10-K document when firm performance is poor (making earnings a strategic determinant of readability), I do not control for current earnings in this model as I would like the "abnormal" readability (i.e., after controlling for its normal determinants) to reflect these strategic incentives.

[33] The *FOG* data was obtained from Li (2008), while *SEO* was constructed using data from the *SDC Global New Issues* database.

Panel A of Table 10 presents the summary statistics. Note that the 10-K filing for an average firm in my final sample consists of a total of about 54,521 words. Panel B presents the results from the empirical estimation of model (5) for each of the three readability measures.[34] Most of the significant control variables load with the expected signs (similar to Li (2008)). As observed in column (1) which states the results for the *FOG* measure, the coefficient on *CEO Early Years* is negative and significant, suggesting that CEOs decrease the complexity of the 10-K document in their early years as compared to the later years. But when *FILESIZE* is used as the dependent variable (column 3), I find that the main coefficient is significantly positive, thereby implying that CEOs decrease annual report readability in their early years by issuing longer 10-K documents. Note that the explanatory power of the model specification in column (3) is the highest, as observed from its adjusted R-squared value of 63.1 per cent. However, since the main result is sensitive to the choice of the readability measure, nothing conclusive can be said with regards to the CEOs' readability management in their early years of tenure as compared to the later years.

In unreported results, I find that 10-Ks are both lengthier (significant positive association with *LENGTH* and *CEO Final Year*) and longer (significant positive association with *FILESIZE* and *CEO Final Year*) in the CEOs' final year as compared to the previous years.

[34] To ensure robustness of results, 203, 230, and 205 firm-year observations were deleted out of a total of 12,004, 13,936, and 13,936 observations in the *FOG*, *LENGTH*, and *FILESIZE* regressions respectively, using the DFBETA cutoff value of 3/sqrt (n).

VIII. CONCLUSION

This study examines changes in CEOs' incentives to engage in tone management during their tenure as CEO. For the sample period 1993-2010, I show that in the early years (the first three years) of their tenure, when the market is still assessing their ability, CEOs attempt to lower outsiders' expectations by depressing firm performance using an overly pessimistic tone (greater downward tone management) for their firms' 10-K narrative disclosures. This is partly explained by their heightened litigation concerns during this time, since optimistic language has been found to be significantly associated with greater litigation risk and CEOs worry about a potential litigation filing due to the resulting loss of shareholders' wealth and personal credibility, both of which could adversely affect their continuance in office. An interesting finding in the paper is that more able CEOs, on the contrary, adopt an overly optimistic tone, and more so in their early years as compared to the later years. These findings are directionally opposite to those reported in some prior impression management studies around CEO turnovers. For instance, Ali and Zhang (2015) focus on earnings management at different stages of the CEOs' tenure, and document that all CEOs, regardless of ability, manipulate earnings upwards in the early years of their tenure.

I also show that departing CEOs of only balance sheet-constrained firms adopt a more optimistic tone in their final year, even though the overall evidence is consistent with greater downward tone management in the final year. These

results are robust to using different definitions of CEOs' early years (using different cutoff points) and alternate proxies for the controls.

The findings in this study have implications for research on CEO behavior over their tenure and especially around turnovers, particularly with regards to impression management. While prior studies on this subject have mostly focused on quantitative tools (like earnings management, write-offs, or abnormal stock returns), this paper projects tone management as another strategic tool in the hands of CEOs to manage perceptions and expectations in their early years and final year. It further documents the role of contextual factors, such as ability and litigation risk, in influencing CEOs' strategic disclosure choices (with regards to tone) at different stages of their tenure.

Given the advancement of textual techniques facilitating ease of computation and the consequent increase in the number of studies that explore myriad attributes of corporate narrative disclosures, there is scope to further extend the analyses in the paper by examining other interesting aspects of the CEOs' qualitative management strategy over their tenure.

CHAPTER 2

DOES GREATER R&D QUALITATIVE DISCLOSURE PROVIDE INFORMATION ABOUT FIRM PROFITABILITY? (with Sanjay Kallapur and Ankit Jain)

1. INTRODUCTION

Merkley (2014) examines whether managers adjust the narrative disclosures of research and development (R&D) activities in their firms' 10-K filings to cater to the investors' information demands based on variation in earnings performance. Besides documenting empirical evidence in support of this hypothesis,[35] he establishes the informativeness of such disclosures by showing a positive association of R&D disclosure quantity with both analyst following and analyst forecast accuracy, and a negative association with information asymmetry around the 10-K filing date. He, however, has not examined the association of R&D disclosures with future firm profitability. This is important as the "proof of the pudding" lies in their association with future earnings, since the role of R&D is essentially to generate earnings in the future.

In this paper, we examine the future effects of R&D disclosures by studying the association of R&D disclosure quantity with future profitability. There are at least two reasons to expect different results to those implied by Merkley (2014) regarding the short-term effects of R&D disclosures. First, it is possible that managers possess superior information about the outcome of R&D investments, but strategically choose whether and what to disclose. Second, it could be that managers do not have significant information about the future success of R&D investments by falsely relying on their intuitive judgments (forecasts and

[35] We replicated the main empirical test in Merkley (2014), and obtained qualitatively similar results.

predictions) in an environment of low validity[36] and deferred (or delayed) outcome feedback, such as R&D. Kahneman and Klein (2009) argue that the necessary conditions for the development of *skilled* intuitions include presence of a high-validity environment (environment of "sufficient regularity", which provides valid cues), and adequate opportunities for learning the environment ("rapid" and "unequivocal" feedback).[37] Given the typical characteristics of the R&D disclosure setting, it seems difficult that precise and accurate intuitions will develop in this environment, thereby leading to mostly biased disclosures.[38]

Using a *bag of words* approach following Loughran and McDonald (2011), we capture the quantity of a firm's 10-K narrative R&D-related disclosure by constructing a *weighted* word-count-based measure, and then examine the association of our measure with future firm profitability (measured by return on assets (ROA)) using 20,655 firm-year observations from publicly-listed US firms during 1993-2006.[39] We obtain a significant (statistical and economic) and negative association between our measure of R&D disclosure quantity and future

[36] Kahneman and Klein (2009) describe *Validity* as the "causal and statistical structure of the relevant environment". It can be inferred from their paper that a low-validity environment is one which is irregular or unpredictable. We label R&D as a low-validity environment because R&D investments are highly firm-specific, challenging, and characterized by high risk and uncertainty. It is, thus, difficult (even impossible) to precisely estimate the outcome of an R&D investment due to the multitude of possible outcome scenarios, which in turn vary by project and across time.

[37] Even though true skill cannot form in unpredictable or irregular environments, sometimes decisions and judgments will be successful by chance, making the decision-makers susceptible to an "illusion of skill" and to overconfidence (Arkes 2001; Kahneman and Klein 2009).

[38] This view about inferiority of judgments due to systematic errors resulting from experts' uncritical reliance on their intuition is also shared by Meehl (1973).

[39] Note that Merkley (2014), on the other hand, looks at the association of current profitability with future R&D disclosure quantity by using a measure of R&D disclosure quantity that is based on *sentence-count* as opposed to *word-count*. Although there are no obvious advantages of using one versus the other, word-count measures are relatively more popular in the narrative disclosures literature. However, our results remain unchanged when we use a sentence-count-based measure instead. Furthermore, although *term-weighting* is desirable (as we argue later in the paper), our results hold when we use an unweighted (plain word-count) measure instead of a weighted one.

ROA (also, *adjusted* ROA – computed by adding back R&D expenditures), even after controlling for the innovation efficiency measures based on patents and citations (Hirshleifer et al. 2013), R&D expenditure, 10-K length, tone of forward-looking disclosures, and industry (or firm) and year fixed effects. Specifically, an increase from the 25th to the 75th percentile in R&D disclosure quantity is associated with 38.27 per cent (33.67 per cent when the sentence-based measure of R&D disclosure quantity from Merkley (2014) was used) decline in ROA for an average firm in our sample. Furthermore, we find that R&D disclosures made in the past three years are also negatively associated with future profitability, and thus conclude that this negative association is persistent.

To rule out any measurement error concerns that the observed negative association may be driven by the amount of negative or uncertainty-related words contained in the R&D narrative disclosures, we show that the association persists even after including two measures of negative R&D sentiment (*R&D Pessimism* and *R&D Uncertainty*) along with their corresponding interaction terms, where both interactions were found to load insignificantly. To further investigate this association, we test for two strategy-based explanations - namely, the strategic disposition of managers (captured by the readability (or complexity) of the 10-K text using the Fog index, similar to Li (2008)), and competitive pressures facing the firm (captured using three different measures of competition, viz., Herfindahl-Hirschman Index (HHI) and two measures from Karuna (2007)). Results indicate that both these explanations fail to explain the observed bias, and our main (negative) association remains highly significant

across all such specifications (although the role of other strategy-based explanations cannot be ruled out). This evidence of a strong negative association is surprising, since R&D disclosures have been found to be positively correlated with future firm fundamentals (Gu and Li 2003), and the current consensus is that managers adjust them in response to earnings performance in order to provide more relevant information to investors (Merkley 2014; Gu and Li 2003). Moreover, the argument about managers' unskilled and biased intuitive judgments in the R&D disclosure environment predicts an insignificant association between R&D disclosure quantity and future ROA. Hence, our finding of a significant negative association implies that it is the worst R&D performers who are also the most biased, which is not surprising as Kruger and Dunning (1999) have argued that the poorest performers hold the least accurate assessments of their skills and thus end up overestimating their performance relative to that of their peers.

We run some additional tests. First, we decompose the R&D disclosures (identified at the sentence level) into forward-looking (FLS) and non-forward-looking (non-FLS), and obtain a significant negative association with future ROA for only the FLS R&D disclosure quantity. If managers cannot skillfully predict the *future* R&D outcome (given the typical features of the R&D environment), then the "bias" (earlier negative association: worst performers also the most biased) should only stem from FLS (which pertains to the future) and not non-FLS. Second, we perform a host of empirical tests to document the robustness of our results to alternate measures of R&D disclosure quantity (sentence-count-based;

unweighted word-count-based), alternate firm profitability measures (adjusted ROA; cash flow from operating activities (CFO)), inclusion of only the R&D-intensive firms (identified from Hirshleifer et al. (2013)), and choice of sample period (using 1993-2010 instead of 1993-2006).

Our study contributes to this literature in several ways. First, we contribute to the extant literature on R&D – an area of great significance that constitutes an important topic of research across management specializations,[40] and one that is interesting on account of its unique characteristics. Prior studies have documented that future earnings depend on increases in R&D expenditures (Eberhart et al. (2004, 2008)), innovative efficiency (Hirshleifer et al. 2013), (adjusted) patent citations (Gu 2005; Pandit et al. 2011), and industry-level technological progress (Matolcsy and Wyatt 2008). In this context, our paper examines whether and how narrative R&D disclosure quantity contributes to that information mix.

Second, as the first large-sample empirical study to examine the earnings-related future consequences of narrative R&D disclosures, we also contribute specifically to the literature on R&D disclosures. Firms' R&D disclosures contained in their annual filings constitute an important disclosure-type because they are a tool to reduce the information asymmetry between managers and investors, which is a typical feature of R&D investments.[41] Prior literature on the determinants and

[40] In the year 2014, for all firms that invested in R&D across all industries, R&D capital as a percentage of book value of equity was 29.3 percent. (Refer to Appendix A1 for variable definitions and measurement.)

[41] Information problems arise as accounting rules are different for R&D investments, mandating immediate expensing without a need to provide enough disclosure to investors about the same (Lev 2001; Morricone and Oriani 2009; Grandi et al. 2009).

consequences of R&D disclosure quantity is limited, comprising mainly of small-sample studies that focus only on specific industries or time periods (such as, Gu and Li (2003), and Jones (2007)). To the best of our knowledge, Merkley (2014) is the only large-sample empirical study on R&D narrative disclosures; although he examines the determinants of such disclosures as opposed to their consequences.

Third and finally, it enhances our understanding of the firms' narrative disclosure environment, and adds to the few studies which show that the information content of these disclosures may not be meaningful, that is, it does not correlate positively with future firm performance.[42] We focus on the R&D disclosure environment; where R&D is characterized by high risk and uncertainty, high information asymmetry, and deferred outcome feedback. These unique characteristics in turn adversely affect the credibility and effectiveness of the firms' R&D-related narrative disclosures by making it difficult for managers to develop skilled intuitive judgments about the future success of R&D investments. Thus, we show that the disclosure-type and the features of its environment are important considerations while assessing the informativeness of firms' narrative disclosures.

[42] The content and lexical features of narrative disclosures have been found to be informative beyond accounting numbers, and most prior literature has documented a positive association with future firm fundamentals. For instance, the positive correlation between references to competition and mean reversion in return on net operating assets (Li et al. 2013); between the frequency of negative words [identified using the dictionary provided by Loughran and McDonald (2011)] and incidence of "liquidity events" such as debt downgrades and dividend cuts (Bodnaruk et al. 2013); between trust in a firm's corporate culture and subsequent share price volatility (Audi et al. 2014); between financial constraints [defined using the textual measure in Bodnaruk et al. (2013)] and aggressive tax planning strategies (Law and Mills 2015) etc.

2. EMPIRICAL METHODOLOGY

Measuring the quantity of 10-K narrative R&D disclosures

We capture the quantity of a firm's R&D-related narrative disclosures by constructing a textual measure from the annual 10-K filing. Textual measures have the advantage of capturing information from many different sources that are hard to identify empirically (Li 2010). Our choice of 10-Ks is motivated by prior research suggesting that 10-K filings are an important source of information (Previts et al. 1994; Leder 2003; Brown and Tucker 2010; Lehavy et al. 2011; Merkley 2014). Furthermore, the information on R&D activities is largely descriptive in nature; a significant portion of which is contained in the qualitative 10-K disclosures (Entwistle 1999; Jones 2007). We define R&D disclosure quantity as the proportion of R&D related words in the 10-K.

We employ the bag of words approach to represent the 10-K text numerically. Under this approach, each document is represented by the words it contains, ignoring any punctuation and ordering. Every word is identified and counted the number of times it appears in the document. We also use an algorithm to reduce each word to its 'stem', so that different forms of the same word are considered as a single word (for example, the words "develop", "developed", "developing", and "development" are stemmed to "develop").

Although textual measures based on a simple word-count have been widely used in prior literature[43], raw word count is not the most accurate measure of a word's information content (Loughran and McDonald 2011). It is critical, therefore, to weight the

[43] Li et al. (2013) construct a measure of competition; Bodnaruk et al. (2013) construct a measure of financial constraints based on the frequency of negative words; Audi et al. (2014) construct a measure of trust in a firm's corporate culture.

word counts properly, since the adoption of an appropriate term-weighting scheme can have a significant impact on the effectiveness of the information thus retrieved (Buckley 1993; Jurafsky and Martin 2009; Loughran and McDonald 2011).

We adopt the following term-weighting scheme from prior literature, but modify it to account for the variation over time [similar to Loughran and McDonald (2011)]:

$$w_{i,j,t} = \begin{cases} \frac{(1 + \log(tf_{i,j,t})}{(1 + \log(a_{j,t})} * \log(\frac{N_t}{df_{i,t}}), & \text{if } tf_{i,j} \geq 1 \\ 0, & \text{Otherwise} \end{cases}$$

where $a_{j,t}$ denotes the average word count of documents in year t, $tf_{i,j,t}$ is the raw count of the i^{th} word in the j^{th} document in year t, $df_{i,t}$ represents the number of documents containing at least one occurrence of the i^{th} word in year t, and finally N_t is the total number of 10-K documents in year t. The above weighting scheme offers several benefits. The term frequency (*tf)* helps attenuate the impact of high-frequency words, which is accomplished by the logarithmic transformation[44]. Furthermore, stop words such as *"the"*, *"of"*, *"all"*, *"for"* etc. are suitably assigned a weight of zero, since these words appear frequently in all documents and do not provide any information. The document frequency (*df*) in the weighting scheme accounts for the commonality of words, that is, it assigns lesser weight to words that are commonly used across the 10-

[44] As an example, consider the word *"gain"* which appears 92,286 times across our 10-K sample in the year 2006, and another word *"supportable"* that appears only 326 times. The weighting scheme will assign a lesser weight to *"gain"*, since it is unlikely that the collective impact of *"gain"* is more than 283 (=92286/326) times that of *"supportable"*.

K documents[45]. Finally, the weighting scheme also accounts for the average length of 10-K document (*a*) and its variation across our sample period, since the 10-Ks have become significantly lengthier over time and it is more likely for a word appearing in 1994 to have a different impact than a word appearing in 2006.

Our textual measure, $R\&D\ QTY_{it}$, is defined as the ratio of the weighted count of R&D-related words in firm *i*'s 10-K document in year *t* to the weighted count of the total number of words in that document.[46] We use Python scripts to search for R&D-related words in the entire document on each 10-K filing.[47] To identify R&D-related words, we refer to the dictionary of commonly-used R&D keywords and phrases developed by Merkley (2014) and modify it to include only words.[48] Examples of R&D-related words from our (modified) dictionary include "research", "innovate", "breakthrough", "development", "clinical" etc. The dictionary in Merkley (2014) was compiled after consultation with industry personnel on R&D disclosure topics, and even compares to those of James and Shaver (2009) and Entwistle (1999), thus assuring us of its credibility.

[45] For example, in the year 2006, the word *"risk"* appears in 7,323 documents across our 10-K sample, while the word *"insurgency"* appears in only 13 documents. The second term in the weighting scheme will adjust the first term (increase for *"insurgency"*, and decrease for *"risk"*) appropriately to reflect this feature.

[46] We additionally scale our measure by a constant (1000) for ease of interpretability of its coefficient estimate.

[47] We do not extract R&D-related words from only a specific 10-K section since the amount of R&D-relevant information is spread over various sections throughout the entire 10-K (Entwistle 1999), including MD&A (22.7%), Corporate Overview (50.8%), signed letters section (18.9%), audited financial statements (1%), and a separate R&D section (6.6%).

[48] Since our measure is based on a word count, unlike sentence count as in Merkley (2014), we modified this dictionary of R&D phrases to include only words. We validated the modified dictionary through a manual examination of a random selection of 100 10-K filings.

The association of narrative R&D disclosure quantity with firm profitability

To examine the association of our textual measure of R&D disclosure quantity (*R&D QTY$_{it}$*), with subsequent firm profitability, we estimate a (OLS) model similar to Hirshleifer et al. (2013), stated as follows:

$$\mathrm{ROA}_{i,\,t+1} = \alpha_0 + \alpha_1 \ln(\mathrm{R\&D\ QTY})_{\mathrm{i,t}} + \alpha_2 \mathrm{ROA}_{i,t} + \alpha_3 \Delta \mathrm{ROA}_{i,t} + \gamma \mathbf{Z}_t + \varepsilon \quad \text{-- (1)},$$

where $\mathrm{ROA}_{\mathrm{i,t+1}}$ ($\mathrm{ROA}_{i,t}$) is firm *i*'s Return on Assets in year *t+1 (t)*; $\Delta\mathrm{ROA}_{\mathrm{i,t}}$ is the change in ROA between year *t* and year *t-1*; and $\mathbf{Z}_t$ is a vector of controls, including the innovative efficiency (IE) measure(s) from Hirshleifer et al. (2013), R&D expenditure, advertising and capital expenditures, other innovation-related variables (such as, R&D growth and change in adjusted patent citations), and firm size.[49] We additionally control for 10-K length and the tone of forward-looking disclosures. We follow Muslu et al. (2015) to identify forward-looking disclosures in the 10-K, and calculate tone of these disclosures using the dictionary of positive and negative words from Loughran and McDonald (2011). All variable definitions and measurement have been outlined in Appendix A1. We include two-digit SIC industry dummies[50] in the regression to account for differences in industry characteristics or environments, along with year dummies and cluster all standard errors by both firm and year.

[49] As the distributions of our textual measure and some of the controls (including IE) are highly skewed and(or) are often zero, we use a logarithmic transformation of these variables, similar to Hirshleifer et al. (2013) and Lerner (1994).

[50] We do not include firm fixed effects in our main model to avoid inducing a Nickell's bias in the coefficient estimates. The estimates of a model having individual fixed effects and lagged value of the dependent variable as an independent variable (as in our case) are biased when estimated using an OLS (Nickell 1981). As a robustness test, we also include firm fixed effect instead of industry fixed effects after dropping $\mathrm{ROA}_{\mathrm{i,t}}$ and $\Delta\,\mathrm{ROA}_{\mathrm{i,t}}$ and get qualitatively similar results.

We include current ROA in the above model to capture the persistence in operating performance (Gu 2005; Pandit et al. 2011). Change in ROA serves as a control to account for the mean reversion in profitability (Fama and French 2000). More importantly, we control for the IE measure based on patents,[51] where IE is defined as the ability of the firm to generate patents per dollar of R&D investment (Hirshleifer et al. 2013). Hirshleifer et al. (2013) find a positive relation between this IE measure and future ROA, suggesting that IE measures contain incremental information about subsequent operating performance of the firms. We control for 10-K length and the tone of forward-looking disclosures to capture narrative disclosures other than R&D in the annual report, as prior literature (Li 2008, Li 2010a) has shown that these variables are informative about firms' profitability beyond accounting numbers. Finally, we control for advertising and capital expenditures as prior studies have found that they explain operating performance of the firm (Lev and Sougiannis 1996; Pandit et al. 2011; Hirshleifer et al. 2013).

We are interested in the coefficient of $ln(R\&D\ QTY_{it})$, which will inform us about the association between the quantity of narrative R&D-related disclosures contained in the firms' 10-K filing and future firm profitability (ROA).

[51] We also run all our main regressions using the other IE measure in Hirshleifer et al. (2013) – one based on citations, but do not tabulate those results in the paper.

3. DATA

Our sample includes firm-year observations from 1993 to 2006, and consists of firms in the intersection of the Compustat and SEC EDGAR databases, matched using the Central Index Key (CIK). We match remaining unmatched observations using the IRS tax identification number (Nini et al. 2012). We obtain relevant accounting data from Compustat files and the patents and citations data from the National Bureau of Economic Research (NBER) database. We remove firms in the finance, insurance and real estate sectors (SIC codes between 6000 and 6999), and those with negative book value of equity. Figure 1 describes our sample selection procedure in detail. Our final sample comprises of 20,655 firm-year observations from 3,703 unique firms.

We download all 10-Ks from the SEC EDGAR database. Following Loughran and McDonald (2011), we remove 10-Ks that contain less than 2,000 words, and only include one filing per firm per year by removing the filings that were filed within 180 days from a prior filing. In case there were multiple 10-Ks filed within a year, we consider only the first filing. The algorithm to parse the 10-K documents is outlined in Appendix B1.

Table 1 presents the summary statistics for our sample. All variables have been winsorized at the 1 per cent and 99 per cent levels. As shown in Panel A, the average size of a firm in our sample in terms of sales, total assets, and market value is $ 1,940 million, $ 1,932 million, and $ 2,683 million, respectively. Furthermore, a firm expends about $ 39 million on its R&D activities on average per year. Also, the mean innovation output (measured by the amount of patents granted) of a firm is 7. In terms of the

textual characteristics, the 10-K filing for an average firm in our sample consists of a total of about 37,098 words, of which nearly 167 are R&D-related.[52]

[52] To validate our measure of R&D disclosure quantity (*ln* (*R&D QTY*$_{it}$)) and ensure that the textual methodology is identifying the R&D-related words accurately, we manually read the10-K filings of the ten firms with the most R&D-related disclosures and the ten with the least amount of such disclosures, as identified by the *ln* (*R&D QTY*$_{it}$) measure, and noticed significant differences between them with regards to the number of R&D-related words contained in them.

4. RESULTS

Narrative R&D disclosure quantity and future ROA

We want to examine the association of *ln* (*R&D QTY*$_{it}$) with future firm profitability (ROA). Panel A of Table 2 presents our main results. We find that *ln* (*R&D QTY*$_{it}$) correlates negatively with subsequent ROA.[53] This association is statistically significant even after controlling for other narrative disclosures in the 10-K, R&D expenditure, the patents-based IE measure in Hirshleifer et al. (2013), and industry and time effects.[54] It is also economically significant – specifically, an increase from the 25th to the 75th percentile in R&D disclosure quantity results in a decrease in ROA of the magnitude of 0.011, which corresponds to a 38.27 per cent decline in ROA for an average firm in our sample (from Column 3, Panel A, Table 2).

Consistent with the literature, we find that the IE measures[55] [ln(1+*Patents/RDC*)] significantly predict higher ROA. The significantly positive slopes on ROA and the significantly negative slopes on change in ROA confirm both persistence and mean reversion in profitability. Firm size (measured as the natural logarithm of total assets) correlates positively with future ROA. Also, the coefficient on R&D-growth is

[53] When we decompose ROA into margin and turnover (DuPont decomposition), we find that the negative effect of R&D disclosure quantity holds for both.

[54] The association remains negative and significant when we deflate the relevant control variables by the market value of equity (similar to the model of operating performance in Hirshleifer et al. (2013)) instead of average total assets.

[55] In untabulated results, we find that the other IE measure [ln(1+Citations/RD)] also significantly predicts higher ROA and our main association remains significantly negative when it is included as a control (instead of the patents-based IE measure).

insignificant, similar to Hirshleifer et al. (2013). The coefficients on capital and advertising expenditures load insignificantly for our sample.[56]

The negative association remains significant when we use the sentence-count-based measure of R&D disclosure quantity used by Merkley (2014). Panel B of Table 2 presents the results. Specifically, an increase from the 25th to the 75th percentile in the sentence-count-based measure of R&D disclosure quantity results in a decrease in ROA of the magnitude of 0.01, which corresponds to a 33.67 per cent decline in ROA for an average firm in our sample (from Column 3, Panel B, Table 2).

Next, we check whether the negative association thus obtained is also persistent. As shown in Table 3, the association between *ln* (*R&D QTY*) and future ROA remains significantly negative till three years prior. These results imply that narrative R&D disclosures in a year can predict negative ROA for up to four years ahead. In unreported tests, we also include all lags of *ln* (*R&D QTY*) from *t* to *t-5* simultaneously, and find that the total association of *ln* (*R&D QTY*) with future ROA still remains negative and significant.

An important concern is that the observed negative association may be driven by the amount of negative words (pessimism) or uncertainty contained in the R&D disclosures. To check for this possible measurement error, we construct two tonal measures from the R&D narrative disclosures[57] (namely, *R&D pessimism* and *R&D uncertainty*[58]), and

[56] Hirshleifer et al. (2013) and Pandit et al. (2011) obtain a positive association between capital expenditures (intensity) and future ROA. For our sample, this association is positive and significant when we use a model specification with firm fixed effects.

[57] We construct these measures from the narrative R&D disclosures, as opposed to the entire 10-K. Narrative R&D disclosures comprise of the sentences containing R&D-related phrases, extracted using the Python algorithm as described in Appendix B1.

interact them with *ln* (*R&D QTY*$_{it}$) in two separate OLS regressions similar to model (1). Specifically, we define R&D pessimism (uncertainty) as the ratio of a weighted count of negative (uncertain) words contained in the narrative R&D-related disclosures to a weighted count of the total words in these disclosures. We employ the financial sentiment dictionary by Loughran and McDonald (2011) to identify the negative and uncertainty-denoting words in the R&D disclosures. This domain-specific dictionary is widely used by researchers to gauge the linguistic tone of text (see, for example, Feldman et al. 2010; Chen et al. 2014; Kearney and Liu 2014 etc.).

Panel B of Table 4 presents the results. For both *R&D pessimism* (column 1) and *R&D uncertainty* (column 2), we still obtain a significant negative association between *ln* (*R&D QTY*$_{it}$) and future ROA and the corresponding interaction terms load insignificantly. Hence, the negative association between *ln* (*R&D QTY*$_{it}$) and future ROA is not driven by the amount of negative or uncertainty-denoting words contained in R&D disclosures, thereby allaying any concerns with regards to measurement.

Furthermore, an alternative explanation for the negative association could be that narrative R&D disclosures comprise mainly of the manager's projection of future R&D spending, and once the firm actually incurs these expenditures in the subsequent period, the ROA for that period declines as a result. In unreported results, we find that when we adjust our ROA measure by adding back R&D expenditure and then rerun model (1) using this alternate measure, the significant negative association between *ln (R&D QTY)* and subsequent-period adjusted ROA still holds, thereby ruling out the

[58] We compute *R&D Uncertainty*, in addition to *R&D Pessimism*, to capture the uncertainty component of R&D disclosures. Some word examples from uncertainty dictionary include "ambiguous", "cautious", "confusion", "doubt", "unexpected" etc.

above possible explanation. Moreover, in an unreported test, we additionally control for other variables from prior literature (summarized and used by Merkley (2014)) which could potentially affect a firm's disclosure choices, such as firm age, outside monitoring (captured by analyst coverage and institutional holding), information uncertainty (captured by standard deviation of monthly returns and standard deviation of ROA), leverage, book-to-market ratio, tangible assets (PP&E and inventories), and stock issuance. We again obtain a significant and negative association between *ln (R&D QTY)* and future (adjusted) ROA.

This evidence of a strong negative association is surprising, since R&D disclosures have been found to be positively correlated with future firm fundamentals (Gu and Li 2003), and the current consensus is that managers adjust them in response to earnings performance in order to provide more relevant information to investors (Merkley 2014; Gu and Li 2003). Next, we test for two strategy-based explanations and then propose a psychology-based explanation in a bid to understand the aforementioned negative association.

Explaining the negative association between R&D disclosure quantity and future ROA

First, we propose that the managers' strategic motives may have a role in explaining the surprising negative association, and test for this using two scenarios – the readability (or complexity) of the 10-K document, and competitive pressures facing the firm. In other words, we hypothesize that either the manager's (overall) strategic disposition (or intent) as gauged from the 10-K complexity, or competition concerns could bias the disclosure of R&D in the firm's 10-K which then reflects negatively in future firm fundamentals.

To begin with, we consider the tendency of managers to disclose strategically by examining the readability (or complexity) of the 10-K document. Readability of text is an important attribute of narrative disclosures, and could be used strategically by managers (Li 2008; Merkley 2014). Specifically, Li (2008) finds that poorly performing firms tend to file annual reports that are complex and difficult to read. In the current context, we hypothesize that such deliberate obfuscation of the 10-K text by managers, which captures their overall strategic disposition (especially with regards to narrative disclosures), could make the aforementioned negative association stronger as R&D-related disclosures are a constituent of the firm's overall 10-K disclosure strategy and thus may be strategically biased.[59] We measure the readability of the 10-K document using the *Fog* index[60] – a popular computational linguistics tool that is a linear combination of number of syllables per word and the number of words per sentence, both directly proportional to document complexity, *ceteris paribus* (Li 2008). Higher values of this measure indicate greater complexity. The variable definition and construction is further outlined in Appendix A1.

Table 5 presents the results. Panel A presents the summary statistics for the *Fog* measure. Overall, the 10-Ks of firms in our sample are very difficult to read, consistent with the findings of Li (2008). The mean and median values for the *Fog* index of the entire 10-K document are 19.35 and 19.18, respectively. Panel B presents the regression results from adding the relevant interaction term in model (1). As shown in the table, the interaction between *In* ($R\&D\ QTY_{it}$) and the *Fog* index is statistically

[59] We consider the complexity of the entire 10-K document, instead of focusing only on the R&D-related disclosures, to capture the manager's overall strategic disposition which may reflect in other types of narrative disclosures too (not just R&D), such as those pertaining to mergers and acquisitions, new markets, firm growth and expansion, segment-wise performance commentary etc.

[60] We obtain fog index of 10-Ks from Li (2008).

insignificant, while the association between *ln (R&D QTY$_{it}$)* and future ROA remains significantly negative. Thus, the overall strategic disposition of managers fails to explain the observed negative bias.

Next, we turn our attention to any competition-related concerns that may lead managers to opportunistically bias their R&D disclosures. Prior studies have shown that competition affects the firms' voluntary disclosure quantity in their SEC filings (Scott 1994; Harris 1998; Botosan and Stanford 2005), with mixed evidence with regards to the direction of the association. In the case of R&D disclosures, James and Shaver (2014) find that firms having a technological advantage increase the quantity of such disclosures in order to deter R&D competition. In the current context, managers' strategic disclosure behavior in response to greater competition could help explain the negative bias, only if the interaction of R&D disclosure quantity with competition loads significantly in the direction of increase ascertained by the type of competition measure used.

The most popular and widely used proxy for competition is industry concentration, measured using the Herfindahl-Hirschman Index (HHI). Higher (lower) values of HHI indicate greater (lesser) industry concentration and thus lesser (greater) competition. (The definition and construction of HHI is outlined in Appendix A1.) However, industry concentration (or HHI) as a measure of competition has been criticized by prior studies since the relation between concentration and competition is not clear (especially when market structure is not exogenous), and this measure fails to capture several important dimensions of competition, including product substitutability, market size, and entry costs (Raith 2003; Karuna 2007). In view of this, we also examine the effect of

competition by including two measures of competition from Karuna (2007) in two separate regressions (after excluding industry effects) – the price-cost margin (*PC_MARGIN*) which captures product substitutability, and market size (*MKT_SIZE*). Lower price-cost margin (or greater product substitutability), and greater market size reflect greater price competition (Karuna 2007). The definition and measurement of these variables is outlined in Appendix A1.

Panel A of Table 6 presents the summary statistics for the competition measures, while Panel B contains the regression results. Since any measure of competition is likely to be highly correlated with the industry dummies, we ran all regressions in this table after omitting the industry effects. The standard errors were clustered at both firm and year levels. As shown in Panel B, all the three interaction terms involving the three measures of competition are statistically insignificant. Therefore, it appears that competitive pressures facing the firm do not lead managers to disclose R&D strategically in our sample. But more importantly, the negative association between *ln (R&D QTY)* and future ROA remains highly significant across all the specifications.

Hence, the two strategy-based explanations we propose and test for fail to explain the association between R&D disclosure quantity and future ROA. Our main (negative) coefficient is both statistically and economically significant even after accounting for firm strategic motives. Note, however, that the role of other strategy-based explanations in this context cannot be ruled out. We now extend a psychology-based explanation in our attempt to unravel the negative association.

It appears that the worst R&D performers are also the most biased, leading to the observed negative correlation. If the bias in R&D disclosures were to be uncorrelated

with R&D performance, we should have found no association in the first place. Even on its own, the argument deserves merit as Kruger and Dunning (1999) have shown that it is the poorest performers who hold the least accurate assessments of their skill and performances, marked in turn by a gross overestimation of their performance relative to that of their peers. This happens because their lack of skills not only makes them amenable to committing many more mistakes, but it also deprives them of the metacognition of recognizing the incorrectness of a particular decision. Intuitively, this line of reasoning should hold even in the context of R&D, where the worst performers may not be capable to accurately judge the future scope and viability of an R&D investment decision.

Additional analysis and robustness checks

First, we decompose the R&D disclosures (identified at the sentence level) into forward-looking (FLS) and non-forward-looking (non-FLS), and then examine the association of these two individual components with future ROA in model (1). If we are arguing that managers cannot skillfully predict the *future* R&D outcome (given the typical features of R&D environment), then the bias (earlier negative association: worst performers also the most biased) should only be driven by FLS (which pertains to the future) and not non-FLS.

Following Merkley (2014), we first count the number of R&D-related sentences in each firm's annual filing, identified using the R&D dictionary described earlier in the paper. Next, we classify these R&D sentences into FLS and non-FLS using the dictionary of future-oriented phrases and keywords from Muslu et al. (2015). Finally, we compute a measure of forward-looking (non-forward-looking) R&D disclosure quantity by taking a

logarithmic transformation of the count of R&D-related FLS (non-FLS) in the 10-K. All variable definitions and measurement are in Appendix A1.

Table 7 presents the summary statistics and regression results. As shown in Panel A, the average firm in our sample discloses 10.94 R&D-related sentences, of which 1.79 are FLS and 9.11 are non-FLS. Panel B contains the regression results. It can be observed that only the FLS component is significant and negatively associated with future ROA, which confirms our assertion.

Second, we rerun model (1) by focusing only on firms from the six largest and most R&D-intensive industries (Gu and Li 2003; Hirshleifer et al. 2013), with two-digit SIC of 28 (chemicals, biotech and pharmaceuticals), 35 (computer hardware and machinery), 36 (electrical and electronics), 37 (transportation equipment), 38 (medical and scientific instruments), and 73 (computer software and data services). More than 80 per cent of the total R&D expenditure comes from these six industries, which justifies our choice of the six R&D-intensive industries. Moreover, we adopt an industry-wise term-weighting scheme for R&D intensive industries where the words could be strongly linked to the language of specific industry segments (Loughran and McDonald 2011). The regression results are reported in Table 8. It can be observed that the negative association between R&D disclosure quantity and future ROA is even stronger for the R&D-intensive firms. In terms of economic significance, a one standard deviation change in R&D disclosure quantity leads to a 46.75 per cent decline in ROA for an average R&D-intensive firm.

Finally, we run a host of robustness and sensitivity checks. In unreported results, we show that our main results are robust to using an unweighted word-count-based

measure of R&D disclosure quantity, and to cash flow from operating activities (CFO) as an alternate firm profitability measure. Specifically, in the case of the latter, an increase from the 25th to the 75th percentile in R&D disclosure quantity results in a decrease in CFO (scaled by average total assets) of the magnitude of 0.00164, which corresponds to a 2.05 per cent decline in CFO for an average firm in our sample. Also, note that the innovative efficiency (IE) measure(s) from Hirshleifer et al. (2013), described earlier and in Appendix A1, is an important control in all our main specifications. To compute the patent-based IE measure, we obtain the relevant data from the NBER database that ends in 2006. However, Kogan et al. (2014)'s database contains information on patents through 2010. Therefore, as a robustness check, we re-conduct our main test using patent and citation data from this extended patent database. Our findings using the Kogan et al. (2014) patent database reconfirm the negative association between R&D disclosure quantity and future ROA (Table 9). Furthermore, since the average time lag between the patent filing (or application) date and its issue (or grant) date is two years, Hall et al. (2001) argue in favor of using the patent filing date as opposed to issue date; rationale being that the particular innovation was materialized in the year of application, thus making it a better indicator of firm patenting activity. Since we have used the patent issue date throughout the paper, following Hirshleifer et al. (2013), we rerun model (1) after computing the IE measure using the patent filing date instead. The results, reported in column (1) of Table 9, reconfirm our earlier finding of a significant negative association.

5. CONCLUSION

In this paper, we examine the association between the quantity of R&D-related narrative disclosures in a firm's 10-K filing and future performance (ROA), and offer new insights on the role, importance, and credibility of narrative disclosures and the R&D disclosure process. We chose the R&D setting due to its typical characteristics and the important role of R&D in the creation of future firm value and growth.

The empirical findings of the paper can be easily summarized. We obtain a persistent and significant (statistical and economic) association between R&D disclosure quantity and future profitability. This negative bias cannot be explained by strategy-based explanations in this context, namely, the manager's overall strategic disposition as gauged from the 10-K complexity, and competitive pressures facing the firm. In a bid to explain the bias, we then offer a psychology-based explanation. Specifically, we argue that the worst R&D performers are also the most biased, which is not surprising as prior literature in social psychology documents that it is the poorest performers who hold the least accurate assessments of their skills and thus end up overestimating their performance relative to that of their peers.

Taken together, the evidence in this paper suggests that a firm's narrative disclosures may not always be meaningful for analyzing current and future firm fundamentals, and the type of a disclosure and features of its environment are important considerations in this regard. Future research on narrative disclosures should take cognizance of this result.

APPENDIX A (CHAPTER 1): VARIABLE DEFINITIONS AND MEASUREMENT

Variable	Notation	Definition/Measurement
Tone of the 10-K document	10K TONE	$\frac{(\text{Weighted count of positive words in 10K} - \text{Weighted count of negative words in 10K})}{\text{Weighted count of all words in 10K}}$
Earnings	EARN	Earnings before extraordinary items/lagged total assets
Annual returns	RET	Buy-and-hold annual returns at the fiscal year-end
Size	SIZE	Log(market capitalization), where market capitalization is defined as common shares outstanding multiplied by annual closing price
Book-to-Market ratio	BTM	Book value of equity divided by the market value of equity at the fiscal year-end
Return volatility	RET VOL	Standard deviation of monthly returns over the fiscal year
Earnings volatility	EARN VOL	Standard deviation of EARN over the last five years
Business segments	BUS SEG	Number of business segments of a firm
Geographical segments	GEOG SEG	Number of geographical segments of a firm
Firm Age	AGE	Log[1 + (number of years since a firm appears in CRSP monthly file)]
10-K abnormal tone	AB_TONE	Residual term from annual cross-sectional regressions of 10K TONE on its determinants (model 1)
CEO's tenure	CEO Tenure	Number of years the CEO has been in that position (if missing, then number of years at the firm) as of the fiscal year-end; computed using the "BECAMECEO" and "JOINED_CO" variables in *Execucomp*
CEO's early years of tenure	CEO Early Years	A dummy variable equal to 1 for firm-years corresponding to the first three years of a CEO's tenure, and 0 otherwise
CEO's final year of tenure	CEO Final Year	A dummy variable equal to 1 for the year prior to the CEO's turnover year, and 0 otherwise (turnover year identified using the "LEFTOFC" variable in *Execucomp*, and observations where LEFTOFC was missing were dropped)
High-litigation firms	HighLitigation	A dummy variable equal to 1 for firms operating in a high-litigation industry (SIC codes 2833–2836; 3570–3577; 3600–3674; 5200–5961; 7370–7374), and 0 otherwise

Net operating assets	NOA	(Operating assets (OA) – Operating liabilities (OL))/ lagged total assets, where OA= total assets – cash and short term investment, and OL = total assets – debt included in current liabilities – long term debt – minority interests – preferred stocks – common equity [using Hirshleifer et al. (JAE 2004)]. Missing values of long-term debt, minority interest, or preferred stock are coded as zero.
Options-based measure of CEO optimism	CEO Optimism	A dummy variable equal to 1 if the CEO holds (vested) stock options that are at least 67% in-the-money* at least twice during the sample period, and 0 otherwise. [CEOs thus identified as optimistic remain so for the rest of the sample period, beginning with the first time they exhibited this behavior] *Average moneyness of the CEO's option portfolio is calculated as (Campbell et al., JFE 2011): Average moneyness = stock price/ strike price – 1, where strike price = fiscal year end stock price - average realizable value, and average realizable value (for each CEO-year) = total realizable options value/ number of options held by the CEO.
Cash flows	CFO	Operating cash flow/lagged total assets
Accruals	ACC	EARN - CFO
Managerial ability	ABILITY SCORE	Residual-based measure of managerial (CEO) ability from Demerjian et al. (2012)
10-K sentence count	10K Sentences	Total number of sentences in the 10-K document
10-K forward-looking sentence count	10K FLS	Count of forward-looking sentences (FLS) in the 10-K document, where FLS were identified using the methodology in Muslu et al. (Management Science, 2015)
10-K non-forward-looking sentence count	10K Non-FLS	Count of non-forward-looking sentences (Non-FLS) in the 10-K document; calculated as total sentences minus FLS
Forward-looking intensity	FLI	Number of forward-looking sentences in 10-K/total number of sentences in 10-K
Analyst following	ANALYST	Log [1 + (number of analysts following the firm)]
Loss	LOSS	A dummy variable equal to 1 if EARN is positive and 0 otherwise
Special items	SI	Sum of unusual or nonrecurring income statement items (SPI)/lagged total assets. Missing values of SPI were coded to zero when SPI was missing but total assets was nonmissing.
Merger-and-acquisition	MA	A dummy variable equal to 1 if aqs/sales>1% or if aqc/total assets>2% and 0 otherwise, where "aqs" is the acquisition-sale contribution and "aqc" is the value of acquisitions. Missing values of aqc and aqs were coded as zero.
Discretionary accruals	DA	Discretionary accruals calculated using the modified Jones model
Investment-based measure of CEO optimism (Capital Expenditures)	CEO Optimism_ CAPX	A dummy variable equal to 1 if capital expenditures divided by lagged total assets is greater than its industry median value, and 0 otherwise
Investment-based measure of CEO optimism (Excess investment)	CEO Optimism_ XSINVEST	A dummy variable equal to 1 if the residual of a regression of total asset growth on sales growth (run at the industry-year level) is greater than zero, and 0 otherwise
Count of words in 10-K	10K Wordcount	Total number of words in the 10-K document

Fog index (Li, JAE 2008)	FOG	FOG = 0.40*(Average Number of Words per Sentence + Average Number of Complex Words per Sentence)
Length of 10-K	LENGTH	Log(10K Wordcount)
10-K file size	FILESIZE	The file size of the 10-K complete submission text file (in megabytes)
Seasoned equity offering	SEO	A dummy variable equal to 1 if a firm has seasoned equity offering in the current year according to the *SDC Global New Issues* database and 0 otherwise
Delaware	DLW	A dummy variable equal to 1 if a company is incorporated in Delaware and 0 otherwise

APPENDIX B (CHAPTER 1): STEPS IN PARSING THE 10-K DOCUMENTS

All 10-K documents filed between the years 1993 to year 2010 were downloaded from the SEC EDGAR. Each document obtained from EDGAR contained a lot of information, including graphics/jpg segments. These 10-K documents were parsed based on the following algorithm using Python scripts:

* Graphic/jpg/xls segments of each document were removed.

* The text part of each document was extracted after removing various HTML tags.

* The words and sentences from each document were extracted as follows:

PART A: EXTRACTION OF WORDS

* Each word in the text part of the document was compared with an English dictionary (Loughran and McDonald 2011), and phrases that are not actual words defined in the dictionary (e.g. proper nouns) were removed.

* Each word was reduced to its ‘stem’, so that different forms of the same word are considered as one word.

* The text part of each document was represented into a vector of words.

* Positive and negative words [identified based on the financial sentiment dictionary provided by Loughran and McDonald (2011)] were counted.

* The weight for each word was calculated using the term-weighting scheme described in Section II.

PART B: EXTRACTION OF SENTENCES

* The text part of each document was represented into a vector of sentences.

* Forward-looking sentences were identified using the dictionary from Muslu et al. (2015) and counted.

APPENDIX A1 (CHAPTER 2): VARIABLE DEFINITIONS AND MEASUREMENT

Variable	Notation	Definition/Measurement
R&D disclosure quantity	ln (R&D QTY)	$\ln(1 + (\frac{\text{Weighted count of R\&D related words in } 10-\text{K}}{\text{Weighted count of all words in } 10-\text{K}}) * 1000)$
R&D disclosure quantity for R&D Intensive firm	ln (R&D QTY) $_{\text{R\&D INTENSIVE}}$	$\ln(1 + (\frac{\text{Weighted count of R\&D related words in } 10-\text{K}}{\text{Weighted count of all words in } 10-\text{K}}) * 1000)$ using industry wise term weighting for R&D intensive firms
R&D Capital	RDC	$\text{RD}_{it} + 0.8 * \text{RD}_{it-1} + 0.6 * \text{RD}_{it-2} + 0.4 * \text{RD}_{it-3} + 0.2 * \text{RD}_{it-4}$
Innovative Efficiency	ln (1+ Patents/RDC)	Patents granted in year t / RDC(t-2)
Length of the 10-K document	10K Length	ln (total words in 10-K document)
Tone of the forward looking disclosures	FLS Tone	Weighted count of optimistic words in 10-K forward-looking disclosures subtracted by the weighted count of pessimistic words in it, divided by weighted count of all words in 10-K forward-looking disclosures
Advertising Expenditures	ln (1+AD/Asset)	ln (1+Advertising Expenditures/Average Total Assets)
Capital Expenditures	ln (1+Capex/Asset)	ln (1+Capital Expenditures/Average Total Assets)
Size	ln (Asset)	ln (Asset)
Operating Performance	ROA	Return on Asset (ROA): Income before extra-ordinary items plus interest expenses divided by average total assets
R&D Intensity	ln (1+ R&D Exp/Asset)	ln (1+R&D Expenditures/Average Total Assets)
R&D Growth	R&D Growth Dummy	For firms that have (as of the beginning of their R&D increase year) an R&D intensity (i.e., the ratios of R&D to assets and R&D to sales) of at least 5 percent, it is equal to 1 when firm increase its dollar R&D by at least 5 percent, and increase its ratio of R&D to assets by at least 5 percent (e.g., from 10 percent to 10.5 percent).

		Otherwise, it is equal to 0.
Adjusted Patent Citation	APC	Citations in year t scaled by total assets averaged over years t-1 and t-2
R&D Pessimism	R&D Pessimism	Weighted count of pessimistic words in R&D related disclosure divided by weighted count of all words in R&D related disclosure.
R&D Uncertainty	R&D Uncertainty	Weighted count of words involving uncertainty in R&D related disclosure divided by weighted count of all words in R&D related disclosure.
Readability	Fog Index	FOG = 0.40 X (Average Number of Words per Sentence + Average Number of Complex Words per Sentence) (Li, 2008)
Herfindahl-Hirschman Index	HHI	Sum of squared market shares, where market share of an individual firm is calculated by using firm's net sales divided by the total sales value of the whole industry
Price-cost margin	PC_MARGIN	Sales/operating costs, for each industrial segment; where operating costs include cost of goods sold, selling, general, and administrative expense, and depreciation, depletion, and amortization.
Market size	MKT_SIZE	Natural log of industry sales
R&D related forward looking disclosure	ln (FLS)	ln (1+ R&D-related forward-looking sentences)
R&D related non-forward looking disclosure	ln (Non-FLS)	ln (1+ R&D-related non-forward-looking sentences)

APPENDIX B1 (CHAPTER 2): STEPS IN PARSING THE 10-K DOCUMENTS

We download all 10-K documents from EDGAR (SEC) filed between year 1994 to year 2011. Each document obtained from EDGAR contains a lot of information which includes graphics/jpg/xls segments.

We parse these 10-K documents based on the following algorithm using Python scripts:

- We remove graphic/jpg/xls segments of the document.
- We extract text part of each document after removing various HTML tags.
- We use following steps to extract words and sentences respectively.

PART A: STEPS FOR WORDS EXTRACTION

- We compare each word in the text part of the document with an English dictionary [Loughran and McDonald, 2011] and remove phrases which are not actual words defined in the dictionary (e.g. proper nouns).
- We reduce each word to its ‘stem’, so that different forms of the same word are considered as one word.
- We represent the text part of each document into a vector of words.
- We count the occurrence of R&D-related words (Based on R&D dictionary [Merkley, 2014] modified for counting words).
- We count the occurrence of positive/negative/uncertain words using the financial dictionary provided by Loughran and McDonald (2011).
- We calculate weight for each word using the term-weighting scheme described under Empirical Methodology (Section 2).

PART B: STEPS FOR SENTENCES EXTRACTION

- We represent the text part of each document into a vector of sentences.
- We extract R&D related sentences using the R&D dictionary provided by Merkley (2014).
- We categorize all words present in R&D related sentences as positive/negative/uncertain/none using the financial dictionary provided by Loughran and McDonald (2011).
- We identify forward-looking R&D sentences using the forward-looking dictionary from Muslu et al. (2015).

REFERENCES

Ahmed, A. S., & Duellman, S. 2013. Managerial overconfidence and accounting conservatism. *Journal of Accounting Research*, *51*(1), 1-30.

Ali, A., & Zhang, W. 2015. CEO tenure and earnings management. *Journal of Accounting and Economics*, *59*(1), 60-79.

Arkes, H.R., 2001. Overconfidence in judgmental forecasting. In *Principles of forecasting* (pp. 495-515). Springer US.

Audi, R., Loughran, T., & McDonald, B. 2014. Trust, But Verify: MD&A Language and the Role of Trust in Corporate Culture. *Available at SSRN 2480766*.

Barton, J., & Simko, P. J. 2002. The balance sheet as an earnings management constraint. *The accounting review*, *77*(s-1), 1-27.

Berger, P. G., Ofek, E., Yermack, D. L. 1997. Managerial entrenchment and capital structure decisions. Journal of Finance 4, 1411-1438.
Bhagat, Sanjai, John M. Bizjak, and Jeffrey L. Coles. 1998. The Shareholder Wealth Implications of Corporate Lawsuits. *Financial Management* 27:5–27.

Bhattacharya, Utpal, Neal Galpin, and Bruce Haslem. 2007. The Home Court Advantage in International Corporate Litigation. *Journal of Law and Economics* 50:625–659.

Black, Bernard, Brian R. Cheffins, and Michael Klausner. 2006. Outside Director Liability. *Stanford Law Review* 58:1055–1160.

Bodnaruk, Andriy, Tim Loughran, and Bill McDonald. 2013. Using 10-K text to gauge financial constraints. *Social Science Research Network*.

Botosan, C. A., & Stanford, M. 2005. Managers' motives to withhold segment disclosures and the effect of SFAS No. 131 on analysts' information environment. *The Accounting Review*, *80*(3), 751-772.

Brickley, J. A., Linck, J. S., & Coles, J. L. 1999. What happens to CEOs after they retire? New evidence on career concerns, horizon problems, and CEO incentives. *Journal of Financial Economics*, *52*(3), 341-377.

Brown, S., and J. Tucker. 2010. Large-sample evidence on firms' year-over-year MD&A modifications. *Journal of Accounting Research* 49 (2): 309–346.

Buckley, C. (1993, March). The importance of proper weighting methods. In *Proceedings of the workshop on Human Language Technology* (pp. 349-352). Association for Computational Linguistics.

Campbell, T. C., Gallmeyer, M., Johnson, S. A., Rutherford, J., & Stanley, B. W. 2011. CEO optimism and forced turnover. *Journal of Financial Economics*,*101*(3), 695-712.

Chen, H., De, P., Hu, Y. J., & Hwang, B. H. 2014. Wisdom of crowds: The value of stock opinions transmitted through social media. *Review of Financial Studies*, *27*(5), 1367-1403.

Coles, J. L., Daniel, N. D., Naveen, L., 2008. Boards: does one size fit all? Journal of Financial Economics 87, 329–356.

Das, S., Kim, K., & Patro, S. 2011. An analysis of managerial use and market consequences of earnings management and expectation management. *The Accounting Review*, *86*(6), 1935-1967.

Davis, A. K., & TAMA-SWEET, I. S. H. O. 2012. Managers' Use of Language Across Alternative Disclosure Outlets: Earnings Press Releases versus MD&A*. *Contemporary Accounting Research*, *29*(3), 804-837.
Davis, A. K., Ge, W., Matsumoto, D., & Zhang, J. L. 2014. The effect of manager-specific optimism on the tone of earnings conference calls. *Review of Accounting Studies*, *20*(2), 639-673.

DeAngelo, L. E. 1988. Managerial competition, information costs, and corporate governance: The use of accounting performance measures in proxy contests. *Journal of Accounting and Economics*, *10*(1), 3-36.

Demerjian, P., Lev, B., & McVay, S. 2012. Quantifying managerial ability: A new measure and validity tests. *Management Science*, *58*(7), 1229-1248.

Eberhart, A., Maxwell, W. and Siddique, A., 2008. A Reexamination of the Tradeoff between the Future Benefit and Riskiness of R&D Increases.*Journal of Accounting Research*, *46*(1), pp.27-52.

Eberhart, A.C., Maxwell, W.F. and Siddique, A.R., 2004. An examination of long-term abnormal stock returns and operating performance following R&D increases. *The Journal of Finance*, *59*(2), pp.623-650.

Elliott, J. A., & Shaw, W. H. 1988. Write-offs as accounting procedures to manage perceptions. *Journal of Accounting Research*, 91-119.

Entwistle, G. 1999. Exploring the R&D disclosure environment. *Accounting Horizons* 13 (4): 323–341.

Fama, E. F. 1980. Agency Problems and the Theory of the Firm. *The journal of political economy*, 288-307.

Fama, E. F., & French, K. R. 2000. Forecasting Profitability and Earnings.*The Journal of Business*, *73*(2), 161-175.

Feldman, R., Govindaraj, S., Livnat, J., & Segal, B. 2010. Management's tone change, post earnings announcement drift and accruals. *Review of Accounting Studies*, *15*(4), 915-953.

Francis, J., Philbrick, D., & Schipper, K. 1994. Shareholder litigation and corporate disclosures. *Journal of accounting research*, 137-164.

Gande, Amar, and Craig M. Lewis. 2009. Shareholder-Initiated Class Action Lawsuits: Shareholder Wealth Effects and Industry Spillovers. *Journal of Financial and Quantitative Analysis* 44:823–850.

Gibbons, R., Murphy, K.J. 1992. Optimal incentive contracts in the presence of career concerns: theory and evidence. Journal of Political Economy 100, 468–505.
Godfrey, J., Mather, P., & Ramsay, A. 2003. Earnings and impression management in financial reports: the case of CEO changes. *Abacus*, *39*(1), 95-123.

Graham, J. R., Harvey, C. R., & Rajgopal, S. 2005. The economic implications of corporate financial reporting. *Journal of accounting and economics*, *40*(1), 3-73.

Grandi, A., B. H. Hall, and R. Oriani. 2009. R&D and financial investors. In Evaluation and Performance Measurement of Research and Development, edited by V. Chiesa, and F. Frattini, 143–165.Cheltenham, U.K.: Edward Elgar.

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.

Gu, F., & Li, J. Q. 2003. Disclosure of innovation activities by high-technology firms. *Asia-Pacific Journal of Accounting & Economics*, *10*(2), 143-172.

Gu, F., 2005. Innovation, future earnings, and market efficiency. *Journal of Accounting Auditing and Finance* 20, 385–418.

Hall, B. H., Jaffe, A. B., & Trajtenberg, M. 2001. *The NBER patent citation data file: Lessons, insights and methodological tools* (No. w8498). National Bureau of Economic Research.

Harris, Mary Stanford. 1998. "The association between competition and managers' business segment reporting decisions." *Journal of accounting research*: 111-128.

Haslem, Bruce. 2005. Managerial Opportunism during Corporate Litigation. *Journal of Finance* 60:2013–2041.

Henry, E. 2008. Are investors influenced by how earnings press releases are written? *Journal of Business Communication* 45 (4): 363–407.

Hermalin, B. E., & Weisbach, M. S. 1998. Endogenously chosen boards of directors and their monitoring of the CEO. *American Economic Review*, 96-118.

Hertzel, Michael G., and Janet Kiholm Smith. 1993. Industry Effects of Interfirm Lawsuits: Evidence from Pennzoil v. Texaco. *Journal of Law, Economics, and Organization* 9:425–444.

Hirshleifer, D., Hou, K., Teoh, S. H., & Zhang, Y. 2004. Do investors overvalue firms with bloated balance sheets? *Journal of Accounting and Economics*, *38*, 297-331.

Hirshleifer, D., Hsu, P. H., & Li, D. 2013. Innovative efficiency and stock returns. *Journal of Financial Economics*, *107*(3), 632-654.

Huang, X., Teoh, S. H., & Zhang, Y. 2014. Tone management. *The Accounting Review*, *89*(3), 1083-1113.

James, S. D., & Shaver, J. M. 2014. *Motivations for voluntary public R&D disclosures*. Working paper, Fisher College of Business, The Ohio State University.

Johnson, Marilyn F., Karen K. Nelson, and A. C. Pritchard. 2000. In Re Silicon Graphics Inc.: Shareholder Wealth Effects Resulting from the Interpretation of the Private Securities Litigation Reform Act's Pleading Standard. *Southern California Law Review* 73:773–810.

Jones, D. A. 2007. Voluntary disclosure in R&D-intensive industries. *Contemporary Accounting Research* 24 (2): 489–522.

Jurafsky, Daniel and James H. Martin, 2009, Speech and Language Processing, Prentice Hall, New Jersey.

Kahneman, Daniel, and Gary Klein. "Conditions for intuitive expertise: a failure to disagree." American Psychologist 64.6 (2009): 515.

Karuna, C. 2007. Industry product market competition and managerial incentives. *Journal of accounting and economics*, *43*(2), 275-297.

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

Kogan, L., Papanikolaou, D., Seru, A., & Stoffman, N. 2014. *Technological innovation, resource allocation, and growth* (No. w17769). National Bureau of Economic Research.

Kothari, S. P., Shu, S., & Wysocki, P. D. 2009. Do managers withhold bad news?. *Journal of Accounting Research*, *47*(1), 241-276.

Kruger, J., & Dunning, D. 1999. Unskilled and unaware of it: how difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of personality and social psychology*, *77*(6), 1121.

Lang, M. H., & Lundholm, R. J. 1996. Corporate disclosure policy and analyst behavior. *Accounting review*, 467-492.

Lang, M. H., and R. J. Lundholm. 2000. Voluntary disclosure and equity offerings: Reducing information asymmetry or hyping the stock? *Contemporary Accounting Research* 17 (4): 623–662.

Law, K. K. F. and Mills, L. F. 2015. Taxes and Financial Constraints: Evidence from Linguistic Cues. *Journal of Accounting Research*.

Leder, M. 2003. *Financial Fine Print: Uncovering a Company's True Value*. John Wiley & Sons.

Lehavy, R., K. Merkley, and F. Li. 2011. The effect of annual report readability on analyst following and the properties of their earnings forecasts. *The Accounting Review* 86 (3): 1087–1115.

Lerner, J., 1994. The importance of patent scope: an empirical analysis. *RAND Journal of Economics* 25, 319–333.

Lev, B. and T. Sougiannis, 1996. The capitalization, amortization, and value-relevance of R&D. *Journal of Accounting & Economics* 21: 107-138.

Lev, B., 2001. Intangibles: Management, Measurement, and Reporting. The Brookings Institution, Washington, DC.

Levy, H., Shalev, R., & Zur, E. 2015. The Effect of CFO Personal Litigation Risk on Firms' Disclosure and Accounting Choices. Working paper, Zicklin School of Business, Stern School of Business, Smith School of Business.

Li, F. 2008. Annual report readability, current earnings, and earnings persistence. *Journal of Accounting and economics*, *45*(2), 221-247.

Li, F. 2010. Textual Analysis of Corporate Disclosures: A Survey of the Literature. *Journal of accounting literature*, *29*, 143-165.

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

Li, F., Minnis, M., Nagar, V., & Rajan, M. 2014. Knowledge, compensation, and firm value: An empirical analysis of firm communication. *Journal of Accounting and Economics*, *58*(1), 96-116.

Li, Feng, Russell Lundholm, and Michael Minnis. 2013. A Measure of Competition Based on 10-K Filings. *Journal of Accounting Research* 51.2: 399-436.

Loughran, T., & McDonald, B. 2014. Measuring readability in financial disclosures. *The Journal of Finance*, *69*(4), 1643-1671.

Loughran, Tim, and Bill McDonald. 2011. When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks. *The Journal of Finance* 66.1: 35-65.

Mather, P., & Ramsay, A. 2007. Do board characteristics influence impression management through graph selectivity around CEO changes? *Australian Accounting Review*, *17*(42), 84-95.

Matolcsy, Z.P. and Wyatt, A., 2008. The association between technological conditions and the market value of equity. *The Accounting Review*, *83*(2), pp.479-518.

Meehl, P.E., 1973. *Psychodiagnosis: selected papers*. U of Minnesota Press.

Merkl-Davies, D. M. & Brennan, N. M. 2007. Discretionary disclosure strategies in corporate narratives: incremental information or impression management?. *Journal of accounting literature*, *27*, 116-196.

Merkley, K. J. 2014. Narrative Disclosure and Earnings Performance: Evidence from R&D Disclosures. *The Accounting Review*, *89*(2), 725-757.

Morricone, S., and R. Oriani. 2009. R&D information. In Evaluation and Performance Measurement of Research and Development, edited by V. Chiesa, and F. Frattini, 166–188. Cheltenham, U.K.:Edward Elgar.

Muslu, V., Radhakrishnan, S., Subramanyam, K. R., & Lim, D. 2015. Forward-looking MD&A disclosures and the information environment. *Management Science*, *61*(5), 931-948.

Nelson, K. K., & Pritchard, A. C. 2007. Litigation risk and voluntary disclosure: The use of meaningful cautionary language. In *2nd Annual Conference on Empirical Legal Studies Paper*.

Nickell, S. 1981. Biases in dynamic models with fixed effects. *Econometrica: Journal of the Econometric Society*, 1417-1426.

Nini, G., Smith, D. C., & Sufi, A. 2012. Creditor control rights, corporate governance, and firm value. *Review of Financial Studies*, *25*(6), 1713-1761.

Pandit, S., Wasley, C. E., & Zach, T. 2011. The effect of research and development (R&D) inputs and outputs on the relation between the uncertainty of future operating performance and R&D expenditures. *Journal of Accounting, Auditing & Finance*, 26, 121-144.

Pourciau, S. 1993. Earnings management and nonroutine executive changes.*Journal of Accounting and Economics*, *16*(1), 317-336.

Previts, G. J., Bricker, R. J., Robinson, T. R., & Young, S. J. 1994. A content analysis of sell-side financial analyst company reports. *Accounting Horizons*,*8*(2), 55.

Rahman, S. 2012. Impression management motivations, strategies and disclosure credibility of corporate narratives. *Journal of Management Research*,*4*(3), 1-14.

Raith, M., 2003. Competition, risk and managerial incentives. *American Economic Review* 93 (4), 1425–1436.

Rogers, J. L., Van Buskirk, A., & Zechman, S. L. 2011. Disclosure tone and shareholder litigation. *The Accounting Review*, *86*(6), 2155-2183.

Romano, Roberta. 1991. The Shareholder Suit: Litigation without Foundation? *Journal of Law, Economics & Organization* 7:55–87.

Rutherford, B.A. 2003. Obfuscation, textual complexity and the role of regulated narrative accounting disclosure in corporate governance. *Journal of Management and Governance* 7: 187-210.

Ryan, H. E., Wang, L., Wiggins, R. A. 2009. Board-of-director monitoring and CEO tenure. Working Paper, Georgia State University, Bentley College.

Schrand, C. M., & Zechman, S. L. 2012. Executive overconfidence and the slippery slope to financial misreporting. *Journal of Accounting and Economics*,*53*(1), 311-329.

Scott, T. W. 1994. Incentives and disincentives for financial disclosure: Voluntary disclosure of defined benefit pension plan information by Canadian firms. *Accounting Review*, 26-43.

Skinner, D. J. 1994. Why firms voluntarily disclose bad news. *Journal of accounting research*, 38-60.

Skinner, D. J. 1997. Earnings disclosures and stockholder lawsuits. *Journal of Accounting and Economics*, *23*(3), 249-282.

Sloan, R. 1996. Do stock prices fully reflect information in accruals and cash flows about future earnings? *Accounting review*, *71*(3), 289-315.

Vancil, R. F. 1987. *Passing the baton: Managing the process of CEO succession*. Harvard Business School Pr.

Weetman, P., & Collins, B. 1996. *Operating and financial review: experiences and exploration*. Institute of Chartered Accountants of Scotland.

Wells, P. A. 2002. Earnings management surrounding CEO changes. *Accounting and Finance*, *42*, 169-193.

Wier, Peggy. 1983. The Costs of Antimerger Lawsuits: Evidence from the Stock Market. *Journal of Financial Economics* 11:207-224.

TABLE 1: Estimation of 10-K Abnormal Tone

The table in Panel B reports the coefficient estimates of a Fama-MacBeth regression of *10K TONE* on its determinants. Panels A and C report the summary statistics of *10K TONE* and *AB_TONE*, respectively. All variable definitions are outlined in Appendix A. t-statistics are reported in brackets in the table in Panel B. ***, **, * represents statistical significance at the 1%, 5% and 10% levels.

Panel A: Summary Statistics - 10K TONE

Variable	n	Mean	Median	S.D.	Min	Max
10K TONE	56820	-37.11	-33.46	27.34	-135.18	25.37

Panel B: Expected Tone Model

Dep. variable	10K Tone
EARN	2.903
	[0.830]
RET	2.009***
	[5.478]
SIZE	-0.718
	[-1.601]
BTM	-1.128**
	[-2.446]
RET VOL	-8.598***
	[-7.454]
EARN VOL	-1.674
	[-0.847]
ln BUS SEG	1.678***
	[2.967]
ln GEOG SEG	3.568**
	[2.581]
AGE	-1.919***
	[-6.890]
Constant	-23.944***
	[-11.721]
Observations	46,756
No. of groups	18
R-squared	0.008

Panel C: Summary Statistics – AB_TONE

Variable	n	Mean	Median	S.D.	Min	Max
AB_TONE	46756	0	3.4	26.73	-110.71	86.15

TABLE 2: Descriptive Statistics

Panel A provides summary statistics for all the variables with nonmissing firm-year observations from 1993-2010. All variable definitions are outlined in Appendix A. Panel B provides the Spearman correlations between the variables of interest, where * represents statistical significance at or below the 5% level.

Panel A: Summary Statistics

Variable	n	Mean	Median	S.D.	Min	Max
CEO Tenure	18576	7.7	5.25	7.34	0.5	36.11
CEO Early Years	18576	0.31	0.00	0.46	0.00	1.00
CEO Final Year	11694	0.18	0.00	0.39	0.00	1.00
CEO Optimism	18576	0.31	0.00	0.46	0.00	1.00
HighLitigation	18576	0.33	0.00	0.47	0.00	1.00
NOA	15633	0.61	0.62	0.31	-0.89	2.05
ACC	18560	-0.07	-0.06	0.11	-1.38	0.42
ABILITY SCORE	17131	0.01	0.00	0.14	-0.41	0.52
EARN	18563	0.04	0.05	0.17	-2.86	0.44
EARN VOL	18319	0.08	0.04	0.21	0.00	5.87
CFO	14625	0.11	0.11	0.13	-3.24	1.15
SIZE	18574	7.24	7.17	1.64	-0.11	10.67
BTM	18574	0.49	0.43	0.77	-7.3	4.57
AGE (non-log)	18448	21.03	19	11.72	1.00	37
RET	17846	0.16	0.09	0.59	-0.9	3.78
RET VOL	17846	0.42	0.36	0.26	0.12	2.03
BUS SEG	18576	1.52	1.00	0.94	1.00	10
GEOG SEG	18576	1.01	1.00	0.08	1.00	3
10K Sentences	18576	1833.19	1546	1137.74	408	6926
10K FLS	18576	216.8	186.5	139.48	31	809
10K Non-FLS	18576	1615.44	1355	1015.26	362	6208
FLI	18576	0.12	0.12	0.03	0.01	0.35

Panel B: Spearman Correlation Matrix

AB_TONE	1						
CEO Early Years	-0.0756*	1					
CEO Final Year	-0.0782*	-0.0432*	1				
HighLitigation	-0.0579*	0.0332*	0.0226*	1			
NOA	0.1460*	-0.0934*	-0.0756*	-0.2057*	1		
ACC	0.0419*	-0.0347*	-0.0418*	-0.1456*	0.1454*	1	
CEO Optimism	0.1003*	-0.2179*	-0.0425*	0.0203	0.0882*	0.0325*	1

TABLE 3: Early years of CEOs' tenure and tone management

This table reports the coefficient estimates of a regression of 10-K abnormal tone (*AB_TONE*) on the *CEO Early Years* variable and a set of controls. The sample period is from 1993 to 2010. All variable definitions are outlined in Appendix A. Model specification (1) includes both year and firm fixed effects, while specifications (2) and (3) only include year fixed effects. t-statistics (in brackets) are based on standard errors that are clustered at the firm level. ***, **, * represents statistical significance at the 1%, 5% and 10% levels.

Model	(1)	(2)	(3)
Dependent variable	AB_TONE	AB_TONE	AB_TONE
CEO Early Years	-1.427***	-2.112***	-3.324**
	[-3.144]	[-3.506]	[-2.171]
HighLitigation		-2.546***	-1.874*
		[-2.629]	[-1.821]
NOA		6.257***	5.329***
		[4.976]	[3.932]
CEO Optimism		3.844***	3.844***
		[6.117]	[6.117]
ACC		7.747***	7.763***
		[3.036]	[3.034]
CEO Early Years*HighLitigation			-2.180*
			[-1.718]
CEO Early Years*NOA			3.268
			[1.514]
Constant	2.963**	-4.331***	-3.974***
	[2.426]	[-4.299]	[-3.762]
Fixed effects	Firm & Year	Year	Year
Clustering	By Firm	By Firm	By Firm
Observations	17,725	15,068	15,068
Adjusted R-squared	0.460	0.018	0.019

TABLE 4: First five years of CEOs' tenure and tone management

This table reports the coefficient estimates of a regression of 10-K abnormal tone (*AB_TONE*) on indicator variables for each of the first five years of CEOs' tenure. *CEO First Year* takes the value of 1 if the observation is for the first year of CEOs' tenure and 0 otherwise; *CEO Second Year* takes the value of 1 if the observation is for the second year of CEOs' tenure and 0 otherwise; and so on. All other variable definitions are outlined in Appendix A. Model specification (1) includes both year and industry fixed effects, while specification (2) only includes year fixed effects. t-statistics (in brackets) are based on standard errors that are clustered at the firm level. ***, **, * represents statistical significance at the 1%, 5% and 10% levels.

Model	(1)	(2)
Dependent variable	AB_TONE	AB_TONE
CEO First Year	-5.121***	-3.277***
	[-6.487]	[-3.506]
CEO Second Year	-4.297***	-2.452***
	[-6.136]	[-3.124]
CEO Third Year	-3.050***	-2.383***
	[-4.438]	[-3.020]
CEO Fourth Year	-2.484***	-1.913**
	[-3.641]	[-2.424]
CEO Fifth Year	-2.105***	-2.051**
	[-2.924]	[-2.509]
HighLitigation		-2.573***
		[-2.661]
NOA		6.183***
		[4.914]
CEO Optimism		3.658***
		[5.720]
ACC		7.600***
		[2.973]
Constant	1.074	-3.724***
	[0.940]	[-3.524]
Fixed effects	Industry & Year	Year
Clustering	By Firm	By Firm
Observations	17,725	15,068
Adjusted R-squared	0.072	0.018

TABLE 5: Early years of CEOs' tenure and forward-looking disclosures

This table reports the coefficient estimates of a regression of 10-K forward-looking intensity (*FLI*) on *CEO Early Years* and determinants of forward-looking disclosures. The sample period is from 1993 to 2010. All variable definitions are outlined in Appendix A. Model specification (1) includes industry and year fixed effects, while specification (2) includes firm and year fixed effects. t-statistics (in brackets) are based on standard errors that are clustered at the firm level. ***, **, * represents statistical significance at the 1%, 5% and 10% levels.

Model	(1)	(2)
Dependent variable	FLI	FLI
CEO Early Years	-0.001**	-0.002***
	[-2.515]	[-2.792]
ANALYST	0.003***	0.002*
	[3.876]	[1.720]
SIZE	-0.001**	-0.001
	[-2.298]	[-1.049]
EARN	-0.018***	-0.013***
	[-6.135]	[-3.998]
LOSS	0.004***	0.001
	[5.056]	[1.605]
RET VOL	0.013***	0.005***
	[8.508]	[3.093]
EARN VOL	0.005**	0.001
	[2.522]	[0.918]
SI	0.002	0.001
	[0.727]	[0.316]
MA	0.001	0.000
	[1.586]	[0.011]
BTM	0.000	0.000
	[0.027]	[0.138]
AGE	-0.002***	-0.007**
	[-2.649]	[-2.222]
ln BUS SEG	-0.002***	-0.000
	[-2.700]	[-0.377]
ln GEOG SEG	0.001	-0.001
	[0.163]	[-0.157]
Constant	0.090***	0.107***
	[26.486]	[11.232]
Fixed effects	Industry & Year	Firm & Year
Clustering	By Firm	By Firm
Observations	13,224	13,224
Adjusted R-squared	0.264	0.504

TABLE 6: Early years of high-ability CEOs' tenure and tone management

This table reports the coefficient estimates of a regression of 10-K abnormal tone (*AB_TONE*) on the *CEO Early Years* variable and a set of controls, including the managerial ability measure (*ABILITY SCORE*) and the corresponding interaction term. The sample period is from 1993 to 2010. All variable definitions are outlined in Appendix A. The model specification includes only year fixed effects. t-statistics (in brackets) are based on standard errors that are clustered at the firm level. ***, **, * represents statistical significance at the 1%, 5% and 10% levels.

Dependent variable	AB_TONE
CEO Early Years	-3.401***
	[-2.742]
ABILITY SCORE	6.546**
	[2.538]
CEO Early Years * ABILITY SCORE	6.858*
	[1.955]
HighLitigation	-1.773*
	[-1.879]
CEO Early Years * HighLitigation	-2.118*
	[-1.846]
NOA	7.119***
	[6.099]
CEO Early Years * NOA	4.263***
	[2.591]
CEO Optimism	3.778***
	[6.378]
ACC	8.451***
	[3.589]
Constant	-3.785***
	[-3.815]
Fixed effects	Year
Clustering	By Firm
Observations	13,556
Adjusted R-squared	0.030

TABLE 7: Final year of CEOs' tenure and tone management

This table reports the coefficient estimates of a regression of 10-K abnormal tone (*AB_TONE*) on the *CEO Final Year* variable and a set of controls. The sample period is from 1993 to 2010. All variable definitions are outlined in Appendix A. Model specification (1) includes both year and firm fixed effects, while specifications (2) - (4) only include year fixed effects. t-statistics (in brackets) are based on standard errors that are clustered at the firm level. ***, **, * represents statistical significance at the 1%, 5% and 10% levels.

Model	(1)	(2)	(3)	(4)
Dependent variable	AB_TONE	AB_TONE	AB_TONE	AB_TONE
CEO Final Year	-2.526***	-4.314***	-11.676***	-11.725***
	[-4.187]	[-5.608]	[-6.286]	[-6.336]
HighLitigation		-2.278*	-2.280*	-1.067
		[-1.844]	[-1.766]	[-0.766]
NOA		5.981***	3.750**	2.528
		[3.871]	[2.298]	[1.422]
CEO Optimism		4.457***	4.456***	3.827***
		[5.678]	[5.696]	[4.852]
ACC		8.973***	9.386***	9.038***
		[2.658]	[2.791]	[2.684]
CEO Final Year*HighLitigation			0.163	0.079
			[0.097]	[0.047]
CEO Final Year*NOA			12.637***	12.533***
			[5.141]	[5.153]
CEO Early Years				-3.557*
				[-1.909]
CEO Early Years*HighLitigation				-3.732**
				[-2.396]
CEO Early Years*NOA				3.363
				[1.308]
Constant	2.874**	-4.483***	-3.063**	-1.663
	[2.127]	[-3.681]	[-2.433]	[-1.196]
Fixed effects	Firm & Year	Year	Year	Year
Clustering	By Firm	By Firm	By Firm	By Firm
Observations	11,212	9,282	9,282	9,282
Adjusted R-squared	0.471	0.021	0.024	0.028

TABLE 8: 10-K abnormal tone and future financial performance

This table reports the coefficient estimates of a regression of future financial performance measures (*EARN* or *CFO*) on *AB_TONE* and a set of controls. The sample period is from 1993 to 2010. All variable definitions are outlined in Appendix A. All model specifications include both year and industry fixed effects. t-statistics (in brackets) are based on standard errors that are clustered at the firm level. ***, **, * represents statistical significance at the 1%, 5% and 10% levels.

Model	(1)	(2)	(3)	(4)	(5)	(6)
Dependent variable	EARN (t+1)	EARN (t+2)	EARN (t+3)	CFO (t+1)	CFO (t+2)	CFO (t+3)
AB_TONE	-0.00008*	-0.00013**	-0.00015**	-0.00008*	-0.00014***	-0.00014**
	[-1.901]	[-2.343]	[-2.165]	[-1.841]	[-2.687]	[-2.047]
DA	-0.188***	-0.131***	-0.137***	-0.202***	-0.161***	-0.162***
	[-7.986]	[-5.899]	[-5.887]	[-7.089]	[-6.918]	[-9.120]
EARN	0.629***	0.517***	0.427***	0.525***	0.496***	0.415***
	[14.433]	[12.738]	[9.028]	[11.655]	[13.066]	[8.426]
SIZE	0.003***	0.004***	0.008***	0.005***	0.005***	0.007***
	[2.634]	[3.619]	[5.835]	[4.817]	[5.060]	[5.051]
BTM	-0.019***	-0.016***	-0.012*	-0.011***	-0.014***	-0.016***
	[-5.127]	[-4.132]	[-1.759]	[-3.608]	[-3.358]	[-3.505]
RET	0.012**	-0.004	-0.010**	0.007**	-0.005	-0.004
	[2.304]	[-0.816]	[-2.318]	[2.139]	[-1.333]	[-1.017]
RET VOL	-0.057***	-0.074***	-0.056***	-0.023***	-0.039***	-0.038***
	[-5.964]	[-6.593]	[-4.953]	[-2.717]	[-4.340]	[-4.029]
EARN VOL	-0.007	0.014	0.010	0.031***	0.035***	0.027**
	[-0.579]	[1.090]	[0.789]	[3.740]	[3.187]	[2.418]
Constant	0.041***	0.036***	-0.007	0.066***	0.069***	0.073***
	[3.986]	[3.715]	[-0.510]	[7.034]	[6.305]	[6.548]
Industry & Year Effects	Yes	Yes	Yes	Yes	Yes	Yes
Clustering by Firm	Yes	Yes	Yes	Yes	Yes	Yes
Observations	13,723	11,843	10,070	13,721	11,062	9,445
Adjusted R-squared	0.435	0.324	0.255	0.402	0.368	0.297

TABLE 9: Robustness of results

This table reports the regression coefficients from an estimation of model (2) using alternate definitions of *CEO Early Years* (Columns 1 and 2) and alternate proxies for *CEO Optimism* (Columns 3 and 4). The sample period is from 1993 to 2010. All variable definitions are outlined in Appendix A. All specifications include year fixed effects. t-statistics (in brackets) are based on standard errors that are clustered at the firm level. ***, **, * represents statistical significance at the 1%, 5% and 10% levels.

Model	(1)	(2)	(3)	(4)
Dependent variable	AB_TONE	AB_TONE	AB_TONE	AB_TONE
CEO Early Years	-1.126*	-2.163***	-3.912**	-3.538**
	[-1.649]	[-2.903]	[-2.555]	[-2.326]
HighLitigation	-1.589	-1.735*	-1.815*	-2.030*
	[-1.550]	[-1.697]	[-1.749]	[-1.957]
CEO Early Years*HighLitigation	-3.158***	-2.747**	-2.290*	-2.275*
	[-2.696]	[-2.309]	[-1.798]	[-1.791]
NOA	6.221***	5.776***	5.098***	4.159***
	[4.853]	[4.470]	[3.725]	[2.968]
CEO Early Years*NOA	0.146	1.531	3.145	2.741
	[0.127]	[1.359]	[1.460]	[1.279]
ACC	7.689***	7.683***	8.970***	6.970***
	[3.010]	[3.004]	[3.540]	[2.725]
CEO Optimism	3.861***	3.799***		
	[6.087]	[6.032]		
CEO Optimism_CAPX			2.198***	
			[3.235]	
CEO Optimism_XSINVEST				3.467***
				[6.212]
Constant	-4.760***	-4.069***	-3.685***	-3.833***
	[-4.764]	[-3.932]	[-3.425]	[-3.686]
Fixed effects	Year	Year	Year	Year
Clustering	By Firm	By Firm	By Firm	By Firm
Observations	15,068	15,068	14,999	14,997
Adjusted R-squared	0.018	0.019	0.016	0.018

TABLE 10: Early years of CEOs' tenure and 10-K readability

This table reports regression coefficients from the estimation of model (5). The sample period is from 1993 to 2010. All variable definitions are outlined in Appendix A. All model specifications in Panel B include both year and industry fixed effects. t-statistics (in brackets) are based on standard errors that are clustered at the firm level. ***, **, * represents statistical significance at the 1%, 5% and 10% levels.

Panel A: Summary statistics

Variable	n	Mean	Median	S.D.	Min	Max
10K Wordcount	18576	54521.65	41100	47788.44	9152	320000
FOG	15863	19.57	19.39	1.62	4.72	34.95
LENGTH	18576	10.66	10.62	0.67	9.12	12.67
FILESIZE	18576	1.48	1.03	1.65	0.12	10.04
SEO	18576	0.03	0.00	0.18	0.00	1.00
DLW	18283	0.62	1.00	0.48	0.00	1.00

Panel B: Regression output

Model Dependent variable	(1) FOG	(2) LENGTH	(3) FILESIZE
CEO Early Years	-0.119***	-0.005	0.040***
	[-3.824]	[-0.422]	[3.002]
SIZE	0.067***	0.109***	0.117***
	[3.812]	[17.823]	[17.350]
BTM	0.058*	0.090***	0.092***
	[1.829]	[5.971]	[5.783]
AGE	-0.004	0.002	0.047***
	[-0.096]	[0.126]	[2.847]
SI	-0.648***	-0.608***	-0.505***
	[-3.008]	[-5.743]	[-4.874]
RET VOL	0.407***	0.269***	0.230***
	[5.213]	[9.515]	[7.456]
EARN VOL	0.219***	0.108***	0.140***
	[2.949]	[3.598]	[4.008]
ln BUS SEG	0.077	0.027	0.029
	[1.415]	[1.468]	[1.398]
ln GEOG SEG	0.021	-0.099	0.011
	[0.068]	[-0.994]	[0.082]
SEO	-0.141*	0.020	0.003
	[-1.660]	[0.758]	[0.098]
MA	0.079**	0.030**	0.046***
	[2.329]	[2.380]	[3.151]
DLW	0.052	0.079***	0.074***
	[0.999]	[4.397]	[3.490]
Constant	18.299***	9.255***	-2.439***
	[90.926]	[126.027]	[-32.185]
Fixed effects	Industry & Year	Industry & Year	Industry & Year
Clustering	By Firm	By Firm	By Firm
Observations	11,801	13,706	13,731
Adjusted R-squared	0.095	0.307	0.631

	Dropped	Samples Size
SEC 10-K files 1994 to 2006		87,132
Drop if number of words in 10-K < 2,000	5,667	81,465
Merge with Compustat	43,863	37,602
(Exclude firms with negative value of book equity)		
(Exclude firms with only one year of data in Compustat)		
Drop financial firms (Two digits SIC: 6000-6999)	10,867	26,735
Include only first filing in a given year	230	26,505
At least 180 days between a given firm's 10-K filings	258	26,247
Missing Control Variables	5,592	20,655
Firm -Year Sample		20,655
Number of unique firms		3,703

Figure 1: The selection of 10-K sample

TABLE 1: Summary Statistics

This table provides summary statistics for the overall sample of 20,655 firm-year observations from 1993-2006. All variable definitions are outlined in Appendix A1.

Variable	n	Mean	S.D.	0.25 Q	Median	0.75 Q
ROA	20,655	0.03	0.12	0.01	0.05	0.09
Sales ($ millions)	20,655	1,940.36	5,336.01	89.48	320.94	1,174.64
Assets ($ millions)	20,655	1,931.82	5,636.34	83.56	289.58	1,076.53
Market Value ($ millions)	20,644	2,683.00	8,897.27	68.25	298.78	1,247.85
R&D Exp ($ millions)	20,655	38.94	159.03	0.00	0.00	8.80
Advertising Exp ($ millions)	20,655	23.53	111.22	0.00	0.00	0.90
Capital Exp ($ millions)	20,655	108.73	345.30	2.79	12.44	55.47
Patent	20,655	6.94	26.15	0.00	0.00	1.00
Total Words (10K Length)	20,655	37,098	29,085	18,517	28,690	45,422
Total Sentences (10K)	20,655	1,379	956	780	1,151	1,680
Forward Looking Tone (FLS Tone)	20,655	-0.06	0.05	-0.09	-0.05	-0.02
R&D related Words	20,655	167	146	66	123	222
R&D Sentences	20,655	11	17	1	4	14
ln (R&D QTY)	20,655	1.50	1.00	0.61	1.37	2.25
ln (R&D Sentences)	20,655	1.66	1.31	0.69	1.61	2.71

TABLE 2: R&D Disclosure Quantity and Future Profitability

Panels A and B report the coefficient estimates of a regression of subsequent-period ROA on current R&D disclosure quantity for both word-count-based and sentence-count-based measures, respectively. All variable definitions are outlined in Appendix A1. The specification includes year and industry fixed effects. t-statistics (in brackets) are based on standard errors that are clustered at the firm and year level. ***, **, * represents statistical significance at the 1%, 5% and 10% levels.

Panel A: Word-count-based measure of R&D disclosure quantity

VARIABLES	ROA (t+1)			
	(1)	(2)	(3)	(4)
ln (R&D QTY)	-0.006***	-0.006***	-0.007***	-0.006***
	[-5.721]	[-5.597]	[-5.869]	[-5.707]
ROA	0.649***	0.644***	0.621***	0.623***
	[31.531]	[32.675]	[33.144]	[34.166]
Δ ROA	-0.109***	-0.107***	-0.097***	-0.107***
	[-6.529]	[-6.555]	[-5.886]	[-6.270]
ln (1+ Patents/RDC)		0.031***	0.024***	0.021***
		[4.995]	[3.767]	[3.216]
ln (1+ R&D Exp/Asset)		-0.025	-0.010	-0.009
		[-0.668]	[-0.267]	[-0.231]
R&D Growth Dummy		0.002	0.002	0.002
		[0.573]	[0.501]	[0.655]
Δ APC		0.326	0.120	0.118
		[1.546]	[0.723]	[0.780]
ln (1+AD/Asset)			0.045	0.042
			[1.603]	[1.461]
ln(1+Capex/Asset)			-0.023	-0.014
			[-0.877]	[-0.569]
ln (Asset)			0.005***	0.006***
			[10.067]	[10.132]
FLS Tone			-0.025	-0.023
			[-1.749]	[-1.578]
10K Length			-0.010***	-0.010***
			[-6.116]	[-6.091]
Fixed Effects:				
Year	Yes	Yes	Yes	Yes
Industry	Yes	Yes	Yes	Yes
Industry * Year	No	No	No	Yes
Clustering of SE	Firm & Year	Firm & Year	Firm & Year	Firm & Year
Observations	20,655	20,655	20,655	20,655
Adjusted R-squared	0.421	0.421	0.427	0.435

Panel B: Sentence-count-based measure of R&D disclosure quantity (Merkley 2014)

VARIABLES	ROA (t+1)			
	(1)	(2)	(3)	(4)
ln (R&D Sentences)	-0.004***	-0.004***	-0.005***	-0.005***
	[-3.783]	[-4.226]	[-5.318]	[-5.236]
ROA	0.650***	0.645***	0.622***	0.625***
	[31.576]	[32.383]	[32.793]	[33.816]
Δ ROA	-0.109***	-0.108***	-0.097***	-0.107***
	[-6.576]	[-6.596]	[-5.897]	[-6.295]
ln (1+ Patents/RDC)		0.032***	0.024***	0.021***
		[5.165]	[3.920]	[3.353]
ln (1+ R&D Exp/Asset)		-0.020	-0.002	-0.001
		[-0.521]	[-0.051]	[-0.038]
R&D Growth Dummy		0.002	0.002	0.002
		[0.675]	[0.581]	[0.747]
Δ APC		0.349	0.124	0.123
		[1.577]	[0.728]	[0.785]
ln (1+AD/Asset)			0.051*	0.048
			[1.836]	[1.692]
ln(1+Capex/Asset)			-0.025	-0.017
			[-0.958]	[-0.652]
ln (Asset)			0.006***	0.006***
			[10.189]	[10.192]
FLS Tone			-0.025	-0.023
			[-1.716]	[-1.544]
10K Length			-0.007***	-0.007***
			[-4.959]	[-5.003]
Fixed Effects:				
Year	Yes	Yes	Yes	Yes
Industry	Yes	Yes	Yes	Yes
Industry * Year	No	No	No	Yes
Clustering of SE	Firm & Year	Firm & Year	Firm & Year	Firm & Year
Observations	20,655	20,655	20,655	20,655
Adjusted R-squared	0.420	0.421	0.426	0.434

TABLE 3: Historical R&D Disclosure Quantity and Subsequent ROA

This table reports the coefficient estimates of a regression of subsequent-period ROA on current (column 1), one-year-lagged (column 2), two-years-lagged (column 3), three-years-lagged (column 4), and four-years-lagged (column 5) R&D disclosure quantity. All variable definitions are outlined in Appendix A1. Each specification includes year and industry fixed effects. t-statistics (in brackets) are based on standard errors that are clustered at the firm and year level. ***, **, * represents statistical significance at the 1%, 5% and 10% levels.

VARIABLES	ROA (t+1)				
	(1)	(2)	(3)	(4)	(5)
ln (R&D QTY) (t)	-0.007***				
	[-5.869]				
ln (R&D QTY) (t-1)		-0.005***			
		[-3.461]			
ln (R&D QTY) (t-2)			-0.005***		
			[-3.995]		
ln (R&D QTY) (t-3)				-0.002*	
				[-1.991]	
ln (R&D QTY) (t-4)					-0.004
					[-1.723]
ROA	0.621***	0.630***	0.636***	0.631***	0.621***
	[33.144]	[28.891]	[31.121]	[36.114]	[24.872]
Δ ROA	-0.097***	-0.088***	-0.098***	-0.109***	-0.090***
	[-5.886]	[-3.679]	[-3.643]	[-4.507]	[-3.589]
ln (1+ Patents/RDC)	0.024***	0.028***	0.025***	0.024***	0.028**
	[3.767]	[4.693]	[4.204]	[4.046]	[3.182]
ln (1+ R&D Exp/Asset)	-0.010	-0.021	-0.009	-0.018	-0.045
	[-0.267]	[-0.407]	[-0.205]	[-0.336]	[-0.998]
R&D Growth Dummy	0.002	-0.001	-0.002	-0.001	-0.000
	[0.501]	[-0.272]	[-0.571]	[-0.328]	[-0.019]
Δ APC	0.120	0.074	0.181	0.058	-0.166
	[0.723]	[0.389]	[0.816]	[0.329]	[-0.643]
ln (1+AD/Asset)	0.045	0.012	0.024	-0.011	0.004
	[1.603]	[0.524]	[1.307]	[-0.320]	[0.099]
ln(1+Capex/Asset)	-0.023	-0.024	-0.001	-0.004	-0.007
	[-0.877]	[-0.911]	[-0.043]	[-0.202]	[-0.353]
ln (Asset)	0.005***	0.005***	0.005***	0.005***	0.005***
	[10.067]	[11.729]	[8.374]	[8.499]	[14.452]
FLS Tone	-0.025	-0.031*	-0.031**	-0.028**	-0.031**
	[-1.749]	[-2.025]	[-2.380]	[-2.911]	[-2.807]
10K Length	-0.010***	-0.009***	-0.009***	-0.010***	-0.010***
	[-6.116]	[-6.354]	[-5.059]	[-5.315]	[-6.525]
Fixed Effects:					
Year	Yes	Yes	Yes	Yes	Yes
Industry	Yes	Yes	Yes	Yes	Yes
Clustering of SE	Firm & Year	Firm & Year	Firm & Year	Firm & Year	Firm & Year
Observations	20,655	15,724	13,092	10,726	8,716
Adjusted R-squared	0.427	0.427	0.431	0.426	0.424

TABLE 4: Measurement error tests for the association between R&D disclosure quantity and ROA

This table reports the summary statistics (Panel A) and coefficient estimates (Panel B) of a regression of subsequent-period ROA on current R&D disclosure quantity after including two measures of negative R&D sentiment [*R&D Pessimism* (columns 1) and *R&D Uncertainty* (columns 2)]. All variable definitions are outlined in Appendix A1. Each specification includes year and industry fixed effects. t-statistics (in brackets) are based on standard errors that are clustered at the firm and year level. ***, **, * represents statistical significance at the 1%, 5% and 10% levels.

Panel A: Summary Statistics

Variable	n	Mean	S.D.	0.25 Q	Median	0.75 Q
R&D Pessimism	15,700	24.88	51.08	0.00	2.76	27.91
R&D Uncertainty	15,700	4.89	12.79	0.00	0.26	3.91

Panel B: Regression Results

VARIABLES	ROA (t+1)		
	(1)	(2)	(3)
ln (R&D QTY)	-0.009***	-0.009***	-0.009***
	[-5.567]	[-5.586]	[-5.596]
R&D Pessimism	-0.000		-0.000
	[-0.131]		[-0.160]
ln (R&D QTY) * R&D Pessimism	0.000		0.000
	[0.305]		[0.215]
R&D Uncertainty		0.000	0.000
		[1.005]	[0.977]
ln (R&D QTY) * R&D Uncertainty		0.000	0.000
		[0.961]	[0.868]
ROA	0.614***	0.614***	0.614***
	[29.975]	[30.206]	[30.029]
Δ ROA	-0.102***	-0.102***	-0.102***
	[-5.972]	[-5.970]	[-5.967]
ln (1+ Patents/RDC)	0.024***	0.025***	0.025***
	[3.855]	[3.894]	[3.878]
ln (1+ R&D Exp/Asset)	-0.008	-0.008	-0.008
	[-0.208]	[-0.216]	[-0.217]
R&D Growth Dummy	0.002	0.002	0.002
	[0.590]	[0.577]	[0.577]
Δ APC	0.126	0.125	0.124
	[0.799]	[0.789]	[0.787]
ln (1+AD/Asset)	0.034*	0.034*	0.034*
	[1.815]	[1.818]	[1.814]
ln(1+Capex/Asset)	-0.012	-0.010	-0.010
	[-0.351]	[-0.315]	[-0.314]
ln (Asset)	0.006***	0.006***	0.006***
	[7.873]	[7.891]	[7.922]
FLS Tone	-0.045**	-0.044**	-0.044**
	[-2.383]	[-2.355]	[-2.346]
10K Length	-0.012***	-0.012***	-0.012***
	[-7.117]	[-7.210]	[-7.184]
Fixed Effects:			
Year	Yes	Yes	Yes
Industry	Yes	Yes	Yes
Clustering of SE	Firm & Year	Firm & Year	Firm & Year
Observations	15,700	15,700	15,700
Adjusted R-squared	0.427	0.428	0.427

TABLE 5: Examining the effect of manager's strategic disposition (Fog) on the association between R&D disclosure quantity and ROA

This table reports the summary statistics for the Fog measure (Panel A) and coefficient estimates (Panel B) of a regression of subsequent-period ROA on current R&D disclosure quantity after including the Fog measure and a corresponding interaction term. All variable definitions are outlined in Appendix A1. Each specification includes year and industry fixed effects. t-statistics (in brackets) are based on standard errors that are clustered at the firm and year level. ***, **, * represents statistical significance at the 1%, 5% and 10% levels.

Panel A: Summary Statistics

Variable	n	Mean	S.D.	0.25 Q	Median	0.75 Q
Fog Index	17,992	19.35	1.52	18.38	19.18	20.10

Panel B: Regression Results

VARIABLES	ROA (t+1)
ln (R&D QTY)	-0.006***
	[-5.049]
Fog Index	0.001**
	[2.437]
ln (R&D QTY) * Fog Index	-0.001
	[-1.559]
ROA	0.610***
	[34.376]
Δ ROA	-0.089***
	[-5.583]
ln (1+ Patents/RDC)	0.023***
	[3.666]
ln (1+ R&D Exp/Asset)	-0.017
	[-0.399]
R&D Growth Dummy	0.003
	[0.936]
Δ APC	0.108
	[0.595]
ln (1+AD/Asset)	0.047
	[1.560]
ln(1+Capex/Asset)	-0.017
	[-0.678]
ln (Asset)	0.005***
	[9.112]
FLS Tone	-0.021
	[-1.299]
10K Length	-0.010***
	[-6.024]
Fixed Effects:	
Year	Yes
Industry	Yes
Clustering of SE	Firm & Year
Observations	17,992
Adjusted R-squared	0.415

TABLE 6: Examining the effect of competition on the association between R&D disclosure quantity and ROA

This table reports the summary statistics for the three competition measures (Panel A) and coefficient estimates (Panel B) of a regression of subsequent-period ROA on current R&D disclosure quantity after including the respective competition measure [HHI (column 1); PC_MARGIN (column 2); MKT_SIZE (column 3)] and a corresponding interaction term. All variable definitions are outlined in Appendix A1. Each specification includes year fixed effects. t-statistics (in brackets) are based on standard errors that are clustered at the firm and year level. ***, **, * represents statistical significance at the 1%, 5% and 10% levels.

Panel A: Summary Statistics

Variable	n	Mean	S.D.	0.25 Q	Median	0.75 Q
HHI	20,133	0.16	0.14	0.07	0.12	0.19
PC_MARGIN	12,455	1.07	0.22	1.02	1.07	1.15
MKT_SIZE	18,090	9.83	2.20	8.81	10.15	11.34

Panel B: Regression Results

VARIABLES	ROA (t+1)		
	(1)	(2)	(3)
	HHI	PC_MARGIN	MKT_SIZE
ln (R&D QTY)	-0.005***	-0.006***	-0.005***
	[-3.426]	[-3.158]	[-3.532]
HHI	-0.012		
	[-1.058]		
HHI * ln (R&D QTY)	0.008		
	[0.999]		
PC_MARGIN		0.016*	
		[2.037]	
PC_MARGIN * ln (R&D QTY)		-0.000	
		[-0.020]	
MKT_SIZE			0.001
			[1.538]
MKT_SIZE * ln (R&D QTY)			-0.001
			[-1.113]
ROA	0.631***	0.622***	0.626***
	[35.432]	[28.003]	[32.121]
Δ ROA	-0.098***	-0.084***	-0.100***
	[-6.214]	[-4.420]	[-6.099]
ln (1+ Patents/RDC)	0.029***	0.036***	0.030***
	[4.333]	[4.480]	[4.106]
ln (1+ R&D Exp/Asset)	-0.006	-0.015	-0.003
	[-0.153]	[-0.309]	[-0.080]
R&D Growth Dummy	0.001	0.003	0.001
	[0.285]	[0.531]	[0.323]
Δ APC	0.077	0.132	0.165
	[0.427]	[0.507]	[0.899]
ln (1+AD/Asset)	0.045	0.050	0.043
	[1.752]	[1.424]	[1.531]
ln(1+Capex/Asset)	-0.013	-0.011	-0.015
	[-0.434]	[-0.279]	[-0.474]
ln (Asset)	0.005***	0.005***	0.005***
	[9.261]	[6.846]	[8.904]
FLS Tone	-0.031	-0.045*	-0.034*
	[-1.774]	[-2.010]	[-1.951]
10K Length	-0.010***	-0.010***	-0.009***
	[-5.857]	[-3.732]	[-5.428]
Fixed Effects:			
Year	Yes	Yes	Yes
Clustering of SE	Firm & Year	Firm & Year	Firm & Year
Observations	20,133	12,455	18,090
Adjusted R-squared	0.422	0.430	0.417

TABLE 7: Decomposing R&D disclosures into forward-looking (FLS) and non-forward-looking (non-FLS)

Panel A of the table reports the summary statistics for the count of forward-looking (FLS), non-forward-looking (Non-FLS), and total R&D-related sentences. Panel B reports the coefficient estimates of a regression of subsequent-period ROA on both FLS and Non-FLS R&D disclosure quantity. All variable definitions are outlined in Appendix A1. Each specification includes year and industry fixed effects. t-statistics (in brackets) are based on standard errors that are clustered at the firm and year level. ***, **, * represents statistical significance at the 1%, 5% and 10% levels.

Panel A: Summary Statistics

Variable	n	Mean	S.D.	0.25 Q	Median	0.75 Q
Forward Looking Sentences (FLS)	20,655	1.79	3.58	0.00	0.00	2.00
Non-Forward Looking Sentences (Non-FLS)	20,655	9.11	13.60	0.00	3.00	13.00
Total R&D Related Sentences	20,655	10.94	16.94	1.00	4.00	14.00

Panel B: Regression Results

VARIABLES	ROA (t+1)
ln (FLS)	-0.010***
	[-5.400]
ln (Non-FLS)	-0.001
	[-1.017]
ROA	0.620***
	[32.969]
Δ ROA	-0.096***
	[-5.877]
ln (1+ Patents/RDC)	0.022***
	[3.592]
ln (1+ R&D Exp/Asset)	0.024
	[0.662]
R&D Growth Dummy	0.002
	[0.680]
Δ APC	0.173
	[1.032]
ln (1+AD/Asset)	0.047
	[1.663]
ln(1+Capex/Asset)	-0.024
	[-0.928]
ln (Asset)	0.006***
	[10.388]
FLS Tone	-0.020
	[-1.397]
10K Length	-0.006***
	[-4.628]
Fixed Effects:	
Year	Yes
Industry	Yes
Clustering of SE	Firm & Year
Observations	20,655
Adjusted R-squared	0.428

TABLE 8: The association between R&D Disclosure Quantity and ROA for only R&D-intensive firms

This table reports the coefficient estimates of a regression of subsequent-period ROA on current R&D disclosure quantity run on the sample of only R&D-intensive firms. All variable definitions are outlined in Appendix A1. The specification includes year and industry fixed effects. t-statistics (in brackets) are based on standard errors that are clustered at the firm and year level. ***, **, * represents statistical significance at the 1%, 5% and 10% levels.

VARIABLES	ROA (t+1)
ln (R&D QTY) $_{\text{R\&D INTENSIVE}}$	-0.011***
	[-5.440]
ROA	0.607***
	[30.006]
Δ ROA	-0.106***
	[-4.941]
ln (1+ Patents/RDC)	0.029***
	[3.528]
ln (1+ R&D Exp/Asset)	-0.003
	[-0.078]
R&D Growth Dummy	0.001
	[0.309]
Δ APC	0.114
	[0.658]
ln (1+AD/Asset)	0.064
	[1.672]
ln(1+Capex/Asset)	-0.009
	[-0.239]
ln (Asset)	0.006***
	[6.116]
FLS Tone	-0.029
	[-1.075]
10K Length	-0.014***
	[-5.572]
Fixed Effects:	
Year	Yes
Industry	Yes
Clustering of SE	Firm & Year
Observations	9,205
Adjusted R-squared	0.433

TABLE 9: Examining the association between R&D Disclosure Quantity and ROA using the extended patent database

Both columns of this table report the coefficient estimates of a regression of subsequent-period ROA on current R&D disclosure quantity using Kogan et al. (2014)'s extended patent database (available until 2010). The IE measure has been computed using the patent filing date in the first column, and the patent issue date in the second column. All variable definitions are outlined in Appendix A1. The specification includes year and industry fixed effects. t-statistics (in brackets) are based on standard errors that are clustered at the firm and year level. ***, **, * represents statistical significance at the 1%, 5% and 10% levels.

VARIABLES	**ROA (t+1)**	
	(1)	**(2)**
ln (R&D QTY)	-0.006***	-0.006***
	[-5.450]	[-5.416]
ROA	0.613***	0.613***
	[25.180]	[25.184]
Δ ROA	-0.114***	-0.114***
	[-6.601]	[-6.592]
ln (1+ Patents Filed/RDC)	-0.010	
	[-1.073]	
ln (1+ Patents Issued/RDC)		-0.018
		[-1.546]
ln (1+ R&D Exp/Asset)	-0.031	-0.031
	[-0.832]	[-0.840]
R&D Growth Dummy	0.002	0.002
	[0.655]	[0.644]
Δ APC	0.146	0.152
	[0.769]	[0.797]
ln (1+AD/Asset)	0.047	0.047
	[1.730]	[1.728]
ln(1+Capex/Asset)	-0.003	-0.003
	[-0.097]	[-0.097]
ln (Asset)	0.006***	0.006***
	[10.414]	[10.400]
FLS Tone	-0.028**	-0.028**
	[-2.150]	[-2.182]
10K Length	-0.010***	-0.010***
	[-7.029]	[-6.984]
Fixed Effects:		
Year	Yes	Yes
Industry	Yes	Yes
Clustering of SE	Firm & Year	Firm & Year
Observations	26,235	26,235
Adjusted R-squared	0.415	0.415